

**Exploring the quality, usability, and use of learning objects in introductory statistics
classrooms**

by

Rachel Jean Graham

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Co-majors: School of Education (Mathematics Education); Human Computer Interaction

Program of Study Committee:

Barbara Dougherty, Co-major Professor (Mathematics Education)

Anne Foegen, Co-major Professor (Education)

Stephen Gilbert, Co-major Professor (Human Computer Interaction)

Alejandro Andreotti (Mathematics)

Connie Hargrave (Education)

Iowa State University

Ames, Iowa

2014

Copyright © Rachel Jean Graham, 2014. All rights reserved.

DEDICATION

I dedicate this dissertation to my husband Mike, who has remained my devoted partner through this entire journey, to my children Brandy, Michael, Kassie, Nathan, and Lexie, who are the main reason for everything I do, and to my students who help me get better every day.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	i
ABSTRACT	ii
CHAPTER 1. GENERAL INTRODUCTION	1
Statement of the Problem	3
Purpose of the Study	4
Dissertation Format	5
References	8
CHAPTER 2. LEARNING OBJECT USE IN INTRODUCTORY STATISTICS CLASSROOMS: A SURVEY OF CURRENT PRACTICES, TEACHERS' PERCEPTIONS AND ADVICE TO PEERS	10
Abstract	10
Theoretical Framework	11
Literature Review	15
Method	17
Results	19
Discussion	24
Conclusion	28
Appendix	29
References	32
CHAPTER 3. EVALUTING THE DESIGN QUALITY OF SAMPLING DISTRIBUTION LEARNING OBJECTS: AN EXPERT EVALUATION	39
Abstract	39
Introduction	39
Existing Evaluation Systems for Learning Objects	40
Method	45
Results	53
Discussion	60
Conclusion	62
References	63
Appendix	67
CHAPTER 4. USABILITY OF A CONFIDENCE INTERVAL APPLET AS EVALUATED BY COLLEGE STUDENTS	69
Abstract	69
Introduction	69
Previous Work Evaluating Learning Objects	70
Method	72
Results	78
Discussion	82

	Page
Conclusion	87
References	89
Appendix	92
CHAPTER 5. GENERAL CONCLUSIONS	93
Introduction	93
Implications for Theory and the Literature	95
Implications for Practice	97
Conclusion	98
References	100

ACKNOWLEDGEMENTS

First, I am most and forever grateful to God for blessing me with the gifts I have been given and to my prayer warriors for lifting me throughout the years of my journey. I would also like to thank my family for loving me through everything and being the best cheerleaders anyone could ask for. In addition, I would like to thank my committee members for their guidance and support throughout the course of my studies. Finally, thank you to Matanuska-Susitna College (University of Alaska Anchorage) for providing my family and I the most wonderful environment for the next chapters in our lives.

ABSTRACT

The studies included in this dissertation were completed for the purposes of exploring the nature of learning object (LO) use in introductory statistics classrooms and providing information about the quality of existing, freely-accessible online statistics LOs. Publication-ready manuscripts were written to describe the procedures and outcomes of each study. The first paper summarizes the results of a survey of LO use completed by high school and post-secondary statistics educators. The second paper in this dissertation describes an expert evaluation to investigate the quality of existing statistics applets designed to explain sampling distributions of sample means. The results of this evaluation are used to discuss the overall quality of these tools and to examine the evaluators' assessments of individual tool quality for use in teaching and learning sampling distributions for means. The third paper focused on evaluating the usability of one LO, an applet designed to simulate the construction of confidence intervals for one-sample means and proportions. University students who had completed an introductory statistics course within the past year served as evaluators and completed tasks in a formative usability test to provide information on the strengths and weakness of the applet, barriers to task completion, and suggestions for improving the usability of the applet. Together the manuscripts in this study can impact the way introductory statistics technologies are developed, evaluated, and incorporated into classrooms. In the future, statistics teachers and LO designers and evaluators can incorporate the findings into their own designs, evaluations, and classroom practices. While the use and effectiveness of LOs in statistics education remains a topic in need of additional research, this study provides an initial step toward possible explorations into the uses, evaluation, and design of statistics LOs.

CHAPTER 1. GENERAL INTRODUCTION

National education standards recommend that students begin learning statistics skills as early as third grade, when they are expected to learn how to represent and interpret data, and that they continue in their studies of statistics through their post-secondary education (National Governors Association Center for Best Practices and Council of Chief State School Officers [NGA Center and CCSSO], 2010). Despite statistics topics being intertwined in modern K-16 curricula, statistics education has been described as “a new and emerging discipline when compared to other areas of study and inquiry” (Garfield & Ben-Zvi, 2007, p. 372). While statistics education is still a rapidly growing area of study, the past 15 years have resulted in an expanding research base, with researchers exploring topics such as statistical reasoning and literacy and the incorporation of computers in the statistics classroom.

The Guidelines for Assessment and Instruction in Statistics Education (GAISE), the National Council of Teachers of Mathematics (NCTM), and statistics education researchers suggest use of technology in the teaching of statistics concepts (Aliaga et al., 2010; Garfield & Ben-Zvi, 2008; Hsu, 2003; NGA Center and CCSSO, 2010; Schenker, 2007; Sosa, Berger, Saw, & Mary, 2011). By incorporating technology into their pedagogy, statistics educators can shift the focus from calculations and verbal descriptions of abstract concepts to interpretations of results and visualizations. The GAISE report also suggests statistics educators use available technologies, including applets, classroom response systems, educational software, graphing calculators, spreadsheets, statistical packages, and web-based resources that provide data sources, online texts, and data analysis routines in the statistics classroom (Aliaga et al., 2010). Statistics education researchers have shown the potential of technology to positively affect

student performance (Hsu, 2003; Schenker, 2007), but there is very limited information available on the effectiveness of specific types of technologies.

Human-computer interaction (HCI) is an interdisciplinary field that studies the dynamics of the communications between humans and any tool. HCI researchers often study the design and evaluation of technologies, emphasizing usability and ethical use of technology (Berg, 2000). Most of the existing research in HCI is focused on developing usable technologies (Vrasidas, 2004). Educators and technology developers can apply principles of HCI in any context with interaction between a person and a technology or machine. Rubin (2008) defines the systems approach to interaction design, based on Bailey's human performance model, which defines any activity as being comprised of the activity, the context, and the user. The systems approach to HCI views the activity, the context, and the user as equally important in any human-machine interaction (Peterson, 2007, Rubin, 2008). Figure 1 depicts the interaction between a human and a machine using the systems approach to HCI.

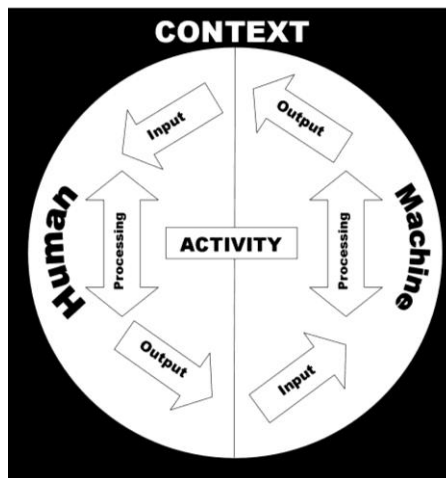


Figure 1. The basic elements of an interaction between a human and a machine. Adapted from “Usability Theory, Practice and Evaluation for Learning Objects” (Peterson, 2007, p. 340).

Educational technology provides an important context in which to study HCI principles (Gerjets & Hesse, 2004; Issroff & Scanlon, 2002; Mishra & Hershey, 2004). In a review of the HCI literature looking for connections between HCI principles and education, Berg (2000) found that “education can learn a great deal from human factors, usability, and interface design approaches to software design” (p. 364). In the past, instructional designers and technology specialists conducted research in separate contexts (Peterson, 2007). By applying the design principles for highly usable instructional materials to research on statistics LOs, I hope to further advance research in both HCI and statistics education.

For this dissertation I chose to focus on one type of educational technology: the learning object (LO). Learning objects (LOs) are defined as any educational technology that an instructor may use to support a specific learning objective (Akpınar, 2008; Baki & Çakıroğlu, 2010; Harvey, 2005; Peterson, 2007; Wiley, 2000; Yacovelli, 2003). Examples of LOs include digital text, pictures, graphics, images, audio or video files (streamed or recorded), live information sources or news feeds, animations, podcasts, applets, wikis, blogs, and forums (Akpınar, 2008; Baki & Çakıroğlu, 2010). Many of the existing technologies freely accessible online to statistics educators and students can be classified as LOs. Existing LOs serve multiple purposes, including data acquisition and storage, calculations, tutorials, simulations, and multi-dimensional analysis. Although LOs have the potential to be efficient, accessible, flexible, and interactive tools (Ben-Zvi, 2000; Peterson, 2007; Wiley, 2000), the absence of research exploring the multi-faceted impact LOs may have on statistics education is a critical gap in the literature.

Statement of the Problem

Researchers in statistics education have found evidence that students who interact with computer-based tools in their learning environments attain greater achievement in statistics

courses than students who learn statistics in a traditional lecture-based course (Hsu, 2003; Schenker, 2007; Sosa et al., 2011). However, Garfield and Ben-Zvi (2008) concluded that statistics education research still needs studies on appropriate uses of technology in statistics classrooms and the most effective ways to implement technology into classrooms. Additionally, the literature calls for studies that evaluate specific technologies for statistics education (such as applets or other statistics LOs), focusing on educational theory and perspectives (Ben-Zvi, 2000; Biehler, 1997; Garfield & Ben-Zvi, 2008). While the literature provides many descriptions of LOs, few studies examine which LOs are the most beneficial or explore the most effective ways to use LOs. Without descriptive studies on the nature of the use of LOs by statistics teachers, researchers have little data on which tools have the most potential to positively impact the classroom. Also, little has been done to evaluate existing statistics LOs. The current statistics education literature shows that there is a need for new studies focusing on how statistics educators evaluate and implement LO technology in their classrooms.

Purpose of the Study

My purpose in producing this interdisciplinary study was to apply foundational HCI principles in a statistics education context in order to provide new information to educators and statistics LO developers about the use and quality of existing LOs. Toward this goal, I completed three studies under the supervision of my dissertation committee and with approval from institutional review boards at both universities associated with the research. In the first study, I surveyed current statistics instructors to gain insight into how educators use statistical LOs in the classes they teach. The next two studies included two different types of evaluation. For the second study, I designed an evaluation where practicing statistics educators used an existing evaluation instrument to assess a group of applets that can be used to illustrate sampling

distributions of sample means. Third, I engaged undergraduate statistics students in a usability test involving an applet designed to simulate the construction of confidence intervals. Separately, each of these studies is unique and provides the foundation for a research trajectory to address the gaps that currently exist in the literature. Together, this work provides an initial look at the intersection of statistics education and HCI. In these studies, I examine the theory behind LO use in a statistics constructivist activity system; the types of LOs statistics educators are using and their methods for implementing them; and the quality of existing LOs designed for use in statistics classrooms.

Dissertation Format

I wrote this dissertation in Iowa State University's approved alternate format which allows writers to produce three journal articles. Each of the three respective articles includes all of the documents I used for data collection as well as the figures, tables, and references I cited. In each article, I described the procedures and outcomes of each study. I used the style required by the respective journals to which I plan to submit the manuscripts; in all three cases, this is the American Statistical Association's Style Guide (American Statistical Association, 2012).

In the first article, "Learning Object Use in Introductory Statistics Classrooms: A Survey of Current Practices, Teachers' Perceptions, and Advice to Peers," I surveyed secondary and post-secondary educators on their use of LOs in their classrooms. The results include descriptions of LOs they used, the ways in which they reported using LOs to promote learning, and the advice they would give their peers on how best to include LOs in statistics instruction.

The second article in this dissertation, "Exploring the Design Quality of Sampling Distribution Learning Objects: An Expert Evaluation," describes an evaluation of existing LOs using data collected to investigate the quality of existing statistics applets that illustrate random

sampling and sampling distributions of sample means. Practicing statistics educators completed an existing evaluation instrument designed by researchers in educational technology to measure the quality of LOs. I used the results to discuss the overall quality of these tools and to examine how the evaluators assessed the quality of individual tools for use in teaching and learning sampling distributions for means.

In the third article, “A Usability Evaluation of a Confidence Interval Applet Using Student Evaluators,” I narrowed the focus of evaluation to the usability of one LO, an applet designed to simulate the construction of one-sample confidence intervals. University students who had completed an introductory statistics course within the past year served as evaluators and completed tasks in a formative usability test to provide information on the strengths and weaknesses of the applet, barriers to task completion, and suggestions for improving the usability of the applet.

I intended this dissertation to relate information from existing learning theories to applications in statistics education and HCI, to provide results that fill gaps in the existing research on statistics LOs, and to inform practice for statistics educators and developers of statistics LOs. This study may provide selection criteria for statistics educators who are interested in choosing the best LOs to use in their classes. Developers of educational technologies will find descriptions of two specific evaluations, an expert-based evaluation using statistics instructors and a formative usability test using statistics students as evaluators. This study’s evaluations can provide statistical LO developers with a set of characteristics to either emphasize or avoid when they develop new or refine existing LOs in the future. This study’s strengths are in its ability to provide statistics educators, educational technology designers, and

LO evaluators with empirical information about how educators are using LOs and the quality of current LOs.

References

- Akpınar, Y. (2008). Validation of a learning object review instrument: Relationship between ratings of learning outcomes and actual learning outcomes. *Interdisciplinary Journal of E-Learning and Learning Objects*, 4, 291-302.
- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., & Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) college report*.
- American Sociological Association. (2007). *American Sociological Association Style Guide*. Ingram.
- Baki, A., & Çakıroğlu, Ü. (2010). Learning objects in high school mathematics classrooms: Implementation and evaluation. *Computers & Education*, 55(4), 1459-1469.
- Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, 65, 167 – 189.
- Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning*, 2(1-2), 127-155.
- Berg, G. A. (2000). Human-computer interaction (HCI) in educational environments: Implications of understanding computers as media. *Journal of Educational Multimedia and Hypermedia*, 9(4), 347-368.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396.
- Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York, NY: Springer.
- Gerjets, P. H., & Hesse, F. W. (2004). When are powerful learning environments effective? The role of learner activities and of students' conceptions of educational technology. *International Journal of Educational Research*, 41(6), 445-465.
- Harvey, B. (2005). Learning objects and instructional design. *The International Review of Research in Open and Distance Learning*, 6(2).
- Hsu, Y. (2003). "The effectiveness of computer-assisted instruction in statistics education: A meta-analysis." PhD dissertation, University of Arizona, Arizona.

- Issroff, K., & Scanlon, E. (2002). Educational technology: The influence of theory. *Journal of Interactive Media in Education*, 2002(6).
- Mishra, P., & Hershey, K. A. (2004). Etiquette and the design of educational technology. *Communications of the ACM*, 47(4), 45-49.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards-Mathematics*. Washington, DC: Author.
- Peterson, D. (2007). Usability theory, practice and evaluation for learning objects. In K. Harman & A. Koochang (Eds.), *Learning Objects: Applications, Implications, & Future Directions* (pp. 337-370). Santa Rosa, CA. Informing Science Press.
- Rubin, J. & Chisnell, D (2008). *Handbook of usability testing: how to plan, design, and conduct effective tests*. John Wiley & Sons,
- Schenker, J. D. (2007). "The Effectiveness of Technology Use in Statistics Instruction in Higher Education: A Meta-Analysis Using Hierarchical Linear Modeling." PhD dissertation, Kent State University College, Ohio.
- 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.
- Vrasidas, C. (2004). Issues of pedagogy and design in e-learning systems. In *Proceedings of the 2004 ACM symposium on applied computing* (pp. 911-915).
- Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In *The instructional use of learning objects: Online version*.
- Yacovelli, S. R. (2003). Understanding learning objects: {The} basic "-chunks." *College & University Media Review*, 10(1), 17-26.

CHAPTER 2: LEARNING OBJECT USE IN INTRODUCTORY STATISTICS CLASSROOMS: A SURVEY OF CURRENT PRACTICES, TEACHERS' PERCEPTIONS, AND ADVICE TO PEERS.

A paper to be submitted to the *Journal of Statistics Education*

Rachel Graham, Iowa State University

Key Words: statistics, education, learning objects, survey, educational technology

Abstract

The purpose of this study was to provide statistics educators with details of learning object use in today's introductory statistics classrooms. A sample of high school and post secondary educators, involved in AP Statistics exam scoring, completed a survey on which learning objects they use, how they use them in their classrooms, and advice they would give to colleagues interested in beginning to use LOs. These unique survey results corroborate information about statistics LOs that can be found in the literature and provide introductory knowledge regarding how these teachers use LOs in the context of introductory statistics.

1. Introduction

Statistics and mathematics standards for preschool grades through postsecondary education include recommendations that teachers should incorporate technology into their pedagogy (Aliaga et al. 2010; Franklin et al. 2007; National Council of Teachers of Mathematics [NCTM], 2000). The *Common Core State Standards for Mathematics* (NGACBP/CCSO 2010), the *Principles and Standards for School Mathematics* (NCTM 2000), and the *Guidelines for Student Assessment and Instruction in Statistics Education (GAISE) Report: Pre K–12 Report* (Franklin et al., 2007) present standards in K–12 mathematics and statistics education that recommend technology as an appropriate tool for the mathematics classroom. The *GAISE College Report* (Aliaga et al. 2010) further recommends using technology to analyze data and to develop understanding of concepts in postsecondary introductory statistics instruction. In addition to recommendations for use, the standards and existing statistics education research also include examples of many different types of technologies designed to help teach statistics topics. However, beyond self reporting by researchers, very little has been done to investigate which of these technologies statistics teachers are using and how they are implementing the technologies in their teaching methods.

Educational technology researchers have identified the learning object (LO) as a classification of educational technology including any reusable digital resource that educators can use to support a learning objective (Apkinar 2008; Baki and Cakiroglu 2010; Butson 2003; Churchill 2007; Harvey 2005; Peterson 2007; Wiley 2000; Yacovelli 2003). Peterson (2007) recommends researchers study LOs within the specific discipline in which educators employ them. Although, the LO is not a popular classification, statistics educators have many technologies available to

them that are learning objects (LOs). The Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), hosts the website CAUSEweb.org, which includes the largest statistics LOR with over 2,000 resources devoted specifically to statistics education. Examples of existing statistics LOs include animations, applets, public blogs, case studies, data sets or repositories, discussion forums, electronic textbooks/material, tutorials, videos, and wikis. The most prevalent type of statistics LO is the applet. Statistics technology developers have created many different applets to help teach most topics covered in introductory statistics courses (i.e. descriptive statistics, basic probability, discrete and continuous probability distributions, and one- and two-sample inference for population means and proportions). The STATistics Applets for Teaching Topics (STAT-ATTIC) website located at <http://sapphire.indstate.edu/~stat-attic/index.php>, provides links to statistics applets exclusively.

If statistics education researchers are to complete relevant empirical studies that provide implications for any significant population of statistics students the existing literature needs more information about effective practices teachers are employing in current statistics classrooms. With the intention of filling a gap in the existing research on current teachers' use of statistics LOs, I surveyed postsecondary introductory statistics teachers and high school AP Statistics teachers who are involved in standardized statistics assessment (i.e. scoring AP Statistics exams) and who have previously used one or more LOs as learning tools in their classrooms. I designed the survey to gain insight into what types of LOs current statistics teachers use, how they use LOs as tools for learning, and their perceived benefits and limitations of LO use. For this article, I first provide a theoretical framework to describe my assumptions about learning in this study. Next, I review existing literature to explore the current research regarding educators' use of LOs in introductory statistics classrooms. I then describe the study methodology as defined by the theoretical framework and literature review, and conclude by providing results of the survey and summarizing the implications of the results for statistics educators and researchers.

2. Theoretical Framework

The intent of this theoretical framework is to outline how students learn including, first, a description of activity theory and constructivism and then an overview of the connection between these ideas. Further, I include a discussion of why LOs are an attractive tool to study inside this theoretical framework.

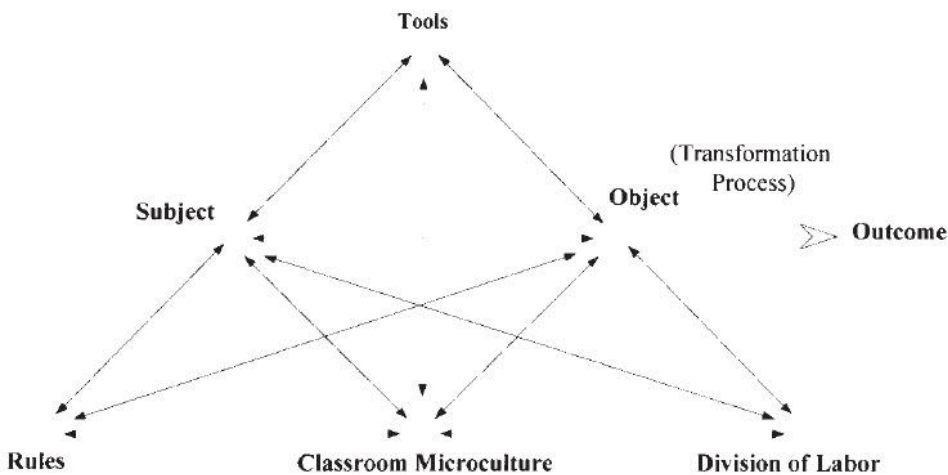
2.1 Activity Theory

Leont'ev (1978) first described activity theory using a dynamic system to illustrate how participants reach a desired outcome through activities that involve subjects interacting with objects and operations that change over time. Activity theory focuses on the interaction between components and both individual and social human actions in a contextualized activity system (Barab et al. 2002; Engeström 1987). The purpose of any activity in a system is to help a subject reach an objective and any outcome remains dependent upon the context in which the activity occurs (Jonassen and Rohrer-Murphy 1999; Kaptelini and Nardi 2012).

The major components of any activity system in a classroom context (shown in Figure 1) are as follows: the subject, object, tools, rules, classroom microculture, and systems educators use to

divide labor (Barab et al. 2002; Engeström 1987; Jonassen and Rohrer-Murphy 1999; Kozulin 1986; Leont'ev 1978). The main activity in a typical classroom is the teaching and learning of a specific concept. The subjects in the activity system are the students and the object is for students to construct deeper understanding of the concept(s). A statistics education activity system's tools are any lecture materials, textbooks, statistical software or LOs, and any other resources instructors use to promote learning; the rules are the social norms in the classroom; the classroom microculture consists of all of the participants involved in the classroom community, including decision makers who might not directly interact with the students during the activity (e.g., department coordinators, administration); the division of labor includes the specific actions of the teachers and students directly involved in the activity; and the outcome is the actual knowledge that the students construct (Barab et al. 2002). Together, the components comprise a complex interactive system in which an activity occurs to help students achieve an established learning outcome.

Figure 1 Modification of Leont'ev's model of an activity system in a classroom context (Barab et al. 2002, p. 79).



From the activity theory perspective, researchers who perform any examination of classroom learning should consider the interactions among all of the components in the activity system (Allen et al. 2011; Kaptelinin and Nardi 2012). While activity theory describes each component of an activity system and includes the principle that these components are all interdependent, it does not describe *how* students construct knowledge. An epistemology, such as constructivism, is needed to provide the explanation of how knowledge is constructed that activity theory is missing.

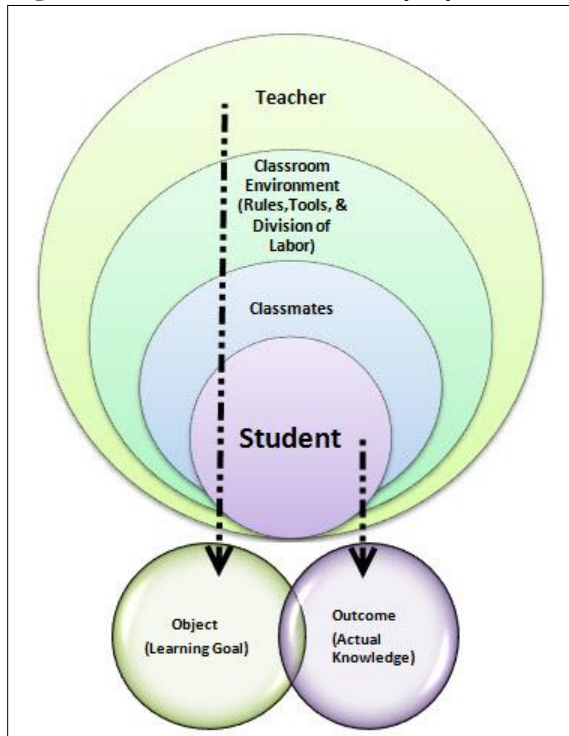
2.2 Constructivism

Many theorists widely accept constructivism as the epistemology for developing knowledge; it explains learning as a complex process where individuals uniquely reconcile new information with past experiences (Boudourides 2003; Cobb 1994; Jonassen et al. 2007; Karagiorgi and Symeou 2005; Lunenberg 1998; Tishkovskaya and Lancaster 2012; Vygotsky 1962; Wilson

1995). Constructivists believe that knowledge is not dependent on the person learning but intertwined in an individual's perceptions, experiences, imaginations and social constructions (Boudourides 2003, Karagiorgi and Symeou 2005; Tishkovskaya and Lancaster 2012). Johnson (2010) explains that humans perceive what they expect to perceive, but these perceptions are biased by past experiences, the present (including the context of interactions), and future goals. When students from the same class individually construct different understandings of the same lesson, their learning supports the individuality of constructivist principles. Even though learning happens on an individual level, many constructivists agree that the student's interactions in a social or collaborative setting can influence a student's individual concept development (Bransford et al. 1999; Jonassen et al. 1995; Morrison 2003; Vygotsky 1981). The social characteristics of constructivist principles reinforce the aspects of activity theory which indicate that all participants involved in the same activity have the potential to affect student learning. Constructivist learning environments are student-centered and in these environments teachers facilitate learning rather than act as the object through which learning happens. Active learning, authentic learning, and collaborative tasks are all indicators of constructivist classroom activities because they are student-focused (Karagiorgi and Symeou 2005).

