SCHENKER, JASON D., Ph.D., December 2007 Educational Foundations and Special Services

THE EFFECTIVENESS OF TECHNOLOGY USE IN STATISTICS INSTRUCTION IN HIGHER EDUCATION: A META-ANALYSIS USING HIERARCHICAL LINEAR MODELING (223 pp.)

Director of Dissertation: Shawn M. Fitzgerald, Ph.D.

The use of various forms of technology to enhance instruction has been on the rise throughout colleges and universities in the United States, including increased use of technology in traditional classrooms, increased enrollment in distance education courses, and increased use of two-way interactive video or teleconferencing (Van Dusen, 1998). Several researchers have found that enhancing instruction with technology leads to greater levels of learning and achievement among students than traditional instruction (Kulik & Kulik, 1986; Bayraktar, 2002; Timmerman & Kruepke, 2006). The purpose of the study was to examine the effectiveness of using technology to enhance statistics instruction, using meta-analytic techniques. Specifically, the focus was to determine whether or not various uses of technology differentially affect statistics achievement and attitudes among undergraduate or graduate students. Also, the relationships between effect size and several intervention and methodological variables across studies were examined.

The mean achievement effect size from 46 studies was found to be 0.239, indicating that technology was modestly effective in improving students’ statistics achievement. Simulations were significantly more effective than other technology types, while online learning was no more effective than traditional instruction. Other variables

related to achievement effect size included course level, student academic standing, publication status, and type of control group. Significantly larger effect sizes were found in introductory courses, in courses with graduate students, in studies published in professional journals, and studies that included a control group that received no instruction.

The mean effect size, across 27 studies, for the effect of technology on attitude was found to be 0.162, and was not statistically significant. Four technology types were significantly, but negatively, related to attitude effect size: statistical analysis software, enhanced lecture, web-based, and online learning. Other variables significantly related to attitude effect size included year, function, duration, academic discipline, course type, and publication status. Studies with significantly larger attitude effect sizes included those that were more recently published, involved technologies that served supplemental rather than substitutive functions, were of longer duration, conducted in mathematics departments, conducted using research methods rather than statistics classes, and published in professional journals.

THE EFFECTIVENESS OF TECHNOLOGY USE IN STATISTICS INSTRUCTION IN HIGHER EDUCATION: A META-ANALYSIS USING HIERARCHICAL LINEAR MODELING

A dissertation submitted to the
Kent State University College and Graduate School
of Education, Health, and Human Services
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

by

Jason D. Schenker

December, 2007

A dissertation written by

Jason D. Schenker

B.A., Heidelberg College, 1996

M.A., University of Akron, 1998

Ph.D., Kent State University, 2007

Approved by

________________________________, Director, Doctoral Dissertation Committee
Shawn Fitzgerald

________________________________, Member, Doctoral Dissertation Committee
Rafa Kasim

________________________________, Member, Doctoral Dissertation Committee
Drew Tiene

Accepted by

________________________________, Chairperson, Department of Educational Foundations and Special Services
Awilda Hamilton

________________________________, Interim Dean, College and Graduate School of Education, Health, and Human Services
Donald L. Bubenzer

ACKNOWLEDGMENTS

First, I would like to thank my wife, Molly, for all the help and support she provided along the way. I would also like to thank my committee members, Shawn Fitzgerald, Rafa Kasim, and Drew Tiene for their willingness to provide suggestions and guidance. Finally, I would like to thank my parents, Gerald and Joy Schenker, and the rest of my family, friends, and co-workers who have provided support in their own ways throughout this journey. Thank you all.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The Topic and Its Context

The use of various forms of technology to enhance instruction has been on the rise throughout colleges and universities in the United States, including increased use of technology in traditional classrooms, increased enrollment in distance education courses, and increased use of two-way interactive video or teleconferencing (Van Dusen, 1998). This increase has been driven, to some extent, by an increased demand for education, especially among older, non-traditional students (Levine, 2001). Colleges and universities must now compete for recruitment, retention, and graduation of students, and the inclusion of technology not only attracts students, but allows colleges and universities to reach a greater number of students who were previously isolated by location or finances (Duhaney, 2005; D. L. Rogers, 2000). In addition, colleges and universities often must integrate technology and gear their instruction toward technology competency standards in order to obtain accreditation (Rogers). Surry and Land (2000) commented that “many university administrators now view technology as a cost-effective and innovative solution to many of higher education’s problems” (p. 145). As early as 1985, Balkovich, Lerman, and Parmelee (1985) cited five uses of computers in academic settings, including in computer science and electrical engineering instruction and research, financial management and record keeping, word processing, spreadsheet analysis, personal database management, electronic mail and other forms of asynchronous communication, and as an integral part of the instructional process. Miller, Martineau, and Clark (2