2.3 Constructivist Activity Systems

Constructivism serves as a lens through which educators can create effective learning activities. Educators can plan classroom activities using constructivism to outline how each individual subject (student) can reach the object (learning goal) while he or she interacts with the teacher and the other students (classroom microculture). Additionally, constructivism can explain why educators choose or design the tools, division of labor, and rules for a particular activity system. Figure 2 shows a model of a constructivist activity system I have organized around an individual student and the knowledge they construct as a result of a particular activity. Teachers in a constructivist activity system design learning experiences that are student focused and with a goal of helping students understand the concept being taught. Students in this system are completing the activity in a classroom with their classmates, under set classroom norms (rules), using tools chosen by the teacher.

Figure 2 Constructivist Activity System

The goal for any constructivist activity system is to have the students understand the learning goal at the completion of an activity. When a student's actual understanding (outcome) overlaps with the learning goal (object) enough to indicate proficiency, this may be evidence of an effective or successful activity. Conversely, if there is not an appropriate amount of overlap between outcome and object, this may indicate an ineffective activity. Measuring the effectiveness of an activity should involve careful assessment how much overlap there is between the object and the outcome of the activity. Ideally, teachers would select tools and tasks that create student-centered learning and align them with their assessment of the outcome. This study focused on LOs because they are a class of tools that researchers describe as having constructivist characteristics which aligns them with this theoretical framework.

2.4 Learning Objects in the Theoretical Framework

Learning objects are one class of educational technology that may bear constructivist attributes. According to Peterson (2007) and Wiley (2000), constructivist tools and LOs have in common: multiple entry points, components with which individual students can interact, and reusability in many different contexts. Wiley (2000) found that educators can reuse and adapt LOs, allowing them to be used by many different students in multiple educational settings. If learning objects contain features educators can use to characterize them as constructivist tools they should ideally show great potential for success in helping students learn statistics topics, making them an attractive technology to study under a constructivist framework. However, it is my opinion that even if an LO has constructivist features it should not be classified as a standalone constructivist

tool. Therefore, I believe it is the activity and not the tool that qualifies a classroom as a constructivist environment.

3. Literature Review

I designed the present study to explore the use of LOs in introductory statistics classes as well as teachers' perceptions of the benefits and limitations of existing LOs. I reviewed the statistics education literature to provide a description of the statistics LOs currently available, the nature of their use in classrooms, and benefits and limitations of LO use. In this review I draw on the literature to describe the types of LOs used in statistics education, the current classroom uses of LOs, and their benefits and limitations.

3.1 Types of Learning Objects Currently Being Used

Although the statistics education literature contains examples of multiple types of LOs freely available online, applets are by far the most common type mentioned. The existing literature is comprised largely of many descriptions of the applets that educators use to teach introductory statistics topics (Briggs and Sheu 1998; Chance and Rossman 2006; Chu et al., 2009; Darius et al. 2000; Darius, et al. 2002; DePaolo 2010; Dinov and Christou 2009; Dinov et al. 2008; Dinov and Sanchez 2006; Lane 1999; Lane and Scott 2000; Mallory and Jensen 2001; Mittag 2002; Mulekar 2000; Phillips, 2003; Saporta 1999; Schenker 2007; Schneider 2008; Schwarz 2007; Tishkovskaya and Lancaster 2012; West and Ogden 1998). Other types of statistics LOs include data libraries (Chance, et al. 2007; Neumann et al. 2010; Saporta 1999), tutorials (Aberson et al. 2003), videos (McDaniel and Green 2012; Schenker 2007), and Wikis (Ben-Zvi 2007). Online statistics tutorials differ from textbooks since they usually present material on individual topics rather than full courses of material (Schenker 2007). Additionally, the research suggested statistics videos available online, including supplementary instruction, statistical reports, and examples of student projects. However, little exists on how statistics educators use animations, blogs or discussion forums.

3.2 Methods for Using Learning Objects

The statistics education literature includes examples of both teacher-centered and student-centered methods to implement LO in statistics classrooms. The teacher-centered methods that I found in the literature were to use LOs to refresh prerequisite material (Anderson-Cook et al. 2003), to introduce new topics (Schneider 2008), and to provide in-class demonstrations of topics such as empirical rule, correlation, sampling distributions for sample statistics, and one- and two-sample inference including confidence intervals, hypothesis testing, p values, and power (Burrill 2010; Mills 2002; Schneider 2008; West and Ogden 1998).

The most common constructivist, or student-centered, method I found in the literature was to use LOs that include simulations that illustrate abstract topics such as, repeated random sampling, confidence intervals, and p -values (Blejec 2003; Bertie and Farrington 2003; Chance and Rossman 2006; delMas et al. 2000; Mills 2002; Schneider 2008). Chance and Rossman (2006) recommend that educators use simulation technologies along with an activity carefully designed by teachers to illustrate abstract concepts that may be difficult for students to understand solely

through lecture based instruction. The second most common method was to use LOs as part of structured activities including guided tutorials (Aberson et al. 2003) and detailed explorations (Chance and Rossman 2006; delMas et al. 1999; McDaniel and Green 2012). Another method was having students compare results from an LO to those of a previously completed hands-on activity. For example, Bertie and Farrington (2003) suggested using an applet to illustrate confidence intervals after students have computed them by hand. Additionally, delMas, Garfield, and Chance (2000) and Chance and Rossman (2006) described in detail how to use a specific LO to teach the concept of a sampling distribution after students completed an in-class, hands-on, sampling activity.

Anderson-Cook and Dorai-Raj (2003) and West and Ogden (1998) described using applets to explain concepts with the purpose of following-up a topic that their students had been exposed to previously. Learning objects can also be used for independent student exploration (Anderson-Cook and Dorai-Raj 2003; West and Ogden 1998) in order to engage students in active learning. The other two methods I found in the literature, collaboration between students (delMas, Garfield and Chance 1999) and group discussion (Schneider, 2008), encompass the social aspects of constructivist activity systems.

3.3 Benefits of Using Learning Objects to Teach Statistics Concepts

The benefits of using LOs in the statistics classroom, as the literature describes, are primarily benefits perceived by researchers based on self-reported observations and literature review rather than on empirical research. The most commonly cited benefit of LOs is the ability to use simulation to illustrate abstract topics (Anderson-Cook and Dorai-Raj 2003; Burrill 2010; Chance and Rossman 2006; Christou et al. 2007; Darius et al. 2002; Hsu 2003; Ng and Wong 1999; Schenker 2007; Schneider 2008; Sturm-Beiss 2005; West and Ogden 1998). Chance and Rossman (2006) noted, when students actively engage with an educational technology, such as an LO, they can possibly create deeper understanding than if they had only watched a demonstration. Students can only have an active engagement with a tool if they are given control of the interaction, allowing them to take ownership of their learning. Garfield, Chance, and Snell (2000) indicate that students who use LOs may benefit from learning through interaction, which can help with active knowledge construction. The perceived benefits found in the literature are overwhelmingly consistent with constructivist activity systems.

I found only two studies indicating benefits for students when instructors used LOs in introductory statistics classrooms. Anderson-Cook and Dorai-Raj (2003) found that students reacted positively to applets in class; the authors observed improved student performance on test items after they implemented the use of LOs into their curriculum. Aberson, Berger, Healy, and Romero (2002) compared two samples: one where students learned hypothesis testing through online tutorials and one where they learned without the tutorials. Using an ANCOVA analysis, Aberson et al. (2002) compared scores from one essay question on a final exam across two groups, using their final course points as a covariate; they found that students using tutorials scored significantly higher than those who did not. Although it is limited, empirical evidence found in the statistics education literature does suggest that LOs can be effective tools for promoting student learning of statistics.

3.4 Limitations of Learning Objects

Limitations of LOs are evident when an LO's features interfere with an educator's ability to effectively communicate concepts to students. Designers should construct reusable LOs by minimizing references to specific contexts; however, removing too much context from learning experiences directly conflicts with the learning theories that show context is important if students are to make connections between tasks they complete in class and activities outside of the classroom (Jonassen 1991; Jonassen and Rohrer-Murphy 1999; Vygotsky 1981). In 2004, Wiley et al. wrote, "if decontextualized learning objects are to be developed and deployed, a method of reintroducing context must be utilized" (p. 509). Teachers need to utilize contextual examples when they are using LOs to compensate for the lack of context imbedded in LOs.

Many statistics educators use LOs to perform calculations and provide simulations (Mills 2002; Ng and Wong 1999; Schenker 2007). However, students using an LO for such purposes without context and further interaction may struggle to build conceptual knowledge, again implying that the constructivist nature of LOs is dependent on the manner in which they are used (Aliaga et al. 2010; Pearl et al. 2012). After students have interacted with an LO, statistics teachers drawing from constructivist theory should be sure to connect the LOs students are using back to the topic they are trying to learn (Schwarz 2007; West and Ogden 1998). If teachers use LOs without carefully considering student interaction with these tools, educators can risk LOs being a barrier to students' statistical literacy.

3.5 Implications for Research

Standards for statistics education and the statistics education literature outline expectations that educators will use technology to help students learn and provide examples of the types of LOs available, links to access them, and anecdotal evidence of the benefits and limitations of LO use. The statistics education literature holds fewer descriptions of *how* educators are using LOs in their classrooms, and there are a limited number of studies showing any evidence of the effectiveness of LOs as tools to help teach statistics. Beyond what researchers have described about their own use of LOs, the literature contains very little evidence about whether the existing descriptions and suggestions accurately reflect what is happening in introductory statistics classrooms. Previous research suggests that statistics education will benefit if educators and researchers know which LOs are being used by teachers, how and why teachers are using them as tools, and what advice experienced users of LOs would offer colleagues interested in implementing them into their own instruction (Hsu 2003; Mills 2002; Schenker 2007; Surry and Land 2000). By describing the LO practices of a sample of both high school and postsecondary introductory statistics teachers this study will broaden the existing knowledge about typical classroom use of LOs and the motivations behind that use.

4. Method

The basis for this study was a survey I conducted to gather opinions and information on classroom practice from educators involved in the process of scoring AP Statistics exams and who have used online learning objects in their classrooms. In this study, I collected and analyzed

data to describe a sample of current statistics teachers' use of LOs and their perceived benefits and limitations of LOs as teaching resources. The research questions were as follows:

1. What are the most popular types of LOs educators are using in introductory statistics classrooms?
2. How are introductory statistics teachers using LOs in their classrooms to support or promote learning?
3. Why do teachers use LOs in introductory statistics classrooms?
4. What advice about using LOs would teachers give to colleagues?

4.1 Participants

Each year, Educational Testing Service (ETS) employs hundreds of introductory statistics teachers, from both postsecondary and high school levels, to read and score the free response questions from the Advanced Placement (AP) Statistics exam for that year. An email list for this group of statistics educators is compiled and distributed by a designated reader to everyone on the list for them to use in educational or personal mass communication. As an AP Statistics Reader for ETS, I have open access to this list. I invited the AP Readers who place themselves on the email list to participate in this research by completing an online survey.

4.2 Instrument

In order to collect data to help investigate the research questions, I created an online questionnaire using Qualtrics, a secure survey creation and management software available to human-computer interaction students at Iowa State University. The survey consisted of ten items designed to collect information to answer the research questions for this study (see appendix). The first five items asked for demographic information, including the level at which the teachers primarily teach statistics (e.g. high school or postsecondary), how many years they have been teaching in general, how many years they have been teaching statistics, and their average class size. The next two items in the questionnaire collected data on the types of LOs teachers had used previously in their introductory statistics classes. The first "type of LO" item asked respondents to select the types of LOs they had used from the list of LO types obtained in the literature review. Respondents had the option of adding additional types of LOs using an "other" field. Next, respondents gave names or urls for specific LOs that they had used. The next item on the survey gathered information on the methods teachers employ when implementing LOs into their classrooms. The "method of LO use" item was a multiple response question with a checklist populated from the literature review. Respondents identified how they used LOs by selecting all that applied to them from a list. An open-ended text-entry question asking why teachers use LOs was the next item on the survey with the purpose of collecting respondents' perceived benefits. The last item asked teachers to describe advice they would give to colleagues who were interested in using LOs in their classrooms in an open-ended question.

4.3 Data Collection

To invite participants to take my survey I sent an email to the list of instructors involved AP Statistics scoring, giving them information about the project, explaining all relevant consent

information, and asking them to participate only if they had used LOs in teaching introductory statistics. To increase response rates, I sent the invitation email in the middle of the week during the afternoon, at a time when emails were not automatically sent (spam) (Boyer and Stron 2012). I sent a reminder email two days later, again in the afternoon. The survey remained open for three weeks. When invited participants responded to the invitation email they were then sent a link to the survey through the secure university email system. The Qualtrics automated data system collected the responses.

4.4 Data Analysis

After the survey closed, I downloaded responses from Qualtrics into one large data set without any identifying characteristics for the respondents. I ran a report in Qualtrics that summarized information collected from the demographic items and the two multiple selection items which collected information on the type of LOs respondents used and the implementation methods they used. I compiled the “other” responses to the multiple selection items into one list and eliminated any that were already represented in the provided selection list retaining responses that were distinct from the options I included in the survey items.

With the data from the free-response questions, which all concerned why respondents used LOs and any advice they would give to colleagues interested in using LOs, I made a list of the distinct responses and tallied frequencies for any similar responses. While respondents provided many reasons to use LOs related to benefits for the teachers or students, most restated methods listed in the previous question, such as “introduction to topics”, “independent exploration”, “review and practice”, and “labs and group activities”. I eliminated the responses that were the same as the methods from prior survey items, and then, I coded the unique responses with either an S, to indicate a constructivist or student-centered approach, or a T to indicate to indicate a teacher-centered approach. With instructor advice data, I sorted similar responses and then coded them as either student-centered or teacher-centered. I also computed the percentages of respondents that indicated each of the unique responses.

5. Results

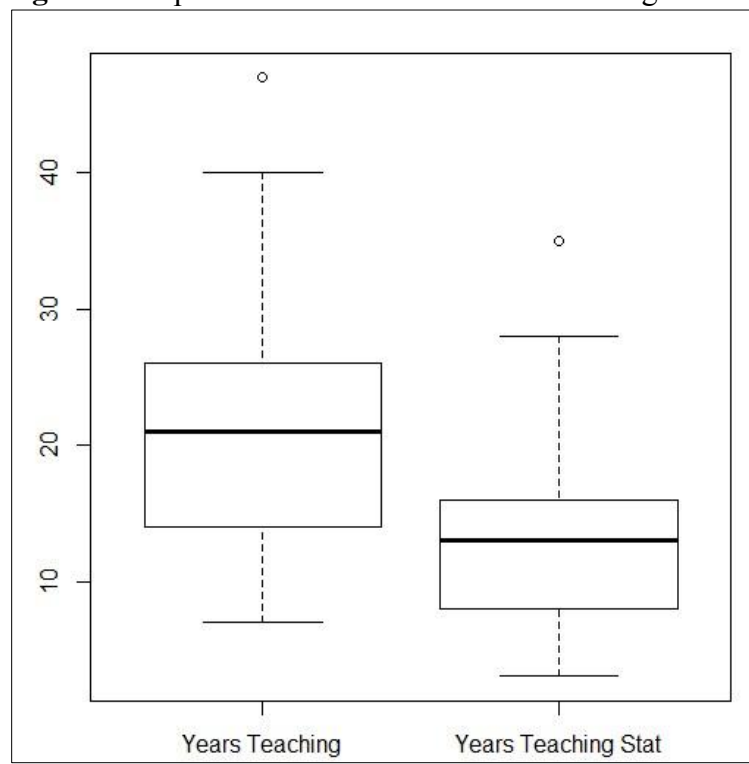
The results of this survey show how LOs are being used in a small group of introductory statistics classrooms. The demographic information for the participants and the results of the survey questions as they connect to research questions are summarized below.

5.1 Participant Demographics

Of the 654 statistics educators on the list of AP Readers I invited to participate, 71 responded to the invitation and 68 completed the survey. While 10.4% of the invited instructors responded, this does not represent an accurate response rate because the invitation to participate in the survey requested responses from only teachers who had experience using at least one LO in their classrooms. Since the number of invited instructors who have used LOs is uncertain, I could not calculate an accurate response rate.

Sixty percent of the 68 respondents taught introductory statistics at the high school level while post-secondary level instructors accounted for the other 40 percent. The mean class size of the respondents' self-reported averages was 28.79 students. There were respondents from 28 different states in the United States and from one country in South America. In Figure 3, side-by-side box plots show the distributions of the number of years teaching in general and the number of years teaching statistics; the distributions for both are positively skewed and each has one outlier. Respondents' years of teaching experience ranged from 7 to 47, with a median of 21 years and an inter-quartile range of 12 years. The maximum number of years any respondent had taught statistics was 35 years, and the minimum was 3 years. The number of years teaching statistics had a median of 13 years and an inter-quartile range of 8 years.

Figure 3 Respondents' Number of Years Teaching and Teaching Statistics.



The side by side box plots show that both the distribution for the number of years teaching and the distribution of the number of years teaching statistics are right skewed. The center of both distributions is greater than ten years, with minimums no less than three years, indicating that the respondents were experienced statistics teachers.

5.2 Types of LOs being used in introductory statistics classrooms

This study's first research question investigated the types of LOs teachers use in their introductory statistics classrooms. Table 1 summarizes the responses to the first survey question: the type of LO, the number of respondents who indicated they have used each type, and the

percentage of respondents. Responses to the survey show that applets, data sets and data repositories, videos, and electronic textbooks and materials each were identified by more than 50 percent of the respondents as being used in their classrooms. Between 25 and 50 percent of the respondents selected three types of LOs: case studies, animations, and discussion boards or forums. Less than 20 percent of the respondents selected each of blogs, wikis, and tutorials. Four teachers indicated they had used online news sources in their “other” responses.

Table 1 Types of Learning Objects Being Used by Survey Respondents

Learning Object Type	# of Respondents (<i>n</i> = 68)	% of Respondents
Applets	57	84%
Data sets/repositories	52	76%
Videos	50	74%
Electronic textbooks/materials	39	57%
Case studies	29	43%
Animations	28	41%
Discussion boards/forums	18	26%
Tutorials	13	19%
Online news sources	4	6%
Wikis	4	6%
Blogs	2	3%

Respondents also provided specific examples of LOs they had used in their classrooms. Responses included references to specific LOs or collections of LOs, URLs, and general comments about the types of LOs used. Respondents gave specific examples of applets, data sets/repositories, electronic textbooks, tutorials, and videos; they also listed case studies and newspaper articles, but no specific information on where they were accessed from. Respondents did not provide any specific examples of blogs, discussion boards, and wikis; however, they did list Moodle and Blackboard, which contain these features but are not freely accessible LOs or LORs. Any further analysis of the specific LOs given by respondents was beyond the scope of this study.

5.3 How LOs are being used to support or promote learning

The next research question investigated *how* statistics educators were using LOs in their classes to promote learning. The responses to the corresponding survey question, summarized in Table 2, include the frequency and percentage of respondents who indicated they used each method. Items denoted with an asterisk indicate methods that are consistent with the constructivist or student-centered characteristics identified in the literature. For the teachers who responded to this survey, in-class demonstration (87 percent), simulation (81 percent), and topic introduction (78 percent) were the most popular methods for using LOs. Two-thirds of the respondents indicated they had used LOs as a follow-up to a previously introduced topic. The next most popular method respondents selected, with approximately half selecting each choice, was to use an LO to have students generate results from an LO and compare these to results from a hands-on activity for the same topic, independent exploration, and collaboration between students. Structured activities, group discussion, and a refresher for previously learned material were respondents' least favored choices, but over a third of respondents even selected these less popular options. The "other" options on this question did not yield any additions to the list.

Table 2 Methods of LO Use Survey Respondents are Employing in Introductory Statistics Classrooms

Method of LO Use (* indicates potential student-centered methods)	# of Respondents	% of Respondents (<i>n</i> = 68)
In-class demonstration	59	87%
Simulation*	55	81%
Introduction to new topic	53	78%
Follow-up to a previously introduced topic	45	66%
Comparison to hands-on activity*	40	59%
Independent exploration*	39	57%
Student collaboration*	33	49%
Structured activities*	30	44%
Group discussion*	26	38%
Refresher for previous material/prerequisites	26	38%

Interestingly, only one of the respondents' four most popular methods can be student-centered. Because simulations can be used to illustrate an abstract concept, they can be used in a student-centered manner if students are directly interacting with the simulation. However, it is unclear when respondents selected this term if they had this use of simulations in mind, or a more passive approach in which students merely observe the instructor demonstrating the simulation.

More than half of the respondents indicated they had used LO methods that are student-driven including, comparison of LO output to results of a hands-on activity and for independent student exploration. Less than 50% of the respondents indicated they used LOs for the other student-centered methods on the survey: student collaboration, structured activities, and group discussion.

5.4 Why teachers are using LOs

To collect data to investigate the third research question, I asked respondents to list the reasons why they use LOs. Most of the respondents provided multiple reasons they had used LOs. The respondents' contributions, summarized in Table 3, were classified as either student-centered or teacher-centered benefits.

Table 3 Survey Respondents' Perceived Benefits of Using LOs

Focus	Perceived Benefit of LO Use	# of Respondents (<i>n</i> = 67)	% of Respondents
Student-Centered	Illustration and /or visualization of topics	32	48%
	Deepen conceptual knowledge	28	42%
	Student engagement, enjoyment, or motivation	26	39%
	Authentic experiences	13	19%
	Student control	7	10%
	Accessibility	5	7%
	Reference points for later in the course	3	4%
Teacher-Centered	Convenient Time Saver	8	12%
	Assess student comprehension	1	1%

The majority of the respondents' reasons given for using LOs are student-centered benefits. Forty-eight percent of the respondents noted illustration and/or visualization of topics as a reason they use LOs, which made it the most popular reason for this group of teachers. Many teachers noted that LOs allow them to show things through visualizations that they cannot relay as effectively in lecture. The second most popular reason, given by 42% of respondents, was that LOs help deepen conceptual understanding. Twenty-six of the survey respondents (39%) gave reasons that related to student engagement, enjoyment, and motivation. Nineteen percent of the respondents stated that LOs allowed them to expose students to authentic experiences through the use of real data. Three teachers indicated that they had used LOs as a point of reference to help students remember a concept they had covered previously in class. The least popular responses, given by at most 10% of the respondents, were student control and accessibility.

Twelve percent of the respondents noted that LOs were a convenient time saver which was the most frequently given reason for teachers to use LOs for their own benefit. Only one teacher indicated that LOs helped him/her assess student comprehension of the material.

5.5 Advice teachers would give to colleagues

To address the final research question, sixty-six respondents shared advice they would give to colleagues who came to them interested in knowing more about implementing LOs in their own classrooms. The majority of respondents' comments represented general teaching practices or common sense and were not directly focused on student learning. The most common piece of advice respondents gave was to try LOs before they are used in the classroom. Multiple teachers noted that educators should repeat this trial directly before instruction to make sure the LO would work correctly during use with the class. Respondents also made recommendations to have a back-up plan, to schedule enough time for LO interactions, and to "shop around" before teachers decide which LOs they will use.

The teachers who gave student-centered represented no more than 12 percent, or 8 of the 66 respondents. Ideas regarding selection criteria for LOs included multiple suggestions that LO selection should occur with student needs and specific learning objectives in mind. Additionally, respondents advised teachers draw data sets from topics that students find interesting and relevant. Respondents also gave recommendations that teachers should reflect on their practice and listen to student feedback. Other advice suggested that educators should use LOs both in independent and guided exploration activities. Only very small percentage of respondents (3%) advised that their colleagues give students clear and precise directions for interacting with LOs or use LOs as reference points, at a time in the course after a topic has previously been covered, to connect to previously learned material.

There were no responses that included negative comments about implementing LOs into the classroom, but three teachers did include words of caution. One teacher indicated that it was difficult to "sift through" all of the options when choosing an LO. A second teacher suggested that using LOs "worked better with more motivated students." Another stated that the quality of LOs varies but then went on to say that "in many cases learning objects allow me to emphasize concepts over computations," implying the quality of an LO may influence teachers' choices when selecting an LO.

6. Discussion

I designed this study to describe how teachers use LOs in secondary and postsecondary introductory statistics classrooms and their perceptions about the use of LOs as educational tools. For this study, activity theory explained why the information collected in this study is important; constructivism provided the reason for choosing LOs as the focal technology. We can see evidence of activity theory and constructivist principles in the literature and through the survey data. This discussion includes a summary of the information that I obtained to answer the research questions that shaped this study, including trends evident in the responses, and connections between results and the findings in the literature. I have concluded the discussion with an explanation of the limitations of this study and recommendations for future research.

6.1 What are the most frequently used types of LOs in introductory statistics classrooms?

This study's results indicated respondents most frequently use applets, data sets and repositories, and videos. Applets, the type of LO identified by the highest percentage of respondents, often contain interactive features allowing them to be reused in multiple different contexts. For example, each student in a class could collect his or her own data and use an applet to create a histogram and to investigate the features of the distribution displayed. While they can be reused in many different contexts, data sets and videos do not involve the interactive features designers usually use in applet design, which may be the reason they are less popular. This study confirmed that the educators who completed the survey apply all types of LOs found in the statistics education literature. The literature had provided descriptions of three of the most popular types of LOs as indicated by the respondents (applets, data sets and electronic textbook materials). The participants also reported using animations, blogs, and case studies, which I did not find in the literature, implying that the literature does not provide an exhaustive list of the types of LOs teachers are using. In addition, responses to the questionnaire showed that statistics educators may want to add online news sources to any list of the types of LOs teachers are currently using.

Despite being scarcely mentioned in the literature, 70 percent of the respondents indicated they have used videos in their statistics classrooms. If the popularity of videos these survey results represent is accurate, future research that focuses on the use of videos may provide better insight into the impact this type of LO can have on students learning statistics. I recommend that any future investigations into the types of LOs statistic teachers are using involve data collection beyond what has been published to ensure accurate information is collected.

6.2 How are introductory statistics teachers using LOs in their classrooms to support or promote learning?

Each of the methods for using LOs that I found in the literature had no less than 38 percent of the respondents. Additionally, the respondents did not indicate any methods for using LOs in addition to what was found in the literature review implying that existing research may provide a good foundation for exploratory studies.

The teachers' responses indicate a wide range of methods for LO use: simulations of abstract concepts, independent and guided explorations, connections back to previously covered material, and collaborative or group work. These results provide evidence of constructivist activities being used as well as information about the division of labor in these teachers' classrooms. However, it is important to note that the majority of the most popular responses were not student-centered or constructivist in nature. Therefore, despite some evidence of constructivist methods, respondents seem to be using predominantly traditional lecture-based methods of instruction. If educators want to design activities that align with a constructivist activity theory framework, they need to select methods that are student-centered whenever possible (delMas et al. 1999; Garfield and Chance, 2000).

6.3 Why do teachers use LOs in introductory statistics classrooms?

When describing their reasons for using LOs, the respondents noted benefits to both students and teachers. The majority of the benefits respondents gave were student-centered. The few number of teacher-centered reasons respondents gave for using LOs may suggest that these respondents are employing teaching methods that are predominantly student-focused which is an attribute consistent with constructivist pedagogies (Karagiorgi and Symeou 2005).

The respondents' perceived benefits for students, including providing authentic experiences and giving students control of their own learning, directly link to constructivist principles. The two most popular benefits of LO use that respondents listed were illustration and/or visualization of topics and deepening conceptual understanding. These responses suggest respondents are using LOs to help their students construct concrete understanding of abstract ideas and of statistics concepts through representations other than the traditional lecture and text-based resources. Results also indicate that the respondents use LOs to reinforce topics with the intent of increasing their students' retention of material. Retention is a critical issue in introductory statistics classes due to many concepts that build on those taught previously.

Another popular reason respondents gave for using LOs was to motivate students, promote engagement, and increase their enjoyment. This may imply that respondents feel student attitudes and interests are important parts of the activities in which they participate. Attitude and motivation are important to study in a constructivist framework because these are intangible attributes that can affect learning outcomes (Carnell 2008).

Interestingly, many of the responses given to the question, "Why do you use LOs in your introductory statistics classrooms?" were restatements of the methods listed previously section. It is unclear if respondents did not understand the distinctions between methods and reasons or if it is because their methods were their reasons for choosing the LOs.

6.4 What advice would teachers give to colleagues about the use of LOs?

The majority of respondents' recommendations can be classified as general teaching practices however, implications can still be drawn from this data. The most common of respondents' advice was to try LOs before they are used in the classroom. Being comfortable with the features and knowing how an LO will work during class time will help teachers prepare materials. Even more, since respondents advised new users to start with simple designs, we can infer that beginning with too many LOs or too many different types of LOs could confuse teachers or students and possibly interfere with student learning processes.

Respondents' recommendations to carefully choose LOs to best fit learning goals may require some research due to the large number of existing statistics LOs. If teachers select LOs in accordance with specific learning objectives this may be an indication that they teach in an activity system that is consistent with constructivist principles. Relevant data will provide students with authentic contexts in which to study statistics. Results may show that respondents recommend exposing students to data that they feel is interesting may in turn increase students'

engagement with an activity, possibly providing more evidence of constructivism in respondents' teaching practices.

The survey data was lacking any negative responses about LOs, which could have many different implications. Perhaps teachers are overwhelmingly pleased with LOs as learning tools or only those with positive feelings about and experiences with LOs responded to the survey. Future research investigating negative aspects of LO use could provide more insight into both positive and negative implications of using LOs.

6.5 Limitations and Recommendations

Due to the survey's low response rate and the bias involved with a voluntary response sample, I do not recommend the results of this survey as appropriate for generalization. With no existing studies of this type in the context of statistics education in the literature and, as an initial study, anyone using these results in future studies will need to interpret them carefully as they are describing a small sample of statistics teachers who are all actively involved in the professional development received from being involved in scoring AP statistics exams. A larger survey, sampling from a general population of statistics teachers, would contribute further to the limited existing knowledge by giving a more accurate representation of how and why introductory statistics teachers are using LOs.

The overlap in instructors' perceived benefits and methods indicates many respondents may view benefits and methods as similar things. I anticipated that benefits of using LOs would refer to outcomes that show increased student understanding while methods of using LOs would include procedures and tasks involved in implementing them into an activity. In order to investigate the overlap, a future study could use interviews or a survey to collect data on teachers' definitions of benefits and methods. Additionally, the survey in this study did not provide respondents with definitions or examples for the types of LOs or use methods that they had to choose from, and there is no guarantee that teachers interpreted all of the terminology in the same manner. I received no clarifying questions from respondents about any of the categories or terms, but definitions of terms such as "active learning" can vary across teachers; thus, terminology is important to define for any future surveys or interviews (Zeiffler et al. 2008). In future research, I plan to include definitions of educational terms to reduce the effects of different interpretations.

In this study, I informally classified teaching practices and reasons for using LOs as either "teacher-centered" and "student-centered," without explicit definitions for what these look like in statistics classrooms the interpretation of the terms may vary. When asked why they use LOs, respondents gave responses that are consistent with student-centered constructivist principles, but when they selected their practices involving LOs, their responses appeared to be less student-centered. In the future this apparent conflict between the nature of these statistics teachers' practices and motivation should be investigated further. If teachers use educational technology for constructivist reasons (e.g. benefits to students), I strongly recommend they implement LOs into their classrooms using student-centered methods.

Both quantitative and qualitative research methods would be appropriate for future studies on the effectiveness of students using LOs. Future work could include learning studies, using pre- and

post-test assessment, comparing the effectiveness of different types of LOs or activities. Future statistics education studies could also compare the differences in student outcomes between classes taught at the secondary and post-secondary levels. Studies focused on which aspects of a constructivist activity system to consider, such as teacher and student attitudes, motivations, and interactions, may be better suited to qualitative research methods. I recommend researchers also use qualitative methods such as case studies or ethnographies to gain in-depth information on the role of LOs in introductory classrooms.

Current calls for empirical statistics education research indicate a need for a focus on specific technologies and how teachers can most effectively use them in the classroom (Schenker 2007; Zieffler et al. 2008). This study contributes to the existing knowledge by giving some indication of the popular types of LOs in statistics classrooms. Researchers can use this information to select popular LOs to study. However, because most survey respondents have used more than one type of LO in their classrooms, researchers should be cautious of excluding the impact that interactions with multiple types of LOs can have on student learning. In the future, statistics education researchers can also focus on the effectiveness of specific types of LOs and implementation methods. As newer technologies, such as tablets and smart phones that use apps, emerge more frequently in classrooms the effect of their different platforms on teachers' LO choices and implementation methods can be another line of research in statistics education.

7. Conclusion

This study forges a new direction in the statistics education literature by moving beyond describing LOs and recommending strategies for their use. The survey responses in this study provide unique information that contributes to existing knowledge regarding the use of LOs in the context of introductory statistics. For the first time, data from a pool of statistics educators regarding their use of LOs corroborated the information about statistics LOs that can be found in the literature. The survey results suggest that at least 10 percent of the AP statistics instructors who were invited to participate in the survey not only use LOs common to the literature, but also use LOs in varied applications.

This study also contributes to existing statistics education knowledge by connecting theory to practice; it provides a framework that educators can use to design effective teaching practices and researchers can use to design learning studies. The results from this study widen understanding of the relationship between activity theory and constructivism in the context of statistics education by connecting the literature on LOs to the tools that practicing statistics educators are using, as well as their reasons for choosing them.

The next step in advancing this line of research is to evaluate the quality of specific LOs and their impact on student learning. By pursuing this line of work, statistics educators and researchers can advance knowledge of the effectiveness of LOs to promote learning and ensure that students and teachers can access better quality tools and information on effectively implementing them into introductory statistics classrooms.

APPENDIX

THE SURVEY:

Directions: A learning object is a classification of education technologies that includes any reusable digital resources that can be used to support a learning objective. For the purpose of this survey please limit your responses to experiences with learning objects that are found online and can be freely accessed by anyone with a computer and an Internet connection.

At what level do you primarily teach introductory statistics?

- ☐ High School
☐ Post-secondary

In which state (or country) do you teach?

How many years have you been teaching?

How many years have you been teaching introductory statistics?

What is your average class size?

Which type(s) of learning objects have you used in your introductory statistics classes? (select all that apply)

- ☐ Animations
- ☐ Applets
- ☐ Blogs
- ☐ Case Studies
- ☐ Data Sets/Repositories
- ☐ Discussion Boards/Forums
- ☐ Electronic Textbooks/Materials
- ☐ Tutorials
- ☐ Videos
- ☐ Wikis
- ☐ Other 1
- ☐ Other 2

List any names and/or urls of specific learning objects you use in your classes.

How do you use learning objects in your introductory statistics classroom? (select all that apply)

- ☐ Collaboration between students
- ☐ Comparison to hands-on activity
- ☐ Follow-up to a previously introduced topic
- ☐ Group discussion
- ☐ In-class demonstration
- ☐ Independent exploration
- ☐ Introduction to new topic
- ☐ Refresher for previous material/prerequisites
- ☐ Simulation
- ☐ Structured activities
- ☐ Other 1

Why do you use learning objects in your introductory statistics classes?

If a colleague came to you interested about learning object use, what advice would you give them about effective strategies for implementing this type of technology into their teaching?

References

- Aberson, C. L., Berger, D. E., Healy, M. R., and Romero, V. L. (2002). An interactive tutorial for teaching statistical power. *Journal of Statistics Education*, 10(3). Retrieved from <http://www.amstat.org/publications/jse/v10n3/aberson.html>
- Aberson, C. L., Berger, D. E., Healy, M. R., and Romero, V. L. (2003). Evaluation of an interactive tutorial for teaching hypothesis testing concepts. *Teaching of Psychology*, 30(1), 75-78.
- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, R., and Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. Retrieved from <http://www.amstat.org/education/gaise/>
- Allen, D., Karanasios, S., and Slavova, M. (2011). Working with activity theory: Context, technology, and information behavior. *Journal of American Society for Information Science and Technology*, 62(4), 776–788.
- Anderson-Cook, C. M., and Dorai-Raj, S. (2003). Making the concepts of power and sample size relevant and accessible to students in intro stats courses using applets. *Journal of Statistics Education*, 11(3). Retrieved from <http://www.amstat.org/publications/jse/v11n3/anderson-cook.html>
- Akpınar, Y. (2008). Validation of a learning object review instrument: relationship between ratings of learning objects and actual learning outcomes. *International Journal of E-learning and Learning Objects*, 4, 91-302.
- Baki, A., and Çakıroğlu, Ü. (2010). Learning objects in high school mathematics classrooms: Implementation and evaluation. *Computers and Education*, 55(4), 1459-1469.
- Ben-Zvi, D. (2007). Using wiki to promote collaborative learning in statistics education. *Technology Innovations in Statistics Education*, 1(1). Retrieved from <http://escholarship.org/uc/item/6jv107c7>
- Bertie, A., and Farrington, P. (2003). Teaching confidence intervals with java applets. *Teaching Statistics*, 25(3), 70-74.
- Blejec, A. (2003). *Teaching statistics by using simulations on the Internet*. In Proceedings of the IASE Satellite Conference on Statistics Education and the Internet, Berlin, Germany.
- Boudourides, M. A. (2003). Constructivism, education, science, and technology. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, 29(3). Retrieved from <http://www.cjlt.ca/index.php/cjlt/article/view/83/77>

Boyer, S., and Stron, M. (2012). Best practices for improving survey participation. *Oracle*. Retrieved from

<http://www.google.com/url?sa=t&drct=j&dq=andsrc=sandsource=webandcd=2andved=0CDEQFjABandurl=http%3A%2F%2Fwww.oracle.com%2Fus%2Fproducts%2Fapplications%2Fbest-practices-improve-survey-1583708.pdf&ei=zpl4UpWGDcL2igKY34HIBwandusg=AFQjCNF2Y8axLwTZFarakxp4eD9RkbDfLAandbvm=bv.55980276,d.cGE>

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. National Academy Press, Washington, D.C.

Briggs, N. E., and Sheu, C.F. (1998). Using java in introductory statistics. *Behavior Research Methods, Instruments, and Computers*, 30(2), 246-249.

Burrill, G. (2010). Using data to make sense of statistics: the role of technology in scaffolding understanding. In *8th International Conference on Teaching Statistics. Ljubljana, Slovenia*.

Butson, R. (2003). Learning objects: Weapons of mass instruction. *British Journal of Educational Technology*, 34(5), 667-669.

Carnell, L. (2008). The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education*, 16(1).

Chance, B., Ben-Zvi, D., Garfield, J., and Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, 1(1). Retrieved from <http://escholarship.org/uc/item/8sd2t4rr>

Chance, B., and Rossman, A. (2006). *Using simulation to teach and learn statistics*. Proceedings of the 7th International Conference on Teaching Statistics, Salvador, Bahia, Brazil.

Christou, N., Dinov, I. D., and Sanchez, J. (2007). *Design and evaluation of SOCR tools for simulation in undergraduate probability and statistics courses*. Paper presented at the International Association of Statistical Education Conference on Statistics Education and the Internet, Satellite Conference to the 56th Session of the International Statistical Institute (ISI), Lisbon, Portugal.

Chu, A., Cui, J., and Dinov, I. D. (2009). SOCR analysis: Implementation and demonstration of a new graphical statistics educational toolkit. *Journal of Statistical Software*, 30(3). Retrieved from <http://www.jstatsoft.org/v30/i03>

Churchill, D. (2007). Towards a useful classification of learning objects. *Educational Technology Research and Development*, 55(5), 479-497.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational researcher*, 23(7), 13-20.

Darius, P., Oltoy, J-P., Solomin, A., Thas, O., Raeymaekers, B., and Michiels, S. (2000). *A collection of applets for visualizing statistical concepts*. In Proceedings in Computational Statistics 2000, Bethlehem, J.G. and P.G.M. van der Heijden (Eds). Physica Verlag.

Darius, P., Oltoy, J-P., Thas, O., Michiels, S., Raeymaekers, B. (2002). *Applets for experimenting with statistical concepts*. Proceedings of the 6th International Conference on Teaching Statistics, South Africa.

delMas, R. C., Garfield, J., and Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3). Retrieved from <http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm>

DePaolo, C. A. (2010). The stat-attic website: links to statistics applets for introductory courses. *Journal of Statistics Education*, 18(3). Retrieved from <http://www.amstat.org/publications/jse/v18n3/depaolo.pdf>

Dinov, I., and Christou, N. (2009). Statistics online computational resource for education. *Teaching Statistics*, 31(2), 49-51.

Dinov, I., Christou, N., and Sanchez, J. (2008). Central limit theorem: New SOCR applet and demonstration activity. *Journal of Statistics Education*, 16(2). Retrieved from <http://www.amstat.org/publications/jse/v16n2/dinov.html>

Dinov, I., and Sanchez, J. (2006). Assessment of the pedagogical utilization of the statistics online computational resource in introductory probability courses: A quasi-experiment. Proceedings of the 7th International Conference on Teaching Statistics, Salvador, Bahia, Brazil.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki, Finland: Orienta-Kosultit Oy.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., and Scheaffer, R. (2007). Guidelines for assessment and instruction in statistics education (GAISE) report. *Alexandria: American Statistical Association*. Retrieved from <http://www.amstat.org/education/gaise/>

Garfield, J., and Chance, B. (2000), Assessment in statistics education: Issues and challenges, *Mathematical Thinking and Learning*, 2(1), 99-125.

Garfield, J., Chance, B., and Snell, J. L. (2000). Technology in college statistics courses. In D. Holton et al. (Eds.), *The teaching and learning of mathematics at university level: An ICMI study* (pp. 357-370). Dordrecht: Kluwer.

Harvey, B. (2005). Learning objects and instructional design. *The International Review of Research in Open and Distance Learning*, 6(2). Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/227/310>

Hsu, Y. (2003). "The effectiveness of computer-assisted instruction in statistics education: A meta-analysis." PhD dissertation, University of Arizona, Arizona.

Johnson, J. (2010). *Designing with the mind in mind: Simple guide to understanding user interface design rules*. Morgan Kaufmann, Boston, MA.

Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm?. *Educational Technology Research and Development*, 39(3), 5-14.

Jonassen, D., Cernusca, D., and Ionas, G. (2007). Constructivism and instructional design: The emergence of the learning sciences and design research. *Trends and Issues in Instructional Design and Technology*, 2, 45-52.

Jonassen, D. H., Davidson, M., Collins, M., Campbell, J., and Bannan Haag, B. (1995). Constructivism and computer-mediated communication in distance education. *The American Journal of Distance Education*, 9(2), 7-26.

Jonassen, D. H., and Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educational Technology Research and Development*, 47(1), 62-79.

Kaptelinin, V., and Nardi, B. A. (2012). Activity theory in HCI: Fundamentals and Reflections. *Synthesis Lectures on Human-Centered Informatics*, 5(1), 1-105.

Karagiorgi and Symeou 2005

Kozulin, A. (1986). The concept of activity in Soviet psychology: Vygotsky, his disciples and critics. *American Psychologist*, 41(3), 264.

Lane, D. (1999). Rice virtual lab in statistics. *Behavior Research Methods, Instruments, and Computers*, 31(1), 24-33.

Lane, D., and Scott, D. W. (2000). Simulations, case studies, and an online text: A web based resource for teaching statistics. *Metrika*, 51, 67-90.

Leont'ev, A. N. (1978). The problem of activity in psychology. *Soviet Psychology*, 13(2), 4-33.

Mallory, T. E., and Jensen, G. C. (2001). Utah virtual lab: JAVA interactivity for teaching science and statistics online. *Behavior Research Methods, Instruments, and Computers*, 33(2), 282-286.

McDaniel, S. N., and Green, L. (2012). Independent interactive inquiry-based learning modules using audio-visual instruction in statistics. *Technology Innovations in Statistics Education* [online], 6(1). <http://www.escholarship.org/uc/item/322385kq>

- Mills, J. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, 10(1). Retrieved from <http://www.amstat.org/publications/jse/v10n1/mills.htm>
- Mittag, H. J. (2002). Java applets and multimedia catalogues for statistics education. In *Invited paper presented at the 6th International Conference on Teaching Statistics, Cape Town, South Africa*.
- Morrison, D. (2003). Using activity theory to design constructivist online learning environments for higher order thinking: A retrospective analysis. *Canadian Journal of Learning and Technology*, 29(3). Retrieved from <http://www.cjlt.ca/index.php/cjlt/article/view/87>
- Mulekar, M. (2000). Internet resources for AP statistics teachers. *Journal of Statistics Education*, 8(2). Retrieved from <http://www.amstat.org/publications/jse/secure/v8n2/mulekar.cfm>
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards-Mathematics*. Washington, DC: NGACBP/CCSSO.
- Neumann, D. L., Neumann, M. M., and Hood, M. (2010). The development and evaluation of a survey that makes use of student data to teach statistics. *Journal of Statistics Education*, 18(1). Retrieved from <http://www.amstat.org/publications/jse/v18n1/neumann.pdf>
- Ng, V. M., and Wong, K. Y. (1999). Using simulation on the Internet to teach statistics. *Mathematics Teacher*, 92(8), 729-733.
- Pearl, D.K., Garfield, J.B., delMas, R., Groth, R.E., Kaplan, J.J., McGowan, H., and Lee, H.S. (2012). Connecting Research to Practice in a Culture of Assessment for Introductory College-level Statistics. [Online: [http:// www.causeweb.org/research/guidelines/ResearchReport_Dec_2012.pdf](http://www.causeweb.org/research/guidelines/ResearchReport_Dec_2012.pdf)]
- Peterson, D. (2007). Usability theory, practice and evaluation for learning objects. *Learning Objects: Applications, Implications, and Future Directions*, 337-370.
- Phillips, B. (2003). *Overview of online teaching and Internet resources for statistics education*. In Proceedings of The IASE Satellite Conference on Statistics Education and the Internet, Berlin. (<http://www.stat.auckland.ac.nz/~iase/publications.php?show=6>)
- Saporta, G. (1999). Teaching statistics with Internet: A survey of available resources and the St@tNet project. (Arias, E. and Manninen, A., Eds.) Proceedings of the 52nd Session of the ISI, Bulletin of the International Statistical Institute (Tome LVII, Book 2, pp. 233-236). Helsinki: International Statistical Institute.

Schenker, J. D. (2007). "The Effectiveness of Technology Use in Statistics Instruction in Higher Education: A Meta-Analysis Using Hierarchical Linear Modeling." PhD dissertation, Kent State University College, Ohio.

Schneider, K. (2008). Two applets for teaching hypothesis testing. *Journal of Statistics Education*, 16(3). Retrieved from <http://www.amstat.org/publications/jse/v16n3/schneider.html>

Schwarz, C.J. (2007). Computer-aided statistical instruction – multi-mediocre techno-trash. *International Statistical Review*, 75(3), 348-354.

Sturm-Beiss, R. (2005). A visualization tool for one-and two-way analysis of variance. *Journal of Statistics Education*, 13(1). Retrieved from <http://www.amstat.org/publications/jse/v13n1/sturm-beiss.html>

Surry, D. W., and Land, S. M. (2000). Strategies for motivating higher education faculty to use technology. *Innovations in Education and Teaching International*, 37(2), 145-153.

Tishkovskaya, S., and Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, 20(2), Retrieved from www.amstat.org/publications/jse/v20n2/tishkovskaya.pdf .
Vygotsky, L. S. (1962). Language and thought. *Massachusetts Institute of Technology Press, Ontario, Canada*.

Vygotsky, L. S. (1981). The genesis of higher mental functions. In J.V. Wertsch (Ed.). *The concept of activity in Soviet psychology*. Armonk, NY: Sharpe. 144-188.

West, R. W., and Ogden, R. T. (1998). Interactive demonstrations for statistics education on the world wide web. *Journal of Statistics Education*, 6(3). Retrieved from <http://www.amstat.org/publications/jse/v6n3/west.html>

Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy, in *The instructional use of learning objects*. Retrieved from <http://reusability.org/read/chapters/wiley.doc>

Wiley, D., Waters, S., Dawson, D., Lambert, B., Barclay, M., Wade, D., and Nelson, L. (2004). Overcoming the limitations of learning objects. *Journal of Educational Multimedia and Hypermedia*, 13(4), 507-521.

Wilson, B. G. (1995). Metaphors for instruction: why we talk about learning environments. *Educational Technology*, 35(5), 25-30.

Yacovelli, S. R. (2003). Understanding learning objects: {The} basic "chunks". *College and University Media Review*, 10(1), 17-26.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., and 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), 1-23. Retrieved from <http://www.amstat.org/publications/jse/v16n2/zieffler.html>

CHAPTER 3. EXPLORING THE DESIGN QUALITY OF LEARNING OBJECTS DESIGNED TO HELP TEACH SAMPLING DISTRIBUTIONS OF SAMPLING MEANS: AN EXPERT EVALUATION

A paper to be submitted to *Technology Innovations in Statistics Education*

Rachel Graham

Abstract

Statistics educators have a large number of tools freely available to them on the Internet, many of which are designed to help teach the same topics. With multiple online tools available to help teach the same topic, it is important to evaluate them, so teachers can make informed decisions about the tools they use in their classes and researchers can investigate which tools most effectively promote learning. Evaluating online learning tools can also provide insight into their strengths and weaknesses and, thus, has the potential to inform developers of these tools about what are the important features of quality educational technologies. This study presents results of an expert-based evaluation of online learning objects designed to help teach sampling distributions of the sample mean. The results are then summarized to discuss the overall quality, the tools that are rated as highest quality, and the common characteristics between the highest and the lowest rated learning tools.

Key words: applets, evaluation, learning objects, statistics education

1. INTRODUCTION

Current K-12 and post-secondary mathematics and statistics education standards include recommendations to use educational technologies to teach statistics (Aliaga et al. 2010; Franklin et al. 2007; National Council of Teachers of Mathematics [NCTM], 2000). Educators can use technology to enhance learning statistics education by assisting students in making calculations, generating graphs, illustrating abstract topics through simulation and animations, and connecting them to data sources. There are thousands of learning tools online to help statistics educators teach many different topics in introductory statistics courses; however, the education literature rarely includes evaluations of statistics education technologies. Also, no standards have been developed specifically for evaluating the quality of statistics education technologies.

When statistics educators want to incorporate technology into their classrooms, one classification of technology available is the learning object (LO). Any reusable digital resource that can be implemented to support a learning objective can be classified as an LO (Akpinar, 2008; Baki & Cakiroglu, 2010; Butson, 2003; Churchill, 2007; Harvey, 2005; Peterson, 2007; Wiley, 2000; Yacovelli, 2003). Examples of LOs include applets, digital images and text, online databases, videos, animations, discussion forums, blogs, and wikis (Akpinar, 2008; Baki & Cakiroglu, 2010; Ben-Zvi, 2007). There are multiple LOs available for many topics commonly taught in introductory statistics courses (statistics LOs) including applets, online data sets, and videos. When there are multiple LOs available for the same topic, it is important to evaluate them, so researchers and teachers can investigate which tools most effectively promote learning. Evaluating LOs can also provide insight into their strengths and weaknesses and, thus, has the potential to inform updates or the designs of new LOs. Expert-based evaluation is one method of

evaluating usability in the field of human-computer interaction (HCI). With established evaluation criteria, experts in any field can use the principles behind expert-based evaluation to compare LOs. An expert-based evaluation of statistics LOs would involve the assessment of technologies using a set of established criteria, also known as heuristics.

The purpose of this evaluation was to gain information on the overall quality of existing LOs that can help statistics educators teach one topic covered in introductory statistics courses. Research has shown that introductory statistics students have difficulty developing conceptual understanding of abstract topics (Darius et al., 2002; Chance & Rossman, 2006). One of the most difficult of these abstract topics is the sampling distribution for a sample statistic because understanding such a distribution requires students to visualize infinite sampling (Chance, delMas, & Garfield, 2004). Visual representations of abstract topics such as simulations of random sampling and the central limit theorem (CLT) have the potential to be much more successful than verbal descriptions when illustrating sampling distributions for students (Chance & Rossman, 2006; Hodgson & Burke, 2000; Ng & Wong, 1999; Schwarz & Sutherland, 1997). Having formal evaluation results for LOs that can help teach sampling distribution of the sample mean (SDMLOs) is essential for teachers, students, and researchers to ensure that the most effective tools to promote learning are being used and studied. This paper summarizes the findings of a literature review completed to locate existing evaluation protocols and instruments designed to evaluate statistics LOs and LOs in general. Second, this study presents results of an expert-based evaluation of LOs designed for use in teaching sampling distributions of the sample mean, using criteria from in the educational technology literature and statistics educators as experts. The following research questions served as a focus for this study:

1. What is the overall quality of existing SDMLOs?
2. Which of the evaluated SDMLOs are rated as the best quality tools?
3. What characteristics differentiate between the highest and the lowest rated SDMLOs?

2. EXISTING EVALUATION SYSTEMS FOR LOs

To acquire information from the literature on the existing evaluation systems for statistics LOs I reviewed education journals, including journals with a focus on statistics education, mathematics education, and technology (see appendix for a complete list of sources included in the search). I was specifically looking for existing instruments designed for evaluating statistics LOs. In the context of statistics education, I found the inclusion criteria and informal evaluation schemes from two repositories of LOs, however, I was unable to find any evaluation instruments designed specifically for evaluating statistics LOs. As a result, I with the widened the search criteria to include evaluation instruments measuring the quality of LOs without a specific context applied. Educational technology, education, and HCI journals were then reviewed. Through this search, I found reports on two evaluation instruments that included empirical data on validation for the instruments. The existing evaluation schemes and the instruments found are described below.

2.1 Learning Object Evaluation Systems Currently Used in Statistics Education

In July 2014, a Google search for “statistics learning objects” yielded more than 7.5 million results and the query “statistics applets” returned over 670,000 results. The statistics education literature also provides many examples of available statistics LOs (Aberson, Berger, Healy, & Romero, 2003; Ben-Zvi, 2007; Blejec, 2003; Chance, Ben-Zvi, Garfield, & Medina, 2007; Chance & Rossman, 2006; Darius et. al, 2002; DePaolo, 2010; Dinov & Christou, 2009; Dinov, Christou, & Sanchez, 2008; Lane, 1999; Lane & Scott, 2000; Mittag, 2002; Mulekar, 2000; Saporta, 1999; Schneiter, 2008; Schenker, 2007; Schwarz, 2007; Tishkovskaya & Lancaster, 2012; West & Ogden, 1998). If the number of Internet search results is overwhelming or time consuming, teachers may benefit from using learning object repositories (LORs), which are collections of LOs that are often searchable by topic to help quickly narrow the results in a user’s search.

The Multimedia Educational Resource for Learning and Online Teaching (MERLOT) hosts an expansive online LOR (MERLOT, 2005). MERLOT provides links to over 34,000 different LOs categorized by subject (i.e., arts, education, mathematics and statistics) and material type (i.e., animation, drill and practice, or simulation). Users specifically interested in ‘mathematics and statistics’ LOs will find over 3,900, including 578 ‘statistics and probability’ LOs. The main evaluation system MERLOT designers built into the LOR is a rated peer review. Each peer review focuses on three broad criteria: quality of content, potential effectiveness as a tool for teaching and learning, and ease of use. The MERLOT peer reviewer rates the LO by giving it a number of stars, 1 to 5, with one star indicating the LO is not valuable and a rating of five stars indicating excellence all around (MERLOT, 2005). The second evaluation scheme in the MERLOT LOR includes informal member ratings and comments. If members leave a comment, they can choose to give an overall rating for the LO along with it. These ratings are based on the same 5-star scale. To summarize peer reviews, the MERLOT LOR apparently averages the five-star scores (including partial stars).

Another large LOR available links specifically to statistics applets; this is the STATistics Applets for Teaching Topics (STAT-ATTIC) website (<http://sapphire.indstate.edu/~stat-attic/index.php>). STAT-ATTIC contains “links, with descriptions, to approximately 600 publicly available applets on topics commonly taught in introductory courses” (DePaolo, 2010, p.1). STAT-ATTIC only includes applets if they are freely accessible to anyone connected to the Internet and do not require any downloads (DePaolo, 2010). STAT-ATTIC allows users to rate any of the applets housed in the repository by giving feedback through an overall score and comments. Feedback score options are a five-point Likert scale with ratings from poor to excellent. STAT-ATTIC displays the average rating for all submitted scores and the number of “votes” used to calculate the average display as part of the descriptive information.

However, both MERLOT and STAT-ATTIC fail to help educators make an informed decision on which LOs are the highest quality because their evaluations are far too informal. The star ratings these LORs utilize provide only a single average score for each LO; in fact, many LOs in both the MERLOT and the STAT-ATTIC LORs have average ratings based on less than ten informal reviews. Ideally, any formal evaluation scheme of online educational technologies would provide evaluators with information regarding multiple aspects of the tools being scored.

The existing evaluation schemes of statistics LOs do not provide ratings on more than overall quality, leading me to conclude that they gave insufficient criteria to design a formal evaluation.

2.2 Stand-Alone Learning Object Evaluation Instruments

To evaluate the quality of statistics LOs, I was interested in finding an existing valid, reliable instrument to gauge the overall quality of online learning tools. After I widened focus of the literature review search to investigate any developed LO evaluation instruments for any subject, I found two stand-alone evaluation instruments in the literature. The most cited stand-alone evaluation instrument for LOs is the Learning Object Review Instrument (LORI) developed by Vargo, Nesbit, Belfer, and Archambault (2003). The most current version, LORI 1.5 (Nesbit & Li, 2004), contains nine evaluation criteria and sixteen evaluation items (see Table 1). Evaluators using the LORI score each of the criteria items using a five point Likert-scale (1 is low, 5 is high).

<u>Evaluation Criteria</u>	<u>Item Description</u>
Content Quality	Accurate information Important ideas emphasized Appropriate level of detail given Unbiased representation of cultural and ethnic diversities
Learning Goal Alignment	Appropriate learning goals declared for users Content aligned with learning goals Achievable learning goals
Feedback and Adaptation	Appropriate feedback for different user inputs: Messages displayed Phenomena based on input is modeled
Motivation	Motivating and interesting topics and content Content relevant to users goals and interests
Presentation Design	Visual and auditory output does not interfere with learning: Legible, error-free text Meaningful headings Narration
Interaction Usability	Quality interface design: Easy and clear navigation Consistent styles
Accessibility	Follows IMS Guidelines for Accessible Learning Applications W3C compliant at 'AAA' level Assistive technologies accommodated
Reusability	Stand-alone resource Usable in many different contexts
Standards Compliance	International standards met: IEEE Learning Object Metadata IMS, IEEE, SCORM, and W3C technical guidelines followed Metadata in tagged code

Table 1: Evaluation Criteria in the Learning Object Review Instrument (LORI), Version 1.5 (Nesbit, Belfer, & Leacock, 2004)

Akpinar (2008) completed a study validating the LORI instrument; over 500 students from grades 4 to 10, and their 24 teachers evaluated 24 LOs designed for biology, chemistry, general science, and mathematics courses. In the validation study, Akpinar gave both pre- and post-tests to each student to assess the change in their content knowledge. Both students and teachers also completed a usability questionnaire to collect opinions on the usability of the LOs included in the study. Akpinar investigated the interactions between the LORI scores and measures from LO designers, teachers, and students. Correlations between scores on all items in the LORI instrument were significant at the 1% level. Results from the usability questionnaires and the LORI ratings did not correlate with increased content knowledge of students. As a result of these

outcomes, researchers concluded that LORI-based assessments should not be used to predict users' learning outcomes.

The other standalone evaluation instrument in the literature was the Learning Object Evaluation Metric (LOEM), created by Kay and Knaack (2008). The LOEM has 17 items, each categorized under one of the four main evaluation criteria: Interactivity, Design, Engagement, and Usability. In Table 2, I have described the items on the instrument along with the criteria they measure.

<u>Evaluation Criteria</u>	<u>Item Description</u>
Interactivity	Meaningful interactions Overall control Multimedia adds to learning value
Design	Consistent look and feel throughout LO Clear and organized layout Effective labeling Readability of text
Engagement	Quality of feedback Modern and appealing look Quality of graphics Learning modalities Motivation
Usability	Natural interface for use Orientation of elements Navigation Cues Helpful instructions Appropriate language level

Table 2: Learning Object Evaluation Measure (LOEM) evaluation criteria (Kay & Knaack, 2008).

Each item on the LOEM instrument has a range of possible scores from 1 to 3 according to descriptive criteria for that item on a rubric. For example, the "Natural to Use" item under the Usability construct on the LOEM measures the intuitiveness of the interface or the ease of use. The Interactivity construct has three items and a range of possible scores from 3 to 9. The Design construct has four items, so the possible scores range from 4 to 12; both Engagement and Usability constructs have five items, so the scores range from 5 to 15. Scores for each construct are summed to determine a total score for the LO being evaluated; total scores range from 17 to 51.

LOEM designers extensively reviewed existing instructional design and LO research as the foundation for the instrument constructs and items. Creators of the LOEM found the instrument to be valid by comparing LOEM scores to evaluation scores on more than 40 learning objects from 1,113 students (grades 6-12) and 33 mathematics and science teachers. To show reliability

the study reported Cronbach's alphas, which they found to be acceptable (> 0.70) for all four constructs; in addition, agreement percentages between their evaluating teachers was greater than 90%. The final version of the LOEM was the result of a principal component factor analysis that confirmed the four significant constructs; Interactivity, Design, Engagement, and Usability. Results of the evaluation led them to conclude that the LOEM effectively measures the quality of LOs.

Although the information available from evaluations of LOs designed specifically for use with statistics learning objectives is limited, this review shows that there are existing stand-alone evaluation instruments available. While this study does not focus on how LOs affect learning, the purpose of any LO is to help students reach proficiency in a learning objective (Peterson 2007) and, therefore, potential to enhance learning should be considered when performing any LO evaluation. When selecting the instrument for this study's evaluation, I considered Akpinar's recommendations discouraging the LORI's use as an instrument to predict learning outcomes, and the promising validation results for the LOEM; I chose the LOEM.

3. METHOD

3.1 Overview

In their study validating the LOEM evaluation instrument, Kay and Knaack (2008) completed an extensive literature review and used the results to help with the LOEM design. Kay and Knaack also compiled a list of characteristics that future evaluations could improve upon based on the LO evaluation studies they researched in their review of the literature. The items included in the LOEM instrument evaluate existing LOs from any discipline. I chose the LOEM as the preferred instrument for this study because Kay and Knaack (2008) found evidence that it was valid and reliable. I obtained permission to use the LOEM for this evaluation from Dr. Robin Kay on December 28, 2012. For this study's evaluation, statistics instructors used the LOEM instrument to assess the quality of nine existing SDMLOs.

3.2 Materials

All of the SDMLOs I included in this evaluation were applets simulating at least one aspect of the random sampling behind a sampling distribution. I compiled a list of LOs specifically designed for use in teaching sampling distributions of sample means from a search of the literature, the Internet and within the LORs, CAUSEweb.org, MERLOT and STAT-ATTIC. I found eleven distinct SDMLOs, but eliminated two of the LOs to avoid potential confusion for the evaluators. One of the eliminated LOs had problems loading and the other was a collection of LOs, of which only one was appropriate for this evaluation. I used a random number generator and put the remaining nine LOs in random order and I requested that all evaluators evaluate the LOs in that order. Brief descriptions of each of the nine LOs I included in the evaluation are below in the order they were evaluated in this study.

3.2.1 LO-A –Central Limit Theorem Applet (CAUSE, [CAUSEweb.org](http://causeweb.org))

When users interact with the single-page applet, available from CAUSE, they can choose between five different distribution shapes and set population parameters for three of these, as seen in Figure 1. After users select this information, the LO displays a histogram of the sampling distribution for the sample means in the middle of the screen with a normal distribution curve fit over the top of the graph. The sample size, sample mean, and variance are also displayed for the last simulated sample drawn.

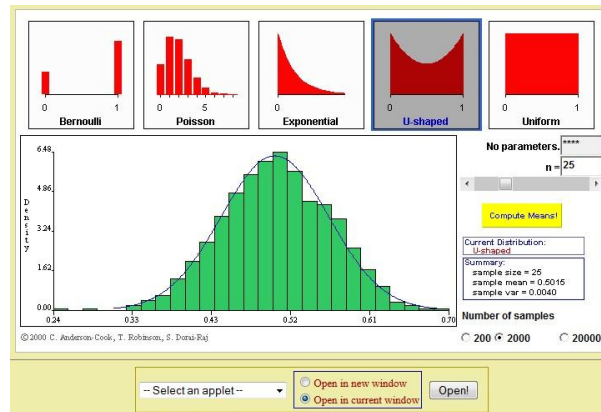


Figure 1: LO-A <http://www.causeweb.org/repository/statjava/CLT2Applet.html>

3.2.2 LO-B – Sampling from a Population (Prentice Hall via Dr. R. Webster West, Texas A &M University)

Dr. R. Webster West, Professor and Associate Director of Online Learning at Texas A&M University, has links to many statistics LOs on his web page, including a set of Prentice-Hall applets. The “Sampling from a Population” LO, is the first of two Prentice-Hall applets from this site included in this evaluation. The applet contains two plots arranged vertically (see Figure 2). The top plot represents the distribution of a population and the lower plot shows the distribution of a randomly selected sample from the given population. Users can choose one of the pre-set shapes for the population distribution from a drop-down menu or custom draw one onto the graph. Descriptive statistics are displayed for both the population and random sample distributions.

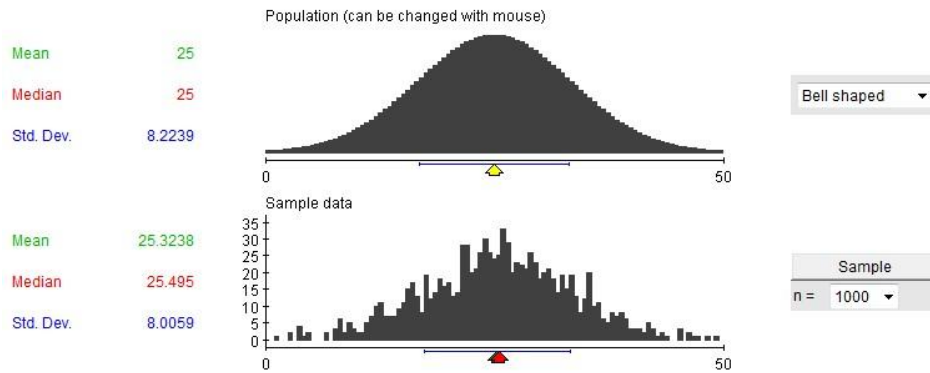


Figure 2: LO-B <http://www.stat.tamu.edu/~west/ph/popsample.html>

3.2.3 LO-C – Random Sampling Animation (Dr. Gary Kerbaugh, Fayetteville State University)

On his personal web page, Dr. Gary Kerbaugh, Professor at Fayetteville State University, provided a link to this applet, which includes an animation that randomly samples from a normal distribution. Numbers representing a population float in a container and when users click the 'New Sample' button, an animated measuring cup dips numbers out of the container and then simulates pouring out the sample (see Figure 3). After the sample simulation, a histogram from the sample data displays with a normal curve fit over the graph. The mean and standard deviation for each random sample appears under the histogram.

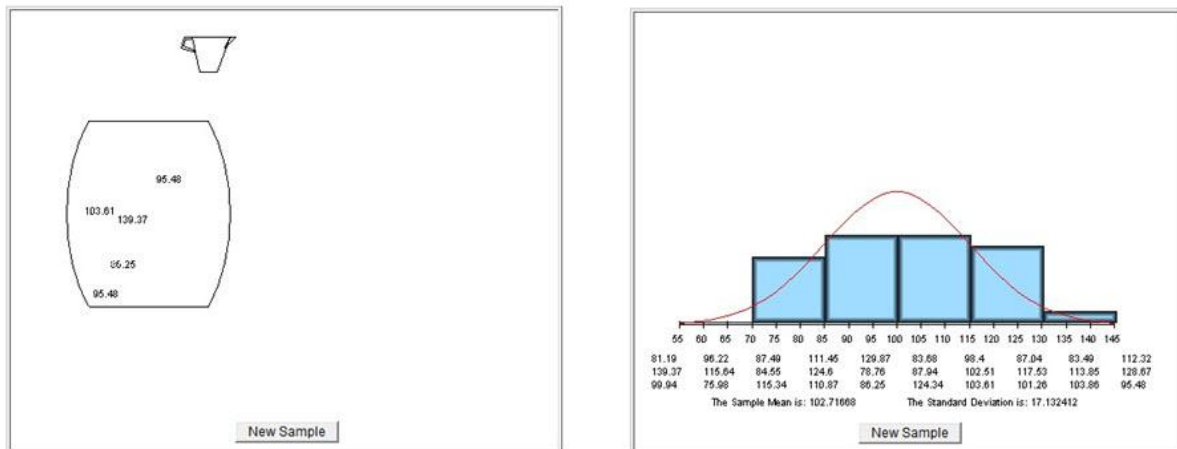


Figure 3: LO-C <http://kerbaugh.uncfsu.edu/stats/barrel/barrel.html>

3.2.4 LO-D – Central Limit Theorem Applet (Department of Chemistry, University of Athens)

The Department of Chemistry at the University of Athens has developed a set of educational applets, available in both English and Greek. One of these applets illustrates the central limit theorem by displaying the effect that sample size has on the distribution of \bar{x} (see Figure 4). The user can choose one of many different population distribution shapes, set the sample size by clicking *increase/decrease* buttons, and select *small*, *medium*, or *large* for the number of samples drawn.

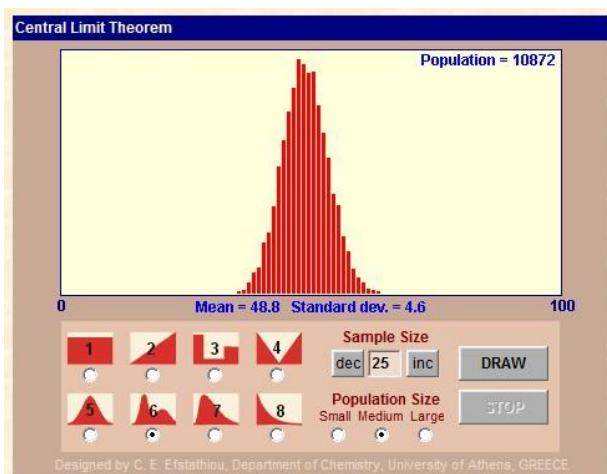


Figure 4: LO-D http://www.chem.uoa.gr/applets/AppletCentralLimit/Apl_CentralLimit2.html

3.2.5 LO-E - Sampling Distributions Applet (Rice Virtual Lab in Statistics)

The most popular SDML, in both an online Google search and in the literature, is from the Rice Virtual Lab in Statistics (RVLS). When users access the applet, they see four plots. At the top is a distribution of a population with descriptive statistics shown to the left (see Figure 5). The second graph displays sample data if the user chooses to animate the sampling. After users simulate sampling, the third and fourth graphs will show the sampling distribution for a choice of the mean or other descriptive statistics.

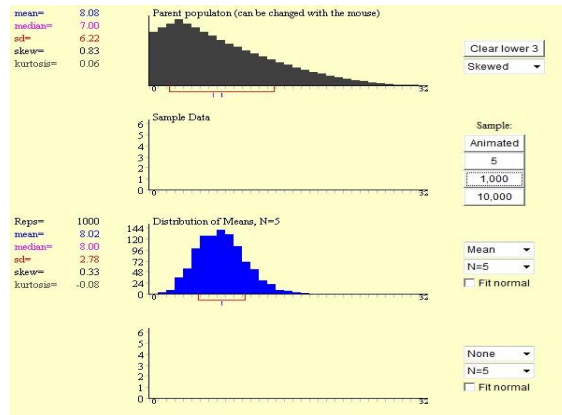


Figure 5: LO-E http://onlinestatbook.com/stat_sim/sampling_dist/index.html

3.2.6 LO-F – Central Limit Theorem (CLT) Applet (intuitor.com)

This LO is an applet that has display boxes, one above the other, each with an option box to its left. A distribution of a population displays in the top plot. The default population is uniformly distributed but the user has the option of changing the distribution to normal, skewed, or bimodal using the options to the left of the graph. Descriptive statistics for the population including sample size, mean, and standard deviation are also shown in the box. At the bottom, the plot there displays the sampling distribution of the sample mean along with descriptive statistics, as shown in Figure 6.

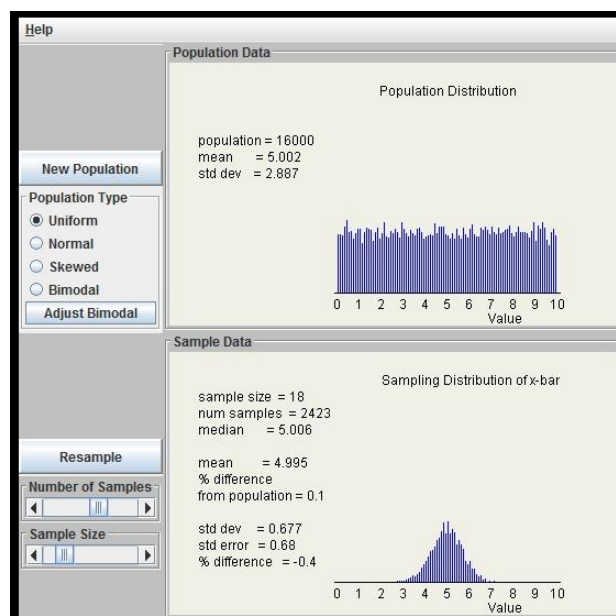


Figure 6: LO-F www.intuitor.com/statistics/CentralLim.htm

3.2.7 LO-G – Central Limit Theorem Applet (Department of Economics, San Jose State University)

The central limit theorem applet from San Jose State University contains a series of images that illustrate the effect that sample size has on sampling distributions of sample means. Each of the illustrations of the central limit theorem assume 2000 samples are drawn, but each graph shows this for a different sample size, ranging from $n = 1$ and increasing by a factor of two until $n = 16$ (see Figure 7). If the user refreshes the page, results are shown for a different 2000 simulated samples.

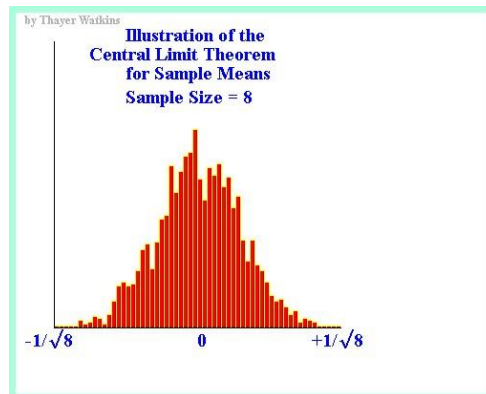


Figure 7: LO-G www.applet-magic.com/samplemean.htm

3.2.8 LO-H – Sampling Distribution of Sample Means Applet (VassarStats: Website for Statistical Computation)

One applet, offered by VassarStats: Website for Statistical Computation, opens with a series of sequential pop-up boxes with prompts for information necessary to simulate the sampling distribution. Users input the population mean, standard deviation, and the sample size of the repeated samples. After a user responds to each of the prompts, a graph of the sampling distribution and descriptive statistics display. Critical values and tail probabilities associated with the sampling distribution appear below the output. Users can reload the series of prompts using a button in the applet.

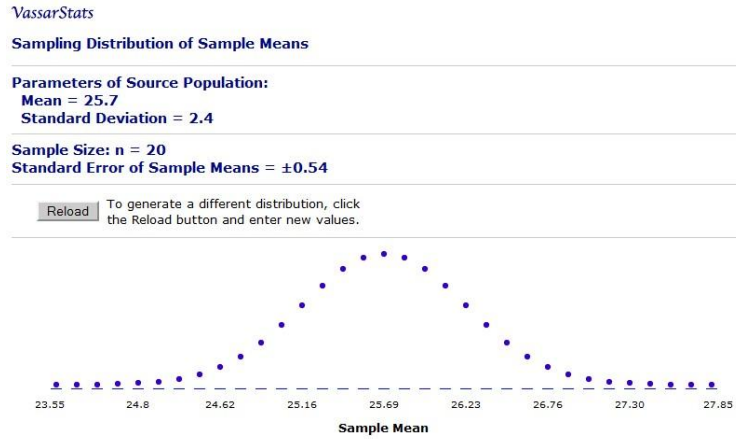


Figure 8: LO-H <http://vassarstats.net/mlsamp.html>

3.2.9 LO-I – Sampling Distributions Applet (Prentice Hall via Dr. R. Webster West, Texas A & M University)

This Sampling Distribution Applet is the second of the Prentice-Hall SDMLLO applets included in this evaluation. The interface of the applet consists of four different plots arranged vertically. The top plot represents the distribution of a population, the second plot shows sample data for the last simulated sample, and the last two plots display a sampling distribution for a selection of statistics. There are drop-down boxes, entry fields, and buttons to allow the user to change input values and manipulate the graphs. Displayed to the left of each of the four plots are their respective means, medians, standard deviations, and, for the last two plots only, sample sizes (see Figure 9).

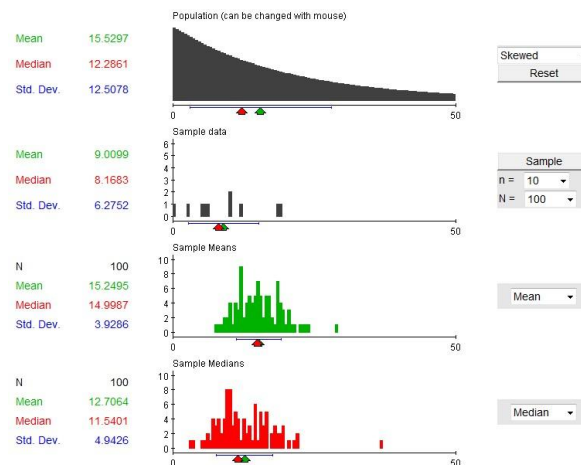


Figure 9: LO-I www.stat.tamu.edu/~west/ph/sampledlist.html

My initial impression of the SDMLLOs chosen for evaluators to score in this study was that there would be a wide range in the scores because of the differences in the features and interactive options between the LOs. After exploring the features of each LO, I found that LOs C and G had

the fewest number of components for users to interact with. Due to their low number of interactivity options I initially predicted that LO-C and LO-G would have the lowest total LOEM scores.

If teachers use LOs that do not show the sampling distributions, students may need to record the means of multiple samples and graph them separately to visualize the distribution of the statistic. LO-B and LO-C are the only two LOs in this evaluation that simulate one sample being drawn at a time but not a sampling distribution, while all of the other LOs illustrate random sampling and the resulting sampling distributions of the means. It was my belief going into this analysis that the LOs that contain a sampling distribution would be evaluated as higher quality learning tools than those that do not (LO-B and LO-C) because their features can provide a bridge between the topics of single sampling and multiple sampling without any additional tasks.

There are visible similarities in the interface designs of some sets of the LOs I included in this evaluation. Similarities between LO-E and LO-I include their four vertically arranged plots, descriptive statistics on the left side of the plots, and similar options in the data collection fields. These LOs also display descriptive statistics for each of their plots to the left and have interactive menus and buttons to the right of the plots. Considering the likenesses between these LOs E and I, I initially predicted that these LOs would receive similar LOEM scores.

3.3 Procedures

To invite evaluators to participate, I sent an email to eighteen of my professional contacts including university colleagues and my fellow Advance Placement (AP) Statistics Readers who had previously shown interest in the study. The invitation email contained information about the evaluation process, a copy of the evaluation instrument (LOEM), and consent information. Additionally, I asked evaluators to read through the LOEM descriptions for each item and to reply giving their consent to participate and with any questions. Fourteen evaluators replied giving their consent. These evaluators sent no questions so no additional preparation or training on the instrument was given. After receiving consent, I emailed the evaluators the list of links to the SDMLs, in the order that they were to evaluate them, along with a demographic questionnaire and a spreadsheet designed for recording their scores (see appendix). Each evaluator recorded scores for each item for the nine LOs into a provided document. After all evaluations were complete, I compiled the scores into an Excel spreadsheet for data analysis.

3.4 Data Analysis

The data analysis for this evaluation included investigations of the LOEM data variables including item, component, and total score across the entire data set, evaluator, and LO. I compiled the evaluators' LOEM scores for each LO into one master data set which I used for a series of analyses using R®, an open-source statistical software freely available to download online. In order to describe the distributions of the component and the total scores I computed descriptive statistics for them across all LOs and evaluators. Additionally, to compare the distributions of each component to those reported in the LOEM validation study (Kay and Knaack, 2008) I ran two-sample mean hypothesis tests to investigate evidence of differences between the mean component scores from the Kay and Knaack LOEM validation study and

those from this evaluation. The validation study did not provide a standard deviation for Total Scores and therefore, I did not complete a two-sample test using that data. To explore similarities to the previously reported statistics for the LOEM, I also tested each evaluator's average total score against the mean total score (35) reported by Kay and Knaack (2008) by completing one-sample mean t-tests.

To measure LOEM component reliability, I first computed Cronbach's alpha to quantify the association between the evaluators' individual item ratings and component scores (Cronbach, 1951; Gliem & Gliem, 2003). I explored inter-rater reliability for each item, component, and the total score using Krippendorff's alphas including 95% confidence intervals from ($n = 10,000$) bootstrapped samples to provide an appropriate interpretation of each alpha (Hayes & Krippendorff, 2007).

Next, I completed analyses focused on the LOs. The possible analyses at the LO level are limited because of the small sample size. I only used the Interactivity component Total Score data in subsequent analyses because the Design, Engagement, and Usability components did not have an acceptable inter-rater reliability, therefore, no further analyses on the LOEM components were done at the LO level. For each LO, total scores for each component and for the overall LO Total Score were generated. To analyze the quality of individual LOs, I computed the mean score was computed across evaluators for each of the nine LOs and an overall average score for each component and the total score. Finally, I ranked each of the evaluator's Total Scores for each LO and reported the average rank for each one.

4. RESULTS

This study was an evaluation by current statistics educators of existing LOs for one specific topic in introductory statistics, the sampling distribution of the sample mean. Fourteen evaluators, who were high school or postsecondary level statistics instructors, completed a total of 124 evaluations on nine SDMLOs. Twelve evaluations contained scores for all nine LOs, however, two evaluators each had one LO they could not get to load (LO-C and LO-G). The two evaluators who had applets that did not load indicated they tried multiple times to get the applet to load with no success. Evaluators all indicated that they had used LOs to support learning in their classrooms prior to this study. The mean number of years the evaluators had taught was 9.2 years. Eleven of the evaluators taught introductory statistics at the postsecondary level and three had taught AP statistics at the high school level. Of the twelve evaluators who provided information on their gender, males and females were equally represented. Eight of the evaluators were currently teaching in Iowa, two in Texas, one in Georgia, and one in Wisconsin. Two of the evaluators chose not to provide information on their locations.

4.1 LOEM Component and Total Score

Possible ranges of scores for each component and total score along with descriptive statistics are displayed in Table 3. The results of the two-sample tests for a difference between the means of the Kay and Knaack LOEM validation study and this evaluation study are also displayed. The observed ranges of scores were nearly identical to the possible ranges of scores on the LOEM. The results showed evidence that there was evidence of a significant difference between the

mean Usability component scores in this study were different from those in the LOEM validation study. However, Interactivity, Design, and Engagement did not show evidence of a difference, indicating the results of this study are consistent with those in the LOEM validation study for those components.

LOEM Component	Total Score				
	Possible Scores	Observed Scores	Median Score ($n = 124$)	Mean Score (SD) ($n = 124$)	z -value ($\mu_1 \neq \mu_2$)
Interactivity	3 – 9	3 – 9	7	6.3 (2.0)	-1.36
Design	4 – 12	5 – 12	9	9.3 (1.8)	0.00
Engagement	5 – 15	5 – 14	8.5	8.8 (2.5)	1.95
Usability	5 – 15	5 – 15	11	11.1 (2.5)	-2.64**
Total Score	17 – 51	19 – 50	35.5	35.4 (7.6)	
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$					

Table 3: Summary of LOEM Component and Total Score data.

The side-by-side box plots (Figure 10) show that the distribution of the Total Scores is roughly symmetric while the distributions for the components are skewed. These results were corroborated by the Shapiro-Wilks tests I ran for each component and Total Score. Total Score was the only variable that showed significant evidence of being normally distributed. Having found the normality assumption violated for all of the LOEM components, I completed minimal analyses to further interpret the findings from the construct-level data.

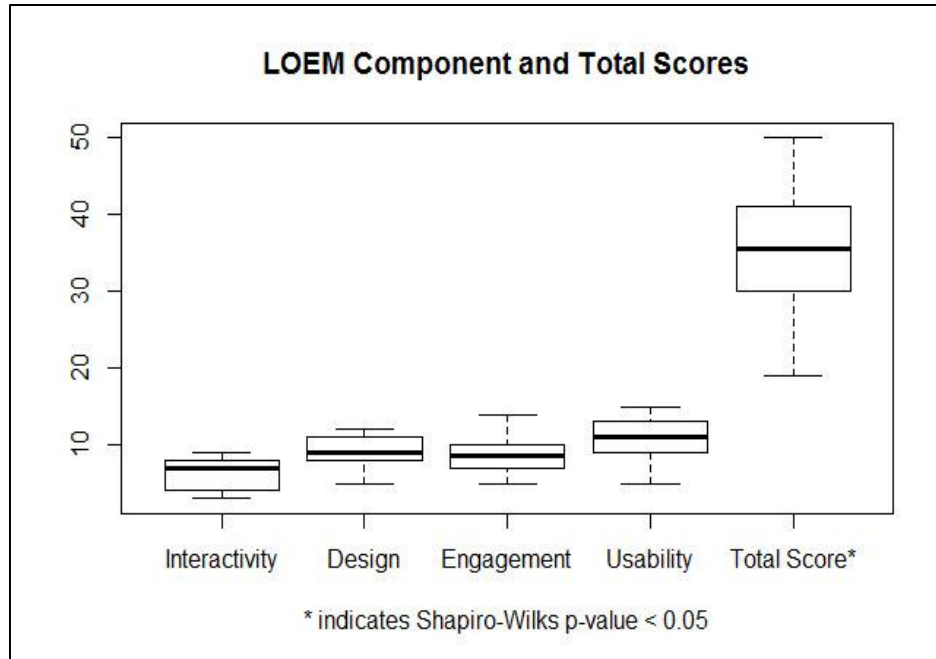


Figure 10: The LOEM component and total score distributions.

To compare the mean Total Scores for each evaluator to the understood average of the LOEM scores, I used one-sample mean t-tests to determine if each evaluator's average total score was significantly different than 35, the hypothesized mean LOEM Total Score (Kay and Knaack, 2008). Results showed evidence that Evaluator 1's average total score was significantly higher than the hypothesized mean LOEM Total Score of 35. The results of the t-test for Evaluator 1 indicated that those scores could inflate the overall scores for each SDMLO and consequently, I excluded Evaluator 1's responses from any further analyses.

4.2 Reliability

The Cronbach's alphas for the components of this evaluation using LOEM were 0.87 (Interactivity), 0.62 (Design), 0.79 (Engagement), and 0.75 (Usability). While the alphas for the Engagement, Interactivity, and Usability components are considered acceptable in the social sciences (Santos, 1999), the Design component produced a questionable level of reliability and any results for this component should be interpreted with caution (George & Mallery, 2003; Gliem & Gliem, 2003; Kay & Knaack, 2008; Tavakol & Dennick, 2011).

Krippendorff's α is a versatile measure of inter-rater reliability that accounts for different levels of data (i.e. nominal, ordinal, interval) and missing data values (Hayes & Krippendorff, 2007). An acceptable level of reliability requires $\alpha > 0.80$ and a reliability $0.67 \leq \alpha \leq 0.80$ should be considered tentative (Hayes & Krippendorff, 2007). The alpha values for the Interactivity component and the Overall Control item indicate tentative reliability between evaluators. Total Score and the Meaningful Interactions item both had $\alpha = 0.66$ which is only one one-hundredth lower than the lowest alpha required to attain tentative reliability. To most accurately interpret the Krippendorff's alpha statistics, I also computed bootstrap confidence intervals, estimating the

true alpha statistics from distributions composed of simulated samples of 10,000 α values. The alpha statistics and the confidence intervals are displayed in Table 4.

LOEM Item	Krippendorff's Alpha	95% CI of Bootstrap Sample (iterations=10,000)
<u>Interactivity</u>	0.70*	(0.21, 0.81)**
Meaningful Interactions	0.66	(0.21, 0.84)**
Overall Control	0.69*	(0.29, 0.85)**
Multimedia adds learning value	0.52	(0.21, 0.63)
<u>Design</u>	0.46	(0.10, 0.46)
Consistency	0.21	(0.00, 0.35)
Layout	0.41	(0.05, 0.59)
Labeling	0.16	(-0.01, 0.29)
Readability	0.21	(0.01, 0.31)
<u>Engagement</u>	0.52	(0.21, 0.65)
Quality of Feedback	0.35	(0.09, 0.52)
Attractive	0.28	(0.08, 0.45)
Graphics	0.16	(0.01, 0.27)
Learning Mode	0.38	(0.13, 0.54)
Motivation	0.43	(0.11, 0.64)
<u>Usability</u>	0.52	(0.24, 0.64)
Natural to Use	0.28	(-0.03, 0.47)
Orientation	0.21	(-0.01, 0.33)
Navigation Cues	0.38	(0.08, 0.50)
Instructions	0.43	(0.17, 0.56)
Appropriate Language	0.42	(0.10, 0.58)
Total Score	0.66	(0.36, 0.75)*

*interval includes tentative alpha (> 0.67) **interval includes acceptable alpha (> 0.80)

Table 4: Krippendorff alpha values and bootstrap confidence intervals for the LOEM items, components, and total score.

The only LOEM components with confidence intervals that included acceptable alpha values were Interactivity and Total Score and the LOEM items with acceptable alpha values were Meaningful Interactions and Overall Control. These confidence interval results were consistent with the alpha values I calculated previously. Individual items on the LOEM instrument have possible scores from 1 to 3, which limits the quantitative analysis that can be done on that level and therefore, I completed no further analysis at the item level. After investigating the normality of the data and the reliability of the responses, I completed the next analyses focusing on the quality of the SDMLOs using only the Interactivity and Total Scores.

4.3 Interactivity Scores

On the LOEM instrument the Interactivity component was the first component evaluators scored; it contains three items resulting in a possible Interactivity score from 3 to 9. The three Interactivity component items are designed to measure whether or not the LO provides meaningful interactions to the user, how much control the user has, and the learning value of the multimedia components of any LO. Table 5 displays the ranges of scores this study's evaluators gave to the Interactivity component (the sum of the scores for the three items), along with the median score for each LO. This study's Interactivity results showed that the highest median Interactivity score (8) in this evaluation was attained by LOs E, F, and I. LOs A and B also had a median Interactivity score (7) that placed them in the upper half of the possible scores. LO-G received the lowest scores for the Interactivity component with a median equal to the lowest possible score for this component (3). Evaluators' median scores for LOs C and H were 4, which is also on the lower end of the possible scores for Interactivity.

Learning Object	Observed Interactivity Scores	Median Interactivity Score
A	6 – 8	7
B	5 – 9	7
C	3 – 5	4
D	5 – 9	6
E	7 – 9	8
F	5 – 9	8
G	3 – 5	3
H	3 – 5	4
I	7 – 9	8

Table 5: Observed ranges of scores and median Interactivity scores for the SMLOs.

Due to the non-normality of the Interactivity data, I restricted my computations for the Interactivity component data to descriptive statistics. Additionally, I used the median as the measure of central tendency for this component. While the information gained from this component's data is limited it does provide a basis for comparison between the LO's scores from their Interactivity component and their Total Scores.

4.4 Total Scores

Initially, I graphed the total scores from all evaluators to describe the distributions of scores for each SDMLO (see Figure 11). While the central tendencies of the scores vary across LOs, the spread in scores for LO-A to LO-H appears to be similar; LO-I's spread appears to be the smallest. To further summarize and describe the Total Scores I computed the mean, standard deviation, median, and range of scores for each SDMLO (see Table 6). Two-tailed one-sample mean tests for a difference between the mean LO Total Score from the LOEM validation study ($\mu = 35$) (Kay and Knaack, 2008). The t-tests show LOs B, E, and I had average scores

significantly higher than the average found in the LOEM validation study ($\mu = 35$), while LOs C, G, and H were significantly lower than average. This analysis also showed no evidence that the mean total scores for LO-A, LO-D, and LO-F were significantly different than 35.

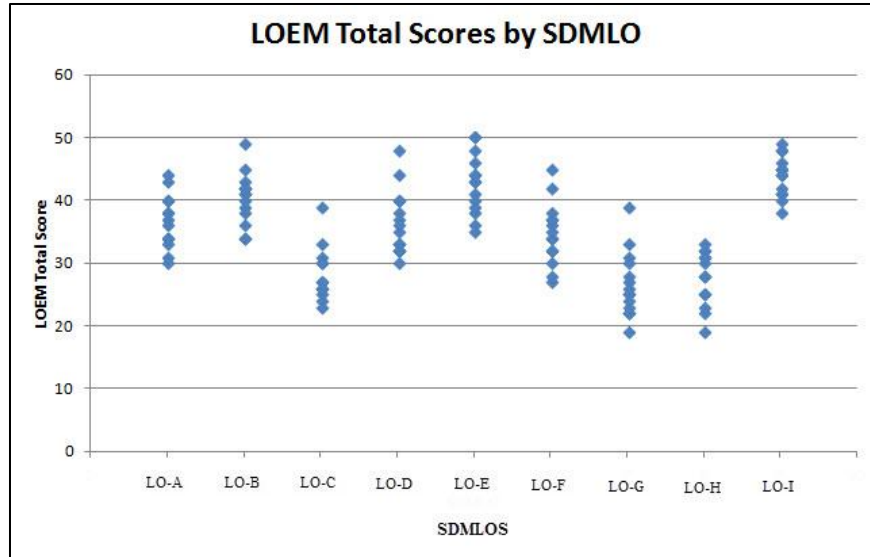


Figure 11: The LOEM total scores by SDMLLO.

Learning Object	Observed Scores	Mean (SD)	Ha: $\mu \neq 35$
A	30 – 44	36.5 (4.4)	t=1.23
B	34 – 49	40.3 (4.3)	t=4.44***
C	23 – 39	28.2 (4.3)	t=-5.70***
D	30 – 48	35.8 (4.9)	t=0.59
E	35 – 50	42.1 (4.5)	t=5.69***
F	27 – 45	34.2 (4.7)	t=-0.61
G	19 – 33	25.8 (4.0)	t=-8.29***
H	19 – 33	27.3 (4.4)	t=-6.31***
I	38 – 49	43.7 (3.2)	t=9.80***

Table 6: Descriptive Statistics for Total LOEM Scores by LO.

The t-test results allowed me to organize the LOs into the categories: “above average,” “average,” and “below average.” Figure 12 displays the mean total scores for each SDMLLO ranked from highest to lowest and color coded according to the LO scores results of the t-tests; yellow indicates the three LOs with mean total scores significantly higher than 35 (above average), orange for the three LOs that did not have significantly different mean total scores (average), and red for the remaining LOs that had mean total scores significantly lower than 35 (below average).

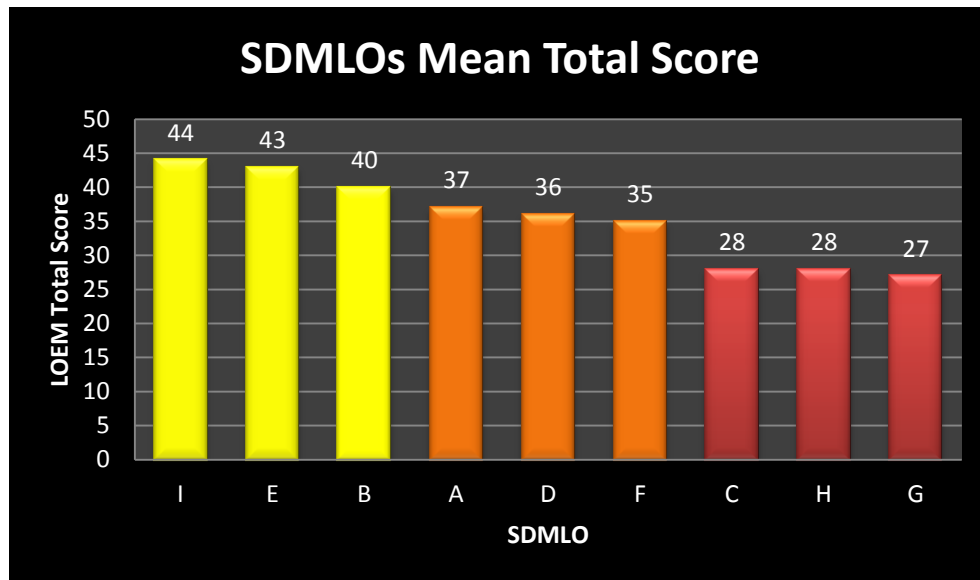


Figure 12: SDMLO Average Total Scores.

In efforts to corroborate the previous results, each evaluator's total scores for each LO were ranked, with 1 being the highest score and 9 being the lowest; any equivalent total scores were given the same rank. Table 7 displays the ranks for the LOs and the average rank across the evaluators. Only ranks of scores from the evaluators who completed the LOEM for all nine LOs were included. The ranked total scores for 5 of the 12 evaluators (42%) matched the top three scores, in the same order, as identified by ranked total average score. Over half (58%) of the evaluator's ranked LO-I the highest quality and exactly half (50%) of the evaluators scores ranked LO-E and B in either second or third place.

Learning Object	Evaluator												Average Rank
	2	3	5	6	7	8	9	10	11	12	13	14	
LO-I	3	2	1	1	1	1	2	1	1	4	1	3	1.8
LO-E	4	3	2	2	2	2	5	3	2	1	2	1	2.4
LO-B	5	1	3	3	3	3	1	1	4	3	3	4	2.8
LO-A	6	3	4	5	4	6	3	3	3	1	4	2	3.7
LO-D	1	6	5	3	5	4	4	4	6	7	5	6	4.7
LO-F	5	6	5	4	5	6	5	6	6	5	7	2	5.2
LO-C	8	5	7	7	7	7	8	7	9	7	7	9	7.3
LO-H	6	8	8	8	8	9	9	8	7	5	8	7	7.6
LO-G	9	8	9	8	9	8	7	9	8	6	9	8	8.2
* The ranks for Evaluators 1 and 4 were excluded because they did not evaluate all nine LOs.													

Table 7: Ranks by Evaluator for the Learning Objects with the above average LOEM scores.

The SDMLOs' average Total scores sorted from highest to lowest correspond exactly with their average rank, which indicates consistency between the two measures of quality for the LOs. Together these results provided the findings I used to answer the research questions that guided this evaluation. The discussion of the research questions and implications of these results for statistics teachers and educational technology developers, including designers and evaluators follows.

5. DISCUSSION

The purpose of this evaluation was to inform the statistics education community, including researchers, educators, and instructional technology developers, about the quality of existing SDMLOs. Despite less than ideal inter-rater reliability results this study's data provide an initial look at heuristic evaluation of LOs for one introductory statistics topic (sampling distribution of a sample mean) and answer a call for research in statistics education focusing on specific types of technologies (Schenker, 2007). In this section, I discuss the overall quality of existing SDMLOs, a comparison of the features of highest and the lowest ranked SDMLOs, and the implications of these results for statistics teachers and online educational technology developers.

5.1 Quality and Comparison of Existing SDMLOs

To investigate the overall quality of existing SDMLOs I computed the range of observed scores for the Interactivity and the Total Score components. The observed ranges of scores for the nine SDMLOs spanned the entire range of possible scores for both the Interactivity and the Total Score components, which may imply that these LOs have different levels of quality. The results corroborated my prediction that the SDMLOs would obtain a wide range of scores based on the LOEM instrument. Assuming the LOEM is measuring quality of LOs according to the intent of its design, the similarities between the average scores for the components and the Total Score from this study and those from the LOEM validation study may imply that the existing SDMLOs are on average approximately the same quality as other LOs for science and mathematics topics such as those Kay and Knaack evaluated.

For the purpose of analyzing the connections between the features of SDMLOs that received similar scores, I first focused on the results from the Interactivity component. The three LOEM items that measure the Interactivity component require scores for the users' potential for meaningful interactions, overall control, and the learning value of the multimedia tool. Therefore, the Interactivity scores for any LO evaluated with the LOEM should correspond to the interaction opportunities the LO provides (e.g., input fields, buttons), the sophistication level of the interaction options (e.g. text entry (low), buttons (medium), sliders/direct control over graphs (high)), and the user's level of control. The five SDMLOs (A, B, E, F, and I) that had median values in the upper half of the range of possible Interactivity scores, 7 or 8, are the only LOs that had interactive features with sophistication that allows users to directly impact the shape of either the population or the sampling distribution graphs in the display. Being able to directly manipulate a graph gives students immediate visual feedback that can help students develop concrete ideas of abstract topics such as infinitely possible populations or random samples (Chance & Rossman, 2006). In addition, LO-E includes an option to animate the sampling, further illustrating the abstract process of repeated random sampling. While LOs D, C,

and H have at least one feature that gives users control (e.g., beginning the simulation), they do not have any interactive features more complex than selecting items from given options or inputting values and clicking a button to begin the simulation. Not surprisingly, evaluators scored LO-G the lowest in the Interactivity component (median = 3; the lowest possible Interactivity score) because the only way to interact with the applet is to refresh the web page.

In addition to Interactivity scores, I analyzed the average Total Scores and compared features between the SDMLOs that evaluators scored the highest. This study identified LOs I, E, and B, respectively, as the highest quality SDMLOs using the LOEM evaluation criteria. Each of these LOs has very similar interface designs including vertically arranged plots with an interactive graph representing the population at the top, input options to the right of the plots, descriptive statistics to the left, and all of their display designs are nearly identical (e.g. yellow background, font, population and sample plots). The main difference found in the top three LOs, as scored in the current evaluation, is that LO-B only displays results of one sample at a time where the other two display the population, the samples (if the correct options are selected), and a graph representing the sampling distribution for the sample mean. This may show evidence counter to my initial prediction that the LOs that only display random samples will score lower than those that show a sampling distribution. Total Score results for each LO led me to conclude that LOs I, E, and B are above average in the existing SDMLOs, which may indicate that these tools have features that educators perceive as higher quality than other SDMLOs.

5.2 Implications for Practice

The results of this evaluation can inform practice in multiple stakeholders statistics education, including teachers, researchers, and LO developers (designers and evaluators). Teachers can utilize information about the top scoring LOs to help make their selections between SDMLOs. As a result of these LOs' classification as above, at, or below average, researchers have the basis from which to outline research questions that investigate a relationship between the quality level of the LOs and students' learning. In addition, now that "above average" tools have possibly been identified it may be very beneficial for statistics education researchers to investigate the difference between student performance using static representations to illustrate the sampling distributions and when using an "above average" SDMLO. Designers can contrast the features between the highest and lowest scoring LOs to consider which features they should incorporate into their designs. Evaluators of LOs can potentially use this study's results to inform future design and evaluation of LOs. For example, the weak reliability results for most of the components were major barriers to data analysis and should be a top consideration in any future work evaluating LOs with the LOEM (or any instrument). Implementing a training component to this evaluation to ensure that the evaluators are scoring consistently according to the LOEM criteria may improve inter-rater reliability results.

5.3 Limitations and Future Work

For this evaluation, I did not observe or interview the evaluators in this study and therefore, no further information beyond the LOEM data collected was available. To improve on my methods, in my future evaluations I would like to incorporate an interview component to help obtain more in-depth data from the evaluators, such as their opinions and comments about the LOs beyond

what is measured on the LOEM. Interviews or focus groups could also provide information on the strengths and weaknesses of LO design quality and LOs' effectiveness as educational tools that LOEM scores alone cannot. Using an iterative development process may bring additional insight into the quality of LOs beyond a single set of scores on the LOEM instrument.

An additional procedure I would recommend for future evaluations of LOs is to add student evaluators, which would provide additional data from an end-user perspective. Student evaluator scores could also be compared with those of teacher experts. Including both statistics teachers and students evaluators is recommended in future evaluations to gain insight from students' scores of existing LOs and to investigate similarities and differences between scores from the two types of evaluators.

The "above average" LOs found in this study had very similar layouts and interactivity features. LO designers would benefit from knowing whether the impact of layout or interactivity was more influential in promoting student learning or if they are equally important to LO design. A study focusing on the association between LOEM scores and the specific features of LO design that lead to high scoring may help designers distinguish which design elements are most important. Potentially, an SMDLO with similar interactive features as the "above average" LOs in this study but with a different layout could be compared to the top scoring LOs. Additionally, LO selection criteria for a future evaluation of LOs, possibly for a different topic in statistics, could focus on acquiring LOs with different layouts and interactive features and the evaluation results for those would serve as a comparison to those in this study.

6. CONCLUSION

The process of evaluation has the potential to greatly expand the research on educational technologies in statistics beyond the descriptions of technologies available and implementation methods that have provided the foundation for this study. Through the evaluation of these SDMLs I have provided the statistics education community with first-of-its-kind information about the quality of this existing set of tools available online. The results of this study begin discourse in statistics education research about evaluating existing LOs that is currently not prevalent in the literature.

Evaluation of all types of statistics LOs should continue in order to provide teachers with knowledge of which of the existing statistics LOs are of the highest quality. The results of this and future evaluations could potentially guide the creation of general design standards for statistics technologies and also provide a foundation for learning studies focusing on different LO characteristics. This line of work will ultimately provide statistics educators with valuable information about which LOs most effectively enhance student learning.

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.
- Akpınar, Y. (2008). Validation of a learning object review instrument: relationship between ratings of learning objects and actual learning outcomes. *International Journal of E-learning and Learning Objects*, 4(1), 91-302.
- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, R., and Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. Retrieved from <http://www.amstat.org/education/gaise/>
- Baki, A., and Çakıroğlu, Ü. (2010). Learning objects in high school mathematics classrooms: Implementation and evaluation. *Computers and Education*, 55(4), 1459-1469.
- Ben-Zvi, D. (2007). Using wiki to promote collaborative learning in statistics education. *Technology Innovations in Statistics Education*, 1(1).
- Blejec, A. (2003). *Teaching statistics by using simulations on the Internet*. In Proceedings of the IASE Satellite Conference on Statistics Education and the Internet, Berlin, Germany.
- Butson, R. (2003). Learning objects: Weapons of mass instruction. *British Journal of Educational Technology*, 34(5), 667-669.
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education (TISE)*, 1(1).
- Chance, B., del Mas, R., & Garfield, J. (2004). Reasoning about Sampling Distributions. In *The challenge of developing statistical literacy, reasoning and thinking*, 295-323, Springer Netherlands.
- Chance, B., & Rossman, A. (2006). Using simulation to teach and learn statistics. *ICOTS-7.4*
- Churchill, D. (2007). Towards a useful classification of learning objects. *Educational Technology Research and Development*, 55(5), 479-497.
- Cronbach L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

- Darius, P., Ottoy, J-P., Thas, O., Michiels, S., Raeymaekers, B. (2002). *Applets for experimenting with statistical concepts*. Proceedings of the 6th International Conference on Teaching Statistics, South Africa.
- DePaolo, C. A. (2010). The stat-attic website: links to statistics applets for introductory courses. *Journal of Statistics Education*, 18(3). Retrieved from <http://www.amstat.org/publications/jse/v18n3/depaolo.pdf>
- Dinov, I., & Christou, N. (2009). Statistics online computational resource for education. *Teaching Statistics*, 31(2), 49-51.
- Dinov, I., Christou, N., & Sanchez, J. (2008). Central limit theorem: New SOCR applet and demonstration activity. *Journal of Statistics Education*, 16(2). Retrieved from <http://www.amstat.org/publications/jse/v16n2/dinov.html>
- Dinov, I., and Sanchez, J. (2006). Assessment of the pedagogical utilization of the statistics online computational resource in introductory probability courses: A quasi-experiment. Proceedings of the 7th International Conference on Teaching Statistics, Salvador, Bahia, Brazil.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., and Scheaffer, R. (2007). Guidelines for assessment and instruction in statistics education (GAISE) report. *Alexandria: American Statistical Association*. Retrieved from <http://www.amstat.org/education/gaise/>
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Gliem, J.A., & Gliem, R.R. (2003). Calculating, interpreting, and reporting chronbach's alpha reliability coefficient for likert-type scales. Proceedings of the Midwest Research to Practice Conference in Adult, Continuing, and Community Education, Columbus.
- Harvey, B. (2005). Learning objects and instructional design. *The International Review of Research in Open and Distance Learning*, 6(2). Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/227/310>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1), 77-89.
- Hodgson, T., & Burke, M. (2000). On simulation and the teaching of statistics. *Teaching Statistics*, 22(3).
- Kay, R. H., & Knaack, L. (2008). An examination of the impact of learning objects in secondary school. *Journal of Computer Assisted Learning*, 24(6), 447-461.

Lane, D. (1999). Rice virtual lab in statistics. *Behavior Research Methods, Instruments, and Computers*, 31(1), 24-33.

Lane, D., & Scott, D. W. (2000). Simulations, case studies, and an online text: a web based resource for teaching statistics. *Metrika*, 51, 67-90.

Mittag, H. J. (2002). Java applets and multimedia catalogues for statistics education. In *Invited paper presented at the 6th International Conference on Teaching Statistics*, Cape Town, South Africa.

Mulekar, M. (2000). Internet resources for AP statistics teachers. *Journal of Statistics Education*, 8(2).

National Council of Teachers of Mathematics (NCTM). (2000). Principles and standards for school mathematics. Reston, VA: Author.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards-Mathematics*. Washington, DC: NGACBP/CCSSO.

Nesbit, J. C., & Li, J. (2004). Web-based tools for learning object evaluation. In *International conference on education and information systems: Technologies and Applications*, 21-25.

Ng, V. M., & Wong, K. Y. (1999). Using simulation on the Internet to teach statistics. *Mathematics Teacher*, 92(8).

Peterson, D. (2007). Usability theory, practice and evaluation for learning objects. *Learning Objects: Applications, Implications, and Future Directions*, 337-370.

Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of extension*, 37(2), 1-5.

Saporta, G. (1999). Teaching statistics with Internet: A survey of available resources and the St@tNet project. (Arias, E. & Manninen, A., Eds.) Proceedings of the 52nd Session of the ISI, Bulletin of the International Statistical Institute (Tome LVII, Book 2, pp. 233-236). Helsinki: International Statistical Institute.

Schenker, J. D. (2007). "The Effectiveness of Technology Use in Statistics Instruction in Higher Education: A Meta-Analysis Using Hierarchical Linear Modeling." PhD dissertation, Kent State University College, Ohio.

- Schneider, K. (2008). Two applets for teaching hypothesis testing. *Journal of Statistics Education*, 16(3). Retrieved from <http://www.amstat.org/publications/jse/v16n3/schneider.html>
- Schwarz, C. (2007). Computer-aided statistical instruction – multi-mediocre techno-trash. *International Statistical Review*, 75(3), 348-354.
- Schwarz, C., & Sutherland, J. (1997). An online workshop using a simple capture-recapture experiment to illustrate the concepts of a sampling distribution. *The Journal of Statistics Education*, 5(1). Retrieved from <http://www.amstat.org/publications/JSE/v5n1/schwarz.html>
- Tavakol, M., & Dennick, R. (2011). Making sense of cronbach's alpha. *International journal of medical education*, 2, 53-55.
- Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, 20(2), Retrieved from www.amstat.org/publications/jse/v20n2/tishkovskaya.pdf .
- Vargo, J., Nesbit, J. C., Belfer, K., & Archambault, A. (2003). Learning object evaluation: computer-mediated collaboration and inter-rater reliability. *International Journal of Computers and Applications*, 25(3), 198-205.
- West, R. W., & Ogden, R. T. (1998). Interactive demonstrations for stat ed on the world wide web. *Journal of Statistics Education*, 6(3). Retrieved from <http://www.amstat.org/publications/jse/v6n3/west.html>
- Wiley, D A (2000) [accessed 17 February 2013] Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy, in *The instructional use of learning objects: Online version*, Retrieved from <http://reusability.org/read/chapters/wiley.doc>
- Yacovelli, S. R. (2003). Understanding learning objects: the basic" chunks". *College & University Media Review*, 10(1), 17-26.

APPENDIX

Sources Searched in Literature Review

- JSTOR
- ERIC
- Google Scholar
- Journal of Statistics Education (JSE)
- Technology Innovations in Statistics Education (TISE)
- The American Statistician (TAS)
- Statistics Education Research Journal (SERJ)
- Journal of Research in Mathematics Education (JRME)

LOs Selected for Evaluation

1. <http://www.causeweb.org/repository/statjava/CLT2Applet.html>
2. <http://www.stat.tamu.edu/~west/ph/popsample.html>
3. <http://kerbaugh.uncfsu.edu/stats/barrel/barrel.html>
4. http://www.chem.uoa.gr/applets/AppletCentralLimit/Apl_CentralLimit2.html
5. http://onlinestatbook.com/stat_sim/sampling_dist/index.html
6. <http://www.intuitor.com/statistics/CentralLim.html>
7. <http://www.applet-magic.com/samplemean.htm>
8. <http://vassarstats.net/m1samp.html>
9. <http://www.stat.tamu.edu/~west/ph/sampledistrib.html>

LOEM URL: <http://faculty.uoit.ca/kay/papers/AppendixB.html>

Scoring Spreadsheet

Applet Link:

Interactivity:

Score

Meaningful Interaction
Overall Control
Multimedia adds Learning
Value

Design:

Score

Consistency
Layout
Labeling
Readability

Engagement:

Score

Quality of Feedback
Attractive
Graphics (not video)
Learning Mode
Motivation

Usability:

Score

Natural to Use
Orientation
Navigation Cues
Instructions
Appropriate Language Level

CHAPTER 4. USABILITY OF A CONFIDENCE INTERVAL APPLET AS EVALUATED BY COLLEGE STUDENTS

A paper to be submitted to the *Journal of Statistics Education*

Rachel Graham, Iowa State University

Key Words: applets, usability testing, confidence intervals

Abstract

It is important for developers and users of statistics education technologies to evaluate existing tools using feedback from the end-users, students. Usability testing has been shown to increase the effectiveness of online learning tools but is relatively new to statistics education research. This study investigated the usability of the “Simulating Confidence Intervals Applet” from the Rossman/Chance applet collection using end-users (students) as the evaluators. The results of the usability test identified strengths and weaknesses of the applet. I also developed a list of recommendations for the improvement of the applet in the next stages of design. The recommendations for improving future designs of the applet have the potential to improve the usability and effectiveness of the applet design.

1. Introduction

The dramatic increase of Internet access in schools across the United States over the past two decades (NCES, 2010) has led to increased development of online digital resources that support specific learning objectives in both face-to-face and distance education classrooms (Goodyear & Retalis, 2010; Schenker, 2007). Educational technologies that are reusable digital resources that teachers can use to support a learning objective are known as learning objects (LOs) (Akpınar 2008; Baki & Cakiroglu 2010; Harvey 2005; Peterson 2007; Wiley 2000; Yacovelli 2003). The main purpose of any LO is to assist the user(s) (e.g., anyone involved in the interaction between the LO and students) in progressing toward proficiency in a specific learning objective (Kay & Knaack 2008; Wiley 2000). If they search online, teachers will find hundreds of free LOs that can help them teach most topics in high school and postsecondary introductory statistics courses (Schenker 2007; Tishkovskaya & Lancaster 2012), including descriptive statistics and graphs, basic probability, discrete and continuous distributions, and one- and two-sample inference for means and proportions. While there are many options of LOs available, the statistics education research lacks studies designed to evaluate and improve these tools. This study seeks to address the gap in the literature by investigating the quality of a single online learning object that can help teach one-sample confidence intervals.

The most common type of LO in statistics education is the applet: a type of web-based software application (Chance et al. 2007; Hsu 2003; Schenker 2007). The *Guidelines for Student Assessment and Instruction in Statistics Education (GAISE) Report: College* (Aliaga et al. 2007) suggest applets as a technology instructors can use when teaching statistics concepts. Applets have the potential to help educators effectively teach multiple topics and methods in statistics. Some applets allow students to automate calculations and graphing (Al-Aziz et al. 2010; Che et

al. 2009), while others simulate abstract concepts, such as repeated sampling (Chance & Rossman 2006; Lane & Scott 2000). Complex topics in inferential statistics, such as estimation of a population parameter using a confidence interval, require students to understand multiple abstract concepts, and applets can provide students with a series of images or animations that are superior to static representations. If educational technology developers, including designers and evaluators, want to maximize a technology's potential as a teaching tool and, therefore, provide statistics teachers with the highest quality resources possible, it is essential that they focus their design objectives on students: the end users of their products.

Usability is a foundational principle of the field of human-computer interaction (HCI), which studies the design and evaluation of technology with a focus on the user and context (Berg, 2000). The International Organization for Standardization (ISO) defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” (ISO 9241-11). Usability testing is an effective method to evaluate the interactive features of any technology (Barnum 2011; Tullis & Albert 2013; Peterson 2007). In the 1990s, Jakob Nielsen and Tom Landauer (1993) revolutionized usability testing by showing effective usability tests can be completed with small numbers of evaluators outside of a controlled lab setting. Usability testing is an evaluation technique that allows designers to incorporate the opinions of end users into the development of the products they are creating, which is important in user-centered design.

The usability of a technology is inherently linked to how well it serves its intended purpose; therefore, increasing the usability of an LO should increase its potential as an effective teaching tool. Any interaction between students and technologies creates a situation where evaluating usability is possible. When developers design LOs and when teachers are selecting LOs to use in their classrooms, usability should be considered (Peterson 2007; Reeves 2008; Wiley 2000). Developers should use usability testing to gain insight into the most effective design aspects of a learning tool. However, the statistics education literature lacks evidence that usability testing methods are present in the design processes of statistics LOs. I designed the present study to introduce user-centered design into the development of statistics LOs by describing the methods and outcomes of a usability test designed to improve the quality of an existing applet that can help teach one-sample confidence intervals.

2. Previous Work Evaluating LOs

To determine the methods of usability testing most appropriate for statistics LOs, it is useful to review the methods educators have used when they implement online media in their teaching practices. Descriptive information obtained from usability tests from previously completed research can provide guidelines for designing effective usability evaluations in new areas. Little work has been published describing usability tests of LOs, and I found no studies involving usability tests of statistics LOs. In response, I widened the scope of the literature review to include studies from any of the STEM fields, a wider classification of science, technology, engineering, and mathematics education disciplines, which includes statistics; this search yielded only three studies. The reviewed papers, summarized below, included reports of usability testing focused on improving the interface design of a course management platform in a high school and usability testing of augmented reality applications.

In their study focused on interface design, Wang and Yang (2005) performed a usability test of an online learning environment using high school students studying fossilization in a science class. Seven sophomore students answered a usability questionnaire after spending 30 minutes independently exploring the environment. Researchers recorded observations about student behaviors and interactions as they explored the environment and interviewed students to collect their opinions of the environment after they had completed the independent exploration and the usability questionnaire. Results showed that the students had an overall positive reaction to the activities and also identified aspects of the interface that could use improvement. The researchers used the results to create a list of design recommendations for instructional software interfaces, including adherence to usability and accessibility standards, keeping information accessible to users, incorporating visually oriented and realistic objects, and conducting usability testing. This research demonstrates a qualitative approach to usability testing, using a questionnaire, observations, and interviews.

Additionally, Kaufmann and Dünser (2007) reported on a series of three evaluations of an augmented reality application (Construct3D) which displays three-dimensional geometric objects to support student explorations in dynamic 3D geometry. For their first informal evaluation, researchers collected feedback from fourteen Austrian high school students who completed a task using Construct 3D and observed the students completing the assigned tasks (Kaufmann et al. 2000). The results led researchers to add features to Construct 3D; the new features included allowing objects to be displayed with partial transparency. More applications, such as visualizing hyperbolic revolutions, were also designed into the program to provide users with more avenues to utilize the Construct 3D program.

Next, Kaufmann and Dünser (2007) described a usability test involving fifteen geometry students who, using Construct 3D, completed practice problems with their instructors over 30 hours (Kaufmann & Schmalstieg 2003). Both teachers and students completed a usability evaluation questionnaire after their interaction with Construct3D. After analyzing the data collected from the questionnaire, the research team changed the labeling of the features of Construct 3D, added a help-box to explain all of the menu items, and restructured the menu system to make frequently used features more easily accessible.

For their third evaluation of Construct 3D, Kaufmann and Dünser (2007) compared usability questionnaire scores from 47 students who used the most current version of Construct3D to 44 students using a three dimensional computer-aided drafting (CAD) program commonly used in Austrian high schools. Both groups of students completed the same geometry problems with a tutor available as the students worked with Construct 3D. Students scored the Construct3D LO significantly higher than the CAD program on all elements in the evaluation except technical aspects such as error tolerance. The conclusions of the study were that the iterative testing process improved the quality of Construct 3D as a learning tool for geometry by increasing its usability (Kaufmann & Papp 2006). The results of these studies imply that iterative design processes may produce better learning tools.

In another relevant usability study, Iordache and Pribeanu (2009), completed end-user evaluations of another augmented reality system they called the Augmented Reality Teaching

Platform (ARTP), a desktop screen that users look through to see images superimposed on the desktop underneath. The test included results from 42 students in Bucharest who went with their teachers to two sessions with the researchers. During these sessions the students completed one demonstration lesson and two exercises utilizing ARTP. They collected measures of efficiency (time on task) and effectiveness (rate of completion) data from user log files. After completing the exercises using ARTP, students were asked to give their opinions of ARTP's three most positive and three most negative aspects. Two usability experts who also completed the activities filled out a usability questionnaire. The researchers grouped results of the expert evaluations into specific usability problems and compared them to the negative aspects of ARTP that students identified in their responses to the usability questionnaire. After comparing student and expert responses, the design team was able to create a list of changes that would possibly improve the usability of ARTP. Changes included having the students use different eye equipment when interacting with the tool to reduce eye strain.

The research methodology in the above studies, similar to what I used in the research presented below, illustrates the process of interpreting usability data to offer design guidelines. Reports on usability testing can inform teachers about which aspects of educational technologies are most important when selecting tools to use in the classroom. Usability testing results also provide evaluators and designers with guidelines by which they can develop new or update existing tools. I chose usability testing for this study's evaluation because the purpose was to inform decisions in the next stage of development, which is similar to the studies I found in the literature.

3. Method

In this study, I facilitated a usability test to investigate the usability of the Rossman/Chance Simulating Confidence Intervals Applet. Postsecondary students who had recently completed an introductory statistics course were the evaluators for this usability test. For the remainder of this paper, the Simulating Confidence Intervals applet will be referred to as "the applet." I designed the usability test in this study to accomplish the following goals: 1) identify the applet's strengths and weaknesses including barriers to completion of tasks designed around the applet, and 2) provide recommendations for improving the usability of the applet.

3.1 The Applet

Rossman/Chance applets are a collection of 35 freely accessible LOs that illustrate a variety of statistics topics. The collection is organized by topic: data analysis, mathematical models, probability and inference, randomization distribution simulations, and sampling distribution simulations. Revisions and newer versions of many of the original applets are also available in the collection. Developers of the Rossman/Chance collection ask for feedback on their website, providing further evidence that they are open to continuing development of the tools they offer. The applets in the Rossman/Chance collection are popular in statistics education; thus, facilitating further stages in the development of any of their tools may have great potential to impact a large audience, making these LOs attractive choices to study.

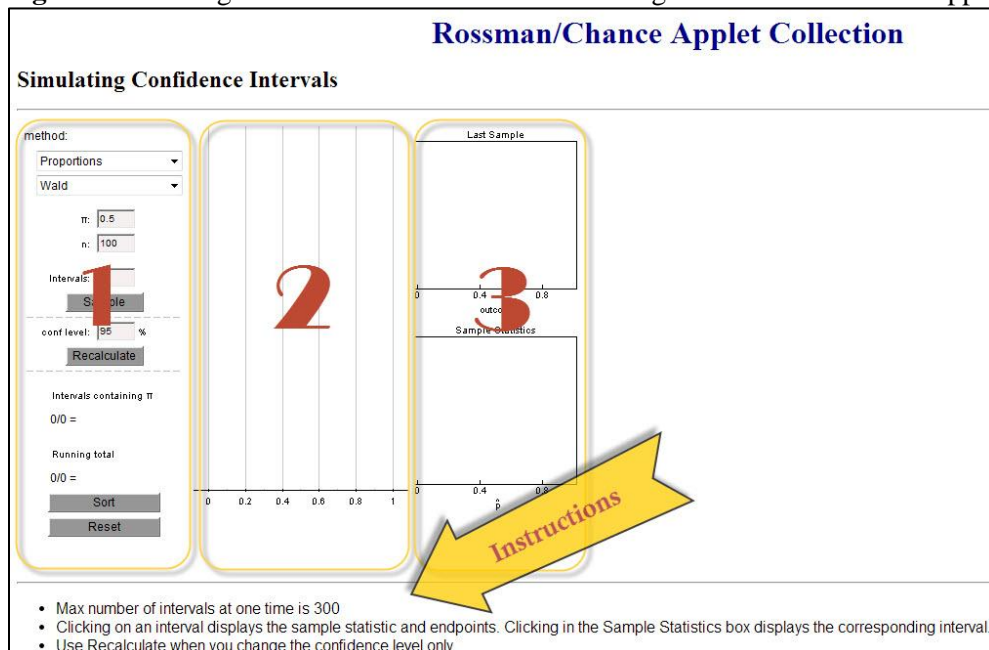
In selecting an applet to use for the study, I first narrowed my focus to topics in inferential statistics, an area of the field with which statistics students often struggle. Ultimately, I chose the

‘Simulating Confidence Intervals’ applet

(<http://www.rossmanchance.com/applets/NewConfsim/Confsim.html>) as the learning object to evaluate in this study because estimating parameters using confidence intervals is a concept in inferential statistics many students often find difficult. Visualization of multiple intervals at one time, a feature of the ‘Simulating Confidence Intervals’ applet, helps students understand how confidence level relates to parameter estimation. I asked permission to complete the usability test on the confidence interval applet in the Rossman/Chance collection from Dr. Beth Chance, which she granted via email in March of 2013.

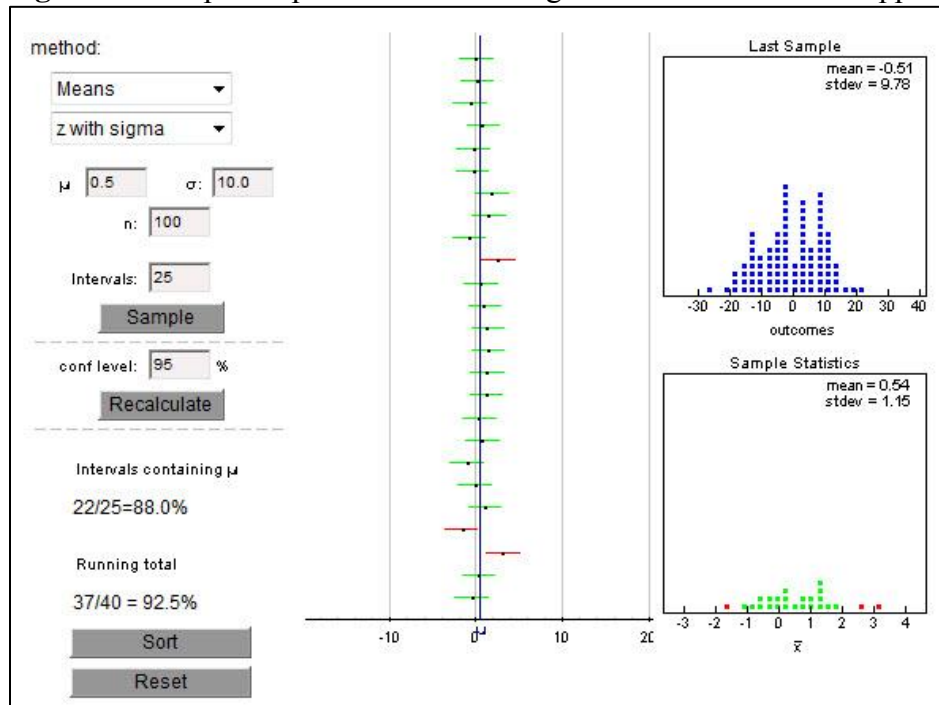
The confidence interval applet has three panels (see Figure 1). The panel on the left (Panel 1), is where users select the parameter they will estimate (mean or proportion), the interval type, population values, sample size, number of intervals to calculate, and confidence level. When a user clicks the “Sample” button in Panel 1, the applet draws random samples from a population with the specifications users have entered (or the default values); it then displays lines representing the confidence intervals in Panel 2: the intervals that do include the population mean are displayed in green while those that do not are displayed in red. The user can continue simulating intervals from the same population by clicking the sample button multiple times without changing any of the specifications. Panel 1 also includes the percentages of intervals containing the parameter for the most recent samples and intervals constructed and a running total percentage to track previous samples taken from the same population but not shown on the screen. After simulating intervals, a user can choose to recalculate the intervals using a different confidence level by selecting the “Recalculate” button. The “Sort” button will sort the intervals shown by their point estimate value, indicated by a black dot in the middle of each interval. The last button in Panel 1, “Reset,” clears the applet.

Figure 1 Main Page of the Rossman/Chance Simulating Confidence Intervals Applet



The panel on the right-most side of the screen (Panel 3) contains two dot plots. The top chart draws a dot plot of the last sample of n and the bottom dot plot displays only the sample statistics from the current random sample. Instructions for the applet load below the three panels of the applet, giving students information about the maximum number of intervals they can create, the interactive features in the graphs, and when to use the “Recalculate” button. All three panels contain interactive capabilities, and together they simulate the process of drawing repeated samples and constructing confidence intervals for means or proportions from each simulated sample.

To illustrate the applet’s capabilities, I calculated twenty-five 95% confidence intervals for a mean from a population with $\mu = 0.5$ and $\sigma = 10$. The applet constructed each interval from 25 simulated random samples of 100. Previously, I had constructed 15 intervals to better show the cumulative features of the applet. Figure 2 shows the output the applet returned for this random trial. At the top of Panel 1, the output shows the method I used (e.g. parameters, number of intervals constructed, and confidence level), the proportions and percentages of intervals containing the population parameter for the current and cumulative samples. Eighty-eight percent of the 25 intervals I constructed contained the parameter ($\mu = 0.5$), meaning that in this set of intervals, there were three intervals that did not include the population parameter. The running total proportion of intervals containing the population parameter (37/40) indicates the 15 samples I calculated prior to the current 25 displayed. Panel 2 includes a visual representation of the 25 intervals. The green intervals in Panel 2 correspond to the 22 intervals that did contain μ while the red intervals represent those that did not. In Panel 3, the plot titled “Last Sample” shows the random sample ($n = 100$) from which the applet calculated the bottom confidence interval (in Panel 2), and the “Sample Statistics” plot displays the sample means from each of the 25 random samples. The green and red data points in the “Sample Statistics” plot correspond to the point estimates from the intervals (in the same color) in the center panel.

Figure 2 Example output from ‘Simulating Confidence Intervals’ Applet

Features of the ‘Simulating Confidence Intervals’ applet illustrate many of the concepts involved in understanding the estimation of a population parameter with a confidence interval. The applet simulates intervals for both means and proportions, the two parameters introductory statistics classes emphasize about one-sample estimation. Additionally, students can visualize the effect of confidence level and sample size on the width of a confidence interval using the applet. Plots showing distributions of the last individual sample drawn and the summary statistics in the applet further allow students to explore the relationships between the sample and the population.

3.2 Participants

Student evaluators in this study included a voluntary response sample of university students from a small liberal arts university in the Midwest. All of the students I invited to participate had completed their most recent statistics course within one calendar year prior to the usability test. The university’s registrar provided a list of email addresses for all of the eligible students as I outlined in my IRB application (approved through both the university where the usability testing took place and the university where I attend) and I then invited (via email) all students who met the criteria to participate. Because single surveys of large populations do not give the intense feedback I needed, I focused on a smaller group of 5-8 participants to test the critical usability issues of the applet (Nielsen & Landauer 1993).

3.3 Testing Procedures

After students responded to the invitation and consented to participate, I scheduled one-on-one sessions with each of them. I designed the usability test in this study so that it would take no

more than an hour to complete; tests that run longer risk student evaluators losing their concentration (Barnum 2011). Student evaluators completed each of the sessions in a computer lab on the campus of the university all of the students attended. The evaluators' mouse movements and audio were recorded through the screen capture system Panopto.

During the usability tests, each student evaluator completed three tasks: an independent exploration with a questionnaire, an activity designed around the features of the applet, and a follow-up questionnaire. First, student evaluators took as much time as they wanted for an independent exploration of the applet. Through this exploration student evaluators were to get familiar with the interface and features without any outside interference. When they were finished exploring the product, student evaluators completed a questionnaire on their initial impressions and their previous exposure to applets in statistics classes. The second task in each session comprised a series of steps and questions that guided student evaluators through using the features of the applet, and I asked them to "think aloud" as they worked through it. Usability experts recommend think-aloud protocols as an effective way to identify what users are thinking and any features they find confusing or unclear (Tullis & Albert 2013). I gave a short introduction to the think-aloud protocol before students began the activity in the second task, including a short example of thinking aloud while I worked through a subtraction problem. Throughout the activity, I prompted them for their thoughts if they paused while working on the tasks. I also answered any questions they had, and provided assistance if at any point the participant indicated that he or she could not proceed. Student evaluators recorded their answers to the activity questions on a worksheet I provided. The third task, a follow-up questionnaire (see appendix) asked students for their overall impressions, their opinions on the strengths and weaknesses of the applet, any barriers they encountered while trying to complete the tasks, and recommendations to improve the design of the applet or its effectiveness as a learning tool. Students completed the questionnaires independently.

3.4 Data Collection

For this usability test, I used four instruments to collect written data from student evaluators. A demographic form gathered information about the student evaluators' age, gender, declared major, and class. The second instrument, an initial impressions questionnaire, asked students to give their initial impressions of the applet and if they had experience with any tool of this type. If students had prior experience with online statistics tools, I asked them to list the tools they had used and their purposes.

A confidence interval activity worksheet I designed to guide student evaluators through using the features of the applet was the third instrument (see Figure 3). The activity consisted of a worksheet with seventeen items: six items instructed the student evaluators to complete specific interactions with the applet and the other eleven items asked the student to report information that could be retrieved from output displayed on the screen after the interaction item had been completed. Two of the interaction items on the confidence interval activity instructed participants to construct a given number of confidence intervals and provided them with values that tied to each of the entry fields on the applet. Other activity items directed student evaluators through using features of the applet that required clicking the buttons: repeated sampling, sorting the intervals, and resetting the applet. The activity did not require participants to make any interpretations of the output.

Figure 3 Task 2 Confidence Interval (Task 2) Activity**Confidence Interval Activity**

Instructions: Complete the following tasks using the Rossman/Chance Applet for Confidence Intervals.

1. Set the population proportion to 47% and the sample size to 125. Construct five (5) 95% confidence intervals for proportions using the Adjusted Wald formula.
2. How many of the constructed intervals contain the population proportion?
3. What was the sample proportion for the last sample in the simulation?
4. What are the mean and the standard deviation for the five sample proportions drawn?
5. Without resetting the applet, continue sampling and constructing five (5) 95% confidence intervals at a time (using the same parameter and sample size) until you have a running total of 40 samples.
6. What percent of the 40 intervals constructed contained the population proportion?
7. Reset the applet.
8. Construct (35) 99% t-intervals for means from simulated samples of 20 drawn from a population with a mean of 75, a standard deviation of 3.
9. How many of the constructed intervals contain the population mean?
10. What are the mean and standard deviation for the last sample of 20?
11. What is the shape of the distribution of the 35 sample statistics?
12. What are the mean and the standard deviation of the 35 sample means drawn?
13. Sort the intervals.
14. By which values are the intervals sorted?
15. Without resetting the applet, continue sampling until you have a running total of 350 samples.
16. What percent of the 350 intervals constructed contained the population mean?
17. Does the distribution of sample statistics include all 350 samples? How do you know?

The fourth data collection instrument was a follow-up questionnaire. The follow-up questionnaire asked the student evaluators to give their overall impressions of the applet, the applet's strengths and weaknesses, any barriers they may have encountered while completing the tasks on the activity, and any recommendations they had to improve the design of the applet in order to make it a more effective learning tool.

In addition to the audio and video recordings of the sessions, throughout each session I made my own observations, including student evaluator questions, any technical difficulties or errors the students encountered. After each testing session, I used the videos to obtain data on the features students explored during the independent exploration, the time they spent on each task, and whether they could successfully complete the items on the second task. I also used the videos to compile the student evaluators' questions and their think-aloud dialogue.

3.5 Data Analysis

To analyze the demographic information for the student evaluators, I summarized the self-reported information, and recorded the responses to questionnaires as well as the Task 2 Activity in a spreadsheet. I also marked each response to the activity items as correct or incorrect. Next, I organized the student evaluator responses from the initial impressions and follow-up questionnaires into the categories: positive comments, negative comments, and suggestions for

improvement. I identified similar categories in the responses from student evaluators and I counted their frequencies.

I collected data from the videos of each session, including the completion time for each task and participant comments made through the think-aloud process. Minimum, maximum, and average completion times were computed for the independent exploration and the confidence interval activities for each participant. I also recorded whether the students needed assistance to complete any of the items on the Task 2 Activity.

4. Results

The main purposes of the usability test were to advance knowledge of the strengths and weaknesses of the applet and to offer suggestions for its increased usability. Six students, two male and four female, replied to my invitation email and completed the usability test. The mean age of the student evaluators was 26 years. Four student evaluators were nursing students, one was a biochemistry major, and one was studying mass communications. When asked to indicate their class standing, four selected “senior,” one selected “junior,” and one participant did not respond. In the following section, I report positive and negative comments the student evaluators gave, descriptive statistics for the Task 2 activity, and my observations (as moderator). The results are organized in the order that the student evaluators completed tasks.

4.1 Task 1: Independent Exploration and Initial Impressions

To begin the usability test, I gave unrestricted time to the student evaluators to explore the applet. Participants spent an average of 1.9 (SD =1.2) minutes on their independent explorations of the applet. The least amount of time spent in an individual exploration was 1.3 minutes and the greatest time spent exploring was 4.3 minutes. After the student evaluator indicated his/her exploration was complete, I provided the initial impressions questionnaire. All six student evaluators indicated they had prior experience using at least one applet in a statistics education context. Five evaluators indicated that they had previously used an applet to calculate p-values for chi-squared hypothesis tests. Half of the student evaluators stated that they had used the “Reece’s Pieces” applet (another Rossman/Chance applet that simulates the sampling distribution of a sample proportion); this was the only applet identified by name. Two of the student evaluators reported that they had used applets to create histograms, and one had investigated aspects of the normal distribution. In summary, all participants had at least minimal experience with applets in statistics contexts; however, none indicated that they had any experience with the applet being evaluated.

Overall, initial impressions of the applet were mixed. Five student evaluators noted that the applet was “easy to follow” or “straight forward.” One participant indicated that the graphs on the applet were a positive aspect of the LO. Negative comments (2) noted that the confidence interval section of the graph was confusing and that the font used in the applet was too small. During the independent explorations, I observed that student evaluators seemed to not really engage during the independent exploration time. Five of the students investigated one or two of the applet’s features and appeared to randomly enter data into a field or check the options in the

drop-down menus without studying the purpose of the applet. Many of them looked at the clock multiple times. One student carefully studied the applet and interacted with most of its features.

4.2 Task 2: Confidence Interval Activity

The second task consisted of the 17-item Confidence Interval Activity. The mean time it took for student evaluators to complete the second task was 12.7 minutes (SD = 2.2). All of the student evaluators completed the activity in less than 17 minutes (see Table 1). None of the students encountered technical difficulties with the applet as they completed the task.

Table 1 Time-on-Task for Confidence Interval Activity

Participant	Time-to-Complete: Task 2 (in minutes)
P1	16.80
P2	11.25
P3	10.62
P4	11.88
P5	12.43
P6	13.02

Six items on the Confidence Interval Activity worksheet contained directions to either input data into the entry fields or manipulate the buttons in the applet. In Table 2, I have listed each of the six item types, descriptions, and numbers, as well as the number of students who successfully completed each item and how many needed assistance. The first interval construction task required students to construct confidence intervals for a proportion. Every student who successfully completed Item 1 needed assistance entering the population parameter for the first interval construction task because they did not recognize the symbol for π as a symbol for a population proportion. One of the students asked, “Where is p ?” and four others asked for clarification on the meaning of the symbol π . When students asked, I let the student evaluators know that π was a symbol for the population proportion and that it was sometimes represented by just the letter p . After receiving this information, those five students entered in the correct information. One participant did not successfully complete the interval construction; however, this was because he/she entered the incorrect sample size, which was a minor input error. Another student needed assistance on the interval construction task for the mean and asked how to switch from proportions to means; the student was able to complete the task after being shown the drop-down box where the user selects the parameter that is to be estimated. Only one of the tasks requiring participants to use the applet’s buttons required intervention; one student asked for clarification on the “Sample” button before he/she proceeded. All of the students were able to complete the tasks requiring buttons correctly, and they asked for no other clarifications.

Table 2 Task 2 Feature Interaction Items

Item Type	Item Description	Item Number	Successful Completion ($n = 6$)	Assistance Needed
Interval Construction	Proportion	1	5	5
	Mean	8	6	1
Button	Repeated Sampling (p)	5	6	1
	Reset	7	6	0
	Sort	13	6	0
	Repeated Sampling (μ)	15	6	0

The remaining eleven items on the Confidence Interval Activity worksheet contained items that asked the students to retrieve information from the applet's output. I have summarized the results by item type, description, and number in Table 3. In this section, all six student evaluators answered seven of the eleven items correctly. However, at least half of the students answered items 2, 14, and 17 incorrectly in the activity.

Item 2 asked students to report the number of intervals that contained the population parameter. The applet displays the information students needed to complete the item in two places: on the bottom of Panel 1, it is displayed as the numerator of the proportion of intervals containing the parameter, and in Panel 2 it is the number of green intervals that contain the parameter. Only two students answered Item 2 correctly, and both asked questions about the meaning of the green intervals before answering. To complete Item 2, all of the students referred to the set of intervals in Panel 2 and did not give any indication that they used the proportion in panel 1 when they gave their answers.

While completing Item 14 of the activity, none of the students asked for assistance, but only one answered the question correctly and two said that they did not know. Item 14 asked the student evaluators what value the intervals were sorted by after they had pressed the sort button. The three student evaluators who did not successfully complete the task indicated that the intervals were sorted by the lower bound value of their interval. The last item on the activity (Item 17) asked students about the content of the "Sample Statistics" Graph in Panel 3. Half of the students answered the item correctly; each of them counted the number of dots in the graph to answer their question and did not ask for assistance. The other three students also did not ask for assistance, but they answered both quickly and incorrectly.

Table 3 Task 2 Output Items

Item Type	Item Description	Item Number	Successful Completion ($n = 6$)	Needed Assistance
Intervals include parameter	Number of intervals containing p	2	2	4
	Percentage of intervals containing p	6	6	3
	Number of intervals containing μ	9	6	1
	Percentage of intervals containing μ	16	6	0
Last sample	Report \hat{p}	3	6	3
	Report mean and SD of \bar{x} sample	10	5	2
Sample statistics	Report mean and SD (\hat{p})	4	6	2
	Shape of the distribution	11	6	2
	Report mean and SD (\bar{x})	12	6	0
	Cumulative distribution	17	3	0
Interval components	How intervals are sorted	14	1	0

As the moderator, I observed that while the students seemed to answer Item 2 using Panel 2, they all answered Item 9 (identical to Item 2) using the proportions in Panel 1. This change is possibly a result of Item 6 which required them to collect information only found in the section with the proportions (in Panel 1). I also noted that students did not take much time to answer Item 14 and those who answered incorrectly were confident of their answers. None of the student evaluators asked for clarification after Item 12; the students answered the questions with more confidence after the first time they used the features in the earlier questions. All but one of the students recorded their responses to the last three items very quickly. The students, in general, did not seem to utilize or notice the percentage output at the bottom of Panel 1 until their attention was called to it by an item later on in the activity. Also, the students rarely used the interactive features of the applet unless they were guided to do so during the activity.

The student evaluators' issues with confidence, content knowledge, and features of the applet became apparent through the think-aloud discussions. Every evaluator began the activity by checking if they were right after many of the first five items on the Task 2 Activity. For example, two of the students who answered Item 2 incorrectly, indicated through their think-aloud process that they thought the "dot in the middle" or the lower limits of the confidence intervals were the intervals. In addition, for these student evaluators, the think aloud protocol appeared to be the most difficult part about the usability test. Each student evaluator read most of the items on the activity out loud as their main participation in the think-aloud protocol. All of the student evaluators asked questions about features of the applet, such as "are the sample statistics here?" or to verify their answers were correct. After all of the tasks of the usability test were completed, three of the students mentioned that they had never done anything like that before.

4.3 Task 3: Follow-Up Questionnaire

Upon completion of the Task 2 activity, I asked the student evaluators to complete a follow-up questionnaire collecting their overall impressions about the applet as well as their ideas about the applet's strengths and weaknesses, barriers they encountered, and any recommendations for the design of the applet as a better learning tool. Students identified ease of use and components of the design as the strengths of the applet. Four of the six student evaluators indicated that the applet was easy to use or follow in response to the question about their overall impressions or feelings about the applet. Other strengths identified by the student evaluators involved visual aspects of the design, including the color coordination of the confidence intervals containing the population parameter, the simulation of the intervals constructed, the variety of the plots, and the interactivity between the plots displayed in Panel 2 and Panel 3.

Barriers to task completion students listed most frequently were inappropriate labeling, inadequate directions, and insufficient content knowledge. Every participant included at least one comment about either the labeling or instructions for the applet. One student wrote, "I had a hard time figuring out what everything was." Two-thirds of student evaluators indicated that they had forgotten what they had learned about confidence intervals, and that this was a barrier to their task completion. The number of labels, the font, the size of the font, and the symbol used for the population proportion all prevented evaluators from completing items. Evaluators' identification of weaknesses and recommendations for improvement coordinated well with the problems students faced while attempting to complete tasks. Every student evaluator indicated that the applet's use of the pi symbol for the population proportion confused them. Comments regarding weaknesses of the applet frequently included the need for more instructions. Suggestions specifically about the existing user interface called for more labeling and larger font size. Several students recommended that the applet could be "spread out" across the entire screen.

5. Discussion

This study's usability test of the "Simulating Confidence Intervals" applet identifies strengths and weaknesses of the design (including barriers to task completion) and provides recommendations for improving the usability of the applet. In this section, I first summarize the applet's strengths and weaknesses identified through the results of the usability test. Next, I give recommendations for the improvement of the applet in the next stages of design. Then, I discuss the general implications of the results for LO developers and statistics educators, and, finally, I report this study's limitations.

5.1 Strengths and Weaknesses

Several student evaluators identified "ease of use" as a positive feature of the applet in both the initial impressions and the follow-up questionnaire. If an LO is easy to use, this can help students feel comfortable during interaction with the tool. The students' comfort with the entry fields and buttons of the applet and little need for assistance when interacting with these features substantiate their feelings that the applet's interface was effective. One of the foundations of gestalt theory, directly applicable to design, is the Law of Prägnanz which explains how humans

naturally prefer objects that are simple and ordered (Wertheimer, 1938). Evaluators indicated that simplicity is an important positive aspect of LO design, which aligns with this principle. Giles Colborne (2011) stated that simplicity both reassures users that the product will be predictable and helps them to feel in control. Thus, in future development of the applet, one goal should be to maintain the level of simplicity found in the current version.

In addition, evaluators identified the strategic color use to connect different pieces of the display as a positive aspect of the design. Colors can help highlight related components of the applet and focus users' attention on connections in the content (Plaisant & Shneiderman 2005). Currently, the applet utilizes the color green to identify which intervals contain the population parameter and red for those that do not. This is consistent with the idea that humans will perceive information according to their past experiences (Wertheimer, 1938; Johnson, 2010). In this case, the green meaning "yes" and the red meaning "no" is consistent with common meanings and symbols attached to those colors in the U.S. today however, the use of green and red can potentially be a barrier for students who are color blind and cannot distinguish between these colors. Therefore, in future designs, the use of alternative colors to connect features of the applet may be more effective for students to recognize relationships between the samples and the intervals.

Color also connects the applet's intervals to the data point in the sampling distributions plot in Panel 3 that corresponds to its sample mean value. Students' positive reaction to the use of color to distinguish between the intervals corroborates existing design principles. As the development of this applet moves forward, I recommend that the designers continue and possibly increase strategic use of color in the LO design. In addition to the current features, the applet could include additional color coordination to connect the last sample drawn to its corresponding confidence interval and the μ input field in Panel 1 to the vertical line that represents μ in Panel 2.

The usability test in this study also revealed aspects of the interface I have identified as weaknesses of the applet or barriers to Task 2 completion. For example, none of the students recognized the symbol π as representing the population parameter for a proportion, a disconnection that may also be explained by the gestalt principle of Past Experiences. Students were unfamiliar with the notation of π to represent a population proportion, a major barrier to any task completion using that statistic. The incorporation of both common symbols (p and π) for the parameter might reduce that confusion. This should be a top priority change in the applet's design because it is a barrier that happens in the initial data input stage of the interaction with the applet, leading to students' ultimate failure to use the LO as a learning tool.

Students identified another weakness in the applet's design: its labeling. The current design of the applet includes inconsistent labeling in a small font (approximately pt. 8). In the current design, only some of the applet's labels are capitalized, and none of the panels are labeled. Labeling is an essential part of the design interface because it communicates the function of each component to the user. The multiple recommendations specifically regarding the font used in the applet implies that labeling was important to the evaluators; changing the font and/or increasing the size of the font may improve the quality of the applet's overall design.

5.2 Implications for Practice

The results of this study have implications for both developers of statistics LOs (e.g. designers and evaluators) and statistics educators. Statistics LO developers can possibly use the strengths and weaknesses evaluators identified in this study as criteria to decide which features they should include or avoid in applets they design or evaluate in the future. Developers can also focus on what the instructions in the LO's interface is communicating to users. Currently, there are only three bullet points for instructions and they are at the bottom of the screen. To improve the design of the applet, developers could use a short introductory tutorial that includes detailed descriptions of each component of the applet, how to use each one, and the connections between them. A voice narration providing users with a preliminary tour of the applet and its features may engage and inform users (Plaisant & Shneiderman 2005).

Student evaluator feedback during the usability test indicated that they had a hard time understanding the relationship between the intervals graphed in Panel 2 and the "Last Sample" and "Sample Statistics" plots. This may be a result of their organization in the current design applet. The plot showing the distribution of the last sample drawn relates to a single sample and its corresponding interval while the plot for the distribution of sample statistics shows data from all samples drawn in the most recent simulation. Despite the graphs looking similar and being displayed close together in Panel 3, they represent very different aspects of the samples the applet simulates. The gestalt principle of "Similarity" explains that if components share similar features, they are perceived as being related. If developers separate the two plots by putting one in Panel 1 and the other in Panel 3, the space between the plots may help communicate the differences in their purpose, which should lead to increased usability.

Additionally, think-aloud comments that students made during the Task 2 activity may give LO developers' insight into what students are thinking and understanding during their interaction with the tool. However, the short introduction to the think aloud process at the beginning of the usability test may not have been sufficient to make this process comfortable for the student evaluators. A longer or more in-depth introduction to the think aloud protocol may produce better results. If I had used a list of specific prompting questions as students moved through Task 2 this may have yielded more substantial data from the think-aloud processes. To maximize the potential of the data evaluators collect in a usability test, evaluators should be careful to plan enough introduction to a think-aloud protocol to increase students' to effective participation.

In the future, statistics educators could use these results to inform their decisions on which tools to select for use in their classrooms. Responses on the Task 2 activity, student evaluator self-diagnosis, and think aloud comments identified content knowledge as a barrier to task completion. The use of an activity in conjunction with the applet may be most effective for teaching and learning of sampling distributions of a sample mean. Assessing an activity in conjunction with the applet will allow teachers to diagnose issues in content knowledge. Statistics teachers may also want to incorporate a think aloud protocol into classroom practices to gain further insight into misconceptions their students may have. The applet itself does not provide students with any opportunity for reflection after interacting with the tool. Teachers can

possibly enhance their students' learning experiences by incorporating a reflection activity after interaction with the applet.

The relatively short period of time student evaluators spent in their independent exploration (less than two minutes, on average) and the number of applet features left unexplored may indicate that the students' exploration may be more beneficial if it had some structure. For example, during the usability test, student evaluators rarely used the interactive features of the applet unless they were guided to them. For developers, this may indicate that open exploration is not a likely source of valuable feedback. Also, statistics teachers can choose or design activities that include specific instructions to use any interactive features desired. A checklist of features to explore or a diagram numbering the different areas of the graph the students are to investigate could possibly enhance the exploration period. However, in order to give students the maximum potential benefits from exploring the applet, teachers who are having their students work independently with LOs may want to give structure to this exploration.

For students to progress in their learning goals through meaningful interactions with an educational technology, they need to be comfortable working with it. As students completed items similar to tasks they had previously seen on the Task 2 activity, their errors seemed to correct themselves. The activity guided students through using the different features of the applet and they seemed to realize new connections between the applet's interactive components as they worked through the exercise. When students progressed through the activity, they seemed more confident of their answers and sought little to no assistance. Also, I did not notice any difference in the comments or the body language between students who answered items correctly and those who answered incorrectly. The purpose of any LO should be to enhance learning (Chance et al., 2007; Friel, 2007; Garfield, Chance & Snell, 2000; Peterson 2007), and therefore, teachers should provide students with multiple exposures to an LO so they can focus on the concept being learned instead of the functions of the tool. Teachers can also incorporate repetition into the tasks they give to their students to help them become more familiar with the features of an LO. If developers create "Back" and "Forward" buttons that allow users to move between previous and current samples and intervals, students may more effectively understand the relationships between the output of the LO and the concepts they are learning.

5.3 Recommendations for Improving Usability

This study evaluated the usability of the Rossman/Chance applet. I used the findings to create a list of potential improvements for the applet's usability in the next stage of its development, as seen in Table 4. I have organized these recommendations into three categories: Display, Layout, and Features. Recommendations regarding labeling and information to be displayed in the interface are in the "Display" category. I categorized recommendations that apply to the physical placement of components on the screen into "Layout." The "Features" category contains any recommendations that involve the creation or modification of the applet's functions.

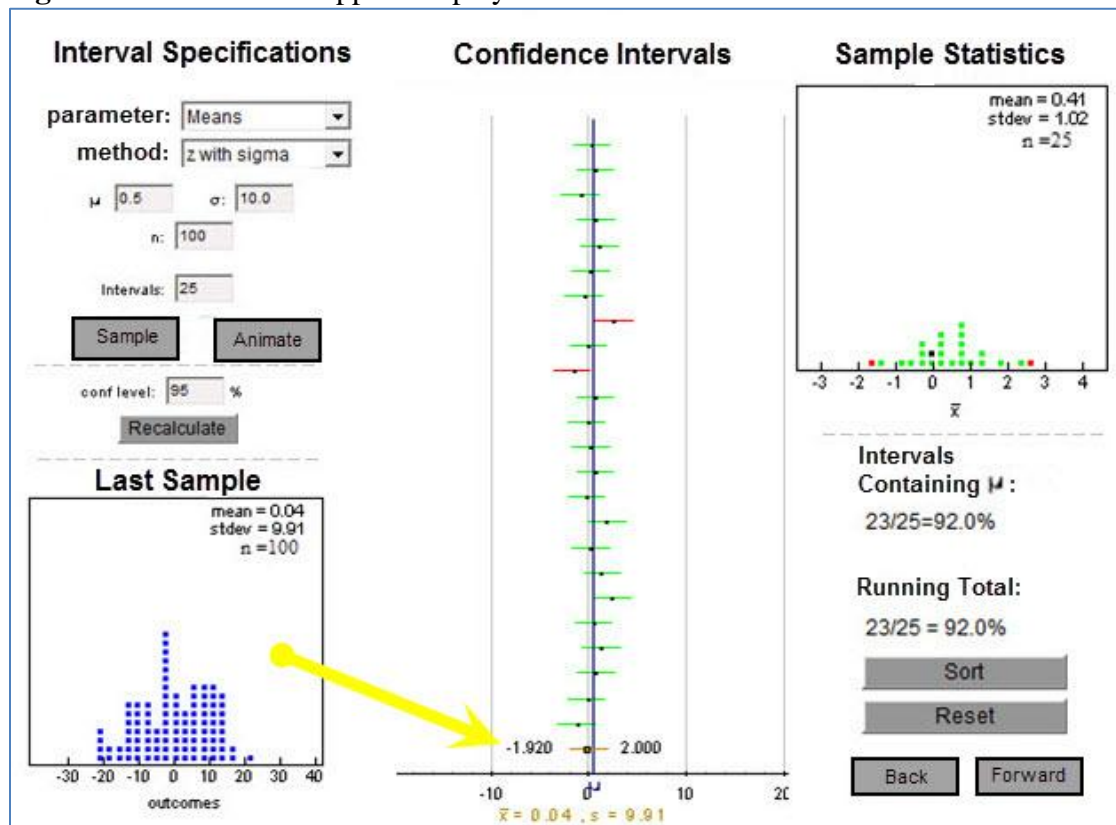
Table 4 Recommendations for Improving Usability

Category	Recommended Change
Display	<ul style="list-style-type: none"> • Add titles for every panel and display • Include the symbol p as a representation for the population proportion • Increase and/or change the font of the labels • Add $n =$ to the “Last Sample” and the “Sample Statistics” plots
Layout	<ul style="list-style-type: none"> • Separate the “Last Sample” and the “Sample Statistics” plots to different panels of the applet • Utilize more of the horizontal screen space available
Features	<ul style="list-style-type: none"> • Create and link to a tutorial activity that guides students through the features • Automatically select the interval calculated from the last sample (endpoints showing) in Panel 2 • Match the color for the last sample and its interval in Panel 2 • Include a tutorial that describes all of the applet’s features • Give users the option to animate the interval construction • Add buttons to the display that allow students to move “back” and “forward” to the previous or next set of results • Highlight the μ input field in Panel 1 and the vertical line representing μ in Panel 2 in the same color • Change the “Last Sample” plot to represent any selected interval

Figure 4 models what the applet might look like for the confidence intervals for means after designers have implemented the majority of the recommended changes. For changes to the display in direct response to student evaluator feedback on labeling, I added titles to each of the panels, increased the font of the existing labels, and added sample size to the “Last Sample” and “Sample Statistics” plots. To address students’ apparent confusion between the information displayed in the applet’s two plots, I moved the “Last Sample” plot into the bottom of Panel 1 to create a visual disconnect between the two plots. I also added an arrow connecting the “Last Sample” plot to the confidence interval constructed from that sample data. Next, I relocated the

“Sample Statistics” plot to the top of Panel 3 with the display that is in the bottom of Panel 1 (in the current design) underneath because the proportion of intervals containing μ is also displayed by the color used in the sample statistics plot and having them close together may help students make the connection more easily than with the current design. I added an “Animate” button next to the existing “Sample” button. If students can have an animated view of the multiple intervals being constructed it may help connect each sample to its corresponding interval. I also included “Back” and “Forward” buttons to Panel 3 which will allow students an opportunity to record and their mental paths while working with the applet.

Figure 4 Possible New Applet Display



In addition to these recommendations, it may benefit the applet developers to strive to maintain the strengths of the existing version of the applet. The simplicity of the interface, the interactive features, the use of color, and the information presented in the graphs should continue as a part of the design. Together, the results of this usability test and the recommendations they guide can serve as an outline for improving the applet.

5.4 Limitations

It is important to note that the results from this usability test should not be directly applied in other contexts. In future usability tests I would replicate these methods with a larger sample size. Additionally, the data collected in this study do not permit an examination of any changes in

student understanding as a result of the features of the applet. To assess student learning, researchers could utilize pre- and post-tests, or interviews with the student evaluators before and after they completed the evaluation in future usability testing of LOs.

6. Conclusion

The usability test completed in this study is the first of its kind in the statistics education literature including results that provide an example of the valuable feedback that can be collected from user-centered evaluation. Improving the usability of LOs based on results from usability testing will potentially make LOs more effective learning tools for statistics students and teachers. Statistics educators can use the results of the usability test as criteria to inform their decisions on which tools to select for use in classrooms. Statistics LO developers can use the list of strengths compiled in the results to decide which features should be included in applets designed in the future. For example, selecting and designing LOs that are easy to use and highly interactive can help ensure the greatest likelihood that those students have a positive experience when interacting with statistics LOs. This study's results also corroborate existing educational design theory. Many of the students' observations and recommendations are consistent with the design principles grounded in gestalt theory. By connecting theory to practice, this study implies that LO developers can use design theory to create and improve effective learning tools.

The recommendations for improving future designs of the applet have the potential to improve the usability and effectiveness of the Rossman/Chance Simulating Confidence Intervals Applet. I plan to work with the developers of the applet to apply the recommended changes to the existing design and, afterward, I would like to continue to test the usability of the updated applet. If the next usability test reveals that the applet improved (as evidenced by increased positive feedback, a decrease in the recorded weaknesses and number of recommended changes), this could lead to additional research comparing between the two versions of the applet. The revised applet could then be used in a study of its effectiveness as a learning tool for teaching confidence intervals or in a study of learning outcomes when the revised applet is compared to other similar LOs.

References

- Akpınar, Y. (2008). Validation of a learning object review instrument: relationship between ratings of learning objects and actual learning outcomes. *International Journal of E-learning and Learning Objects*, 4(1), 91-302.
- Al-Aziz, J., Christou, N., & Dinov, I. D. (2010). SOCR motion charts: An efficient, open-source, interactive and dynamic applet for visualizing longitudinal multivariate data. *Journal of Statistics Education*, 18(3), 1-29.
- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., & Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. Retrieved from <http://www.amstat.org/education/gaise/>
- Baki, A., & Çakıroğlu, Ü. (2010). Learning objects in high school mathematics classrooms: Implementation and evaluation. *Computers & Education*, 55(4), 1459-1469.
- Barnum (2011). *Usability Testing Essentials*. Morgan Kaufmann, Waltham, MA.
- Berg, G. A. (2000). Human-computer interaction (HCI) in educational environments: Implications of understanding computers as media. *Journal of Educational Multimedia and Hypermedia*, 9(4), 347-368.
- Chance, B., Ben-Zvi, D., Garfield, J., and Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, 1(1). Retrieved from <http://escholarship.org/uc/item/8sd2t4rr>
- Chance, B., and Rossman, A. (2006). *Using simulation to teach and learn statistics*. Proceedings of the 7th International Conference on Teaching Statistics, Salvador, Bahia, Brazil.
- Che, A., Cui, J., & Dinov, I. D. (2009). SOCR analyses: implementation and demonstration of a new graphical statistics educational toolkit. *Statistics Online Computational Resource*. Retrieved from <http://escholarship.org/uc/item/0jx0p32c>
- Colborne, G. (2010). *Simple and usable web, mobile, and interaction design*. New Riders, Berkeley, CA
- Goodyear, P., & Retalis, S. (2010). *Technology-enhanced learning*. Sense Publishers, Boston, MA
- Harvey, B. (2005). Learning objects and instructional design. *The International Review of Research in Open and Distance Learning*, 6(2).
- Hsu, Y. (2003). "The effectiveness of computer-assisted instruction in statistics education: A meta-analysis." PhD dissertation, University of Arizona, Arizona.

Iordache, D. D., & Pribeanu, C. (2009). Comparison of Quantitative and Qualitative Data from a Formative Usability Evaluation of an Augmented Reality Learning Scenario. *Informatica Economică*, 13(3), 67-74.

Johnson, J. (2010). *Designing with the mind in mind: Simple guide to understanding user interface design rules*. Morgan Kaufmann, Waltham, MA.

Kaufmann, H., & Dünser, A. (2007). Summary of usability evaluations of an educational augmented reality application. In *Virtual Reality* (pp. 660-669). Springer Berlin Heidelberg.

Kaufmann, H., & Papp, M. (2006). Learning Objects for Education with Augmented Reality. In *Proceeding of EDEN (European Distance and E-Learning Network) Conference, Vienna*, 160 – 165.

Kaufmann, H., & Schmalstieg, D. (2003). Mathematics and geometry education with collaborative augmented reality. *Computers & Graphics*, 27(3), 339-345.

Kaufmann, H., Schmalstieg, D., & Wagner, M. (2000). Construct3D: a virtual reality application for mathematics and geometry education. *Education and Information Technologies*, 5(4), 263-276.

Kay, R. H., & Knaack, L. (2008). A multi-component model for assessing learning objects: The learning object evaluation metric (LOEM). *Australasian Journal of Educational Technology*, 24(5), 574-591.

Lane, D., & Scott, D. W. (2000). Simulations, case studies, and an online text: a web based resource for teaching statistics. *Metrika*, 51, 67-90.

Nielsen, J., & Landauer, T. K. (1993, May). A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, 206-213.

Petersen, D. (2007). Usability theory, practice and evaluation for learning objects. *Learning objects: Applications, implications, & future directions*, 337-370.

Plaisant, C., & Shneiderman, B. (2005). Show me! Guidelines for producing recorded demonstrations. In *Visual Languages and Human-Centric Computing, 2005 IEEE Symposium*, 171-178.

Reeves, T. (2008). Evaluation of the Design and Development of information technology tools in education. *International Handbook of Information Technology in Primary and Secondary Education*, 20 (pp. 1037-1051).

Schenker, J. D. (2007). "The Effectiveness of Technology Use in Statistics Instruction in Higher Education: A Meta-Analysis Using Hierarchical Linear Modeling." PhD dissertation, Kent State University College, Ohio.

Schwarz, C.J. (2007). Computer-aided statistical instruction – multi-mediocre techno-trash. *International Statistical Review*, 75(3), 348-354.

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: a review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, 20(2), 2012.

Tullis, T., & Albert, W. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann, Burlington, MA.

Wang, S. K., & Yang, C. (2005). The interface design and the usability testing of a fossilization web-based learning environment. *Journal of Science Education and Technology*, 14(3), 305-313.

Wertheimer, M. (1938). Laws of organization in perceptual forms. *A source book of Gestalt psychology*, 71-88.

Wiley, D A (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy, in *The instructional use of learning objects: Online version*, Retrieved from <http://reusability.org/read/chapters/wiley.doc>

Yacovelli, S. R. (2003). Understanding learning objects: the basic" chunks". *College & University Media Review*, 10(1), 17-26.

Zaharias, P. A. (2004). Usability and e-learning: The road towards integration. *eLearn Magazine*, 4(6). Retrieved from <http://elearnmag.acm.org/featured.cfm?aid=998345>

APPENDIX

Rossman/Chance CI applet: <http://www.rossmanchance.com/applets/NewConfsim/Confsim.html>

Initial Impressions Questionnaire Items

- What are your initial impressions of the applet?
- Have you ever used an applet (or similar learning object) like this before?
- What type of applet(s) have you used in the past? (only answered if previous answer was affirmative)
- In what context(s) did you use the applet(s)?

Follow-Up Questionnaire Items

- What are your overall impressions of (or your feelings about) the applet you tested today?
- List any strengths and/or weaknesses you see for using this applet as a learning tool for teaching one-sample confidence intervals.
- List any barriers you encountered while trying to complete the tasks on the activity.
- What recommendations do you have to improve the design of this applet?
- What recommendations do you have to make this applet a more effective learning tool?

CHAPTER 5. GENERAL CONCLUSIONS

INTRODUCTION

It is hard to ignore the enormous impact that technology has made on the teaching and learning of statistics in the past twenty years (Belli, 2003). However, the lack of empirical evidence in the literature regarding effects on learning indicates that statistics education research has yet to advance much beyond descriptive studies about the types of technologies available and suggestions for using them (Garfield & Ben-Zvi, 2008; Garfield, 2006; Schenker, 2007). The field of human computer interaction (HCI) provides theories and principles that directly apply to any interaction between a human and a machine and, therefore, can be incorporated into studies focusing on the use of technology in statistics classrooms. One goal for HCI researchers is to develop new or to improve existing technologies by using approaches that focus on the end-users. Therefore, integrating HCI principles into statistics education studies on technology can provide a theoretically sound foundation to build on. Currently, research explicitly tying together the fields of statistics education and HCI is virtually nonexistent. With so much uncharted territory, it is an exciting time to be doing research in these areas because there is great potential to impact both fields.

The purpose of this study was to explore technology use in statistics education based on principles of human computer interaction (HCI) in order to advance current knowledge in both disciplines. Together the three articles in this dissertation investigate the use of learning objects (LOs) currently available for introductory statistics topics, evaluate LOs for one topic according to an established rubric, and measure the quality of one LO through a student-centered usability test. In this chapter, I provide a brief description of the three papers and then I discuss how this

body of work may inform theory, gaps in the existing research, and future practices for both statistics educators and educational technology developers.

The first article, “Learning Object Use in Introductory Statistics Classrooms: A Survey of Current Practices, Teachers’ Perceptions, and Advice to Peers,” summarizes the results of a survey completed by high school and post-secondary statistics educators regarding the nature of their LO use. The results of the survey showed an overlap between the LOs described in the statistics education literature and the LOs the respondents indicated they were using. More importantly, the results identified educators’ use of LOs not represented in the literature, including blogs, case studies, and online news sources. The four most popular types of LOs indicated by the respondents were applets, data sets, videos, and electronic textbook materials.

The second article, “Exploring the Design Quality of Sampling Distribution Learning Objects: An Expert Evaluation,” describes an expert-based heuristic evaluation used to investigate the quality of statistics applets designed to explain sampling distributions of sample means. Fourteen secondary and postsecondary statistics educators used the Learning Object Evaluation Metric (LOEM, Kay & Knack, 2008) to evaluate the quality of the existing sampling distribution of the sample mean learning objects (SDMLOs). The LOEM Total Score data allowed me to classify the SDMLOs as “above average,” “average,” and “below average.” I also found that there were similarities in the designs for both the highest and the lowest scored LOs. An examination of similarly rated LOs revealed that those with the highest ratings gave users multiple options to control the output and had sophisticated interactive features (e.g. “real-time” manipulation of a graph), while those LOs with the lowest ratings allowed users minimal opportunities to control and interact with the tools.

In the third article, “Usability of a Confidence Interval Applet as Evaluated by College Statistics Students,” I reported the results of a formative usability test completed by student evaluators. Student evaluators noted the strengths of the applet were its visual representations of the topic, interactivity, and ease of use. From the results of the usability test I created a list of recommendations for improving the usability to share with the developers as they pursue the next stage of the applet’s development.

IMPLICATIONS FOR THEORY AND THE LITERATURE

Theories are important in education because they provide explanations and predictions for how people behave and learn. In this study I described activity theory, which outlines a system of interconnected components that are involved in a learning environment. I also outlined how constructivism, an epistemology to explain how individual students construct knowledge by reconciling new information with their past experiences and understandings, can be infused into activity theory to create a constructivist activity system. To incorporate HCI theory, I connected design principles based in gestalt theory to the usability test results.

The detailed theoretical framework I described in the first article in this dissertation infuses constructivism into activity theory in a way that may provide an outline for educators to use in conceptualizing future studies. The model of a hybrid constructivist activity system includes a visualization of the relationship between the contextual factors in a learning activity system and the overlap between desired learning goals and actual knowledge outcomes. My diagram of the hybrid theory should easily extend to future studies in education and technology design because the activity theory element applies to all components that influence the learning environment and the constructivism element directly ties to each student’s individual understanding of knowledge.

In addition to learning theory, I connected theory based design principles to the results of the usability test in the third article. The student evaluators in that study identified strengths and weaknesses of the applet that corroborate many of the principles of design that originate in gestalt theory (Wertheimer, 1938). Consistency between student evaluator comments and the design principles reflect that these theories may be directly applicable to LO design and evaluation. I recommend gestalt principles be considered in any design standards developed for educational technologies and in any future evaluations.

The educational technology literature I reviewed for these studies has shown that LOs often include features that are reusable and interactive, which promote learning as described by constructivist principles (Peterson, 2007; Wiley, 2000). Despite the availability of many statistics LOs (Schenker, 2007; Hsu, 2003), the current statistics education literature is largely comprised of descriptions of LOs and reports of their perceived benefits and limitations of use. Only a handful of studies show evidence of the level of LOs' effectiveness as learning tools.

Each article in this study fills a gap in the existing literature by addressing LO quality and use at a level of detail that has not been done before. The first article connects the teaching practices of over sixty teachers to the LO use suggested and described in the literature by providing information on the specific types of LOs respondents are using along with how and why they are using them. The evaluation results in the second and third articles are the first of their kind in the statistics education literature and have the potential for being a foundation for researchers to start an extensive line of research investigating the quality of the many classifications or individual LOs for statistics.

Throughout this study, I observed that much of the statistics education literature did not provide a theoretical framework or simply included an understanding of how learning happened,

such as constructivism, without any explicit connection between theory and results or practice. I hope that the theoretical framework in this dissertation's first study will serve as an example of how to explicitly incorporate more theory into statistics education research. In addition to contributing to the existing literature, the results from this study have the potential to start lines of research that will add tremendous depth to the existing statistics education and HCI literature.

IMPLICATIONS FOR PRACTICE

This dissertation's synthesis of ideas from the fields of statistics education and human-computer interaction can provide insight for statistics educators and developers of statistics LOs. Together the articles inform statistics educators about multiple aspects of the interaction between their students, instructional contexts, and statistics LOs, as well as selection criteria to ensure the most effective tools are being chosen. The LO is a classification of technology that statistics education researchers can examine as they build knowledge about which tools are most effective in helping students meet their learning goals. However, the present studies did not investigate the direct effect statistics LOs have on students' conceptual understanding of statistics topics. These effects could be studied in statistics classes offered in in-class, hybrid, and distance education modalities. For example, learning studies could be conducted, possibly using pre-test/post-test designs, to investigate the effect different LOs can have on student learning in statistics courses. Groups could be compared using LOs that scored similarly or differently on the expert evaluation. Another study could compare scores of students who used a newly developed version of the applet evaluated in the third article with those who used an older version of the applet.

Future studies could more deeply explore the use of LOs in statistics education by widening the scope of the studies reported here. The survey in the first article was limited to respondents who had used LOs, so no information on the percentage of statistics instructors

using LOs can be drawn from this data. A more broad survey of statistics instructors could investigate the proportion of instructors who are using LOs in order to provide this information. A review summarizing the presence of LOs in published curriculum materials would provide further information on the availability of these tools and the degree to which the promotion of LOs in curriculum materials influences their use by instructors. The present study focused on the introductory statistics level and it may be beneficial to investigate LO use in more advanced statistics courses.

For developers of statistics LOs, including designers and evaluators, this study can provide insight into who is using statistics LOs, their decision making considerations when choosing which LOs to use, and aspects of design and usability that impact users' perceptions of the quality of educational technologies. The rubric-based heuristic evaluation in the second article provided a basic overview of the quality of a group of existing SDMLOs. The usability test including student (end-user) evaluators in the third article outlined a procedure that can be used to assist in the refinement of existing LOs or the design and development of new LOs. The extensions from both studies are limited because of the small sample size of evaluators. In future studies, larger sample sizes and training evaluators on the rubric are recommended to generate stronger results and provide information that could be compared to the results of this study. Further evaluation research could extend the investigations in these studies to LOs on different statistical topics.

CONCLUSION

Incorporating technology into the teaching and learning of statistics gives students opportunities "to visualize the results of varying assumptions, explore consequences, and compare predictions with data" (CCSSM; National Governors Association Center for Best

Practices and Council of Chief State School Officers [NGA Center and CCSSO], 2010). Many of the technologies freely accessible on the Internet and designed to assist in the teaching and learning of introductory statistics topics can be classified as learning objects (LOs). Together the articles in this study can impact the way introductory statistics technologies are developed, evaluated, and incorporated into classrooms. Statistics teachers and LO designers and evaluators can incorporate the findings into their own designs, evaluations, and classroom practices. While the use and effectiveness of LOs in statistics education remains a largely unstudied topic, this study has provided an initial example of possible explorations into the uses, evaluation, and design of statistics LOs.

References

- Belli, G. (2003). *Finding, evaluating, and organizing Internet resources: issues for statistics instruction*. Paper presented at the International Association of Statistical Education Conference on Statistics Education and the Internet, Satellite Conference to the 54th Session of the International Statistical Institute (ISI), Berlin, Germany.
- Garfield, J. B. (2006), "Collaboration in Statistics Education Research: Stories, Reflections, and Lessons Learned," in Proceedings of the Seventh International Conference on Teaching Statistics, eds. A. Rossman and B. Chance, Salvador, Bahia, Brazil: International Statistical Institute, pp. 1-11.
- Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York, NY: Springer.
- Hsu, Y. (2003). "The effectiveness of computer-assisted instruction in statistics education: A meta-analysis." PhD dissertation, University of Arizona, Arizona.
- Kay, R. H., & Knaack, L. (2008). A multi-component model for assessing learning objects: The learning object evaluation metric (LOEM). *Australasian Journal of Educational Technology*, 24(5), 574-591.
- Petersen, D. (2007). Usability theory, practice and evaluation for learning objects. *Learning objects: Applications, implications, & future directions*, 337-370.
- Schenker, J. D. (2007). "The Effectiveness of Technology Use in Statistics Instruction in Higher Education: A Meta-Analysis Using Hierarchical Linear Modeling." PhD dissertation, Kent State University College, Ohio.
- Wertheimer, M. (1938). Laws of organization in perceptual forms. *A source book of Gestalt psychology*, 71-88.
- Wiley, D A (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy, in *The instructional use of learning objects: Online version*, Retrieved from <http://reusability.org/read/chapters/wiley.doc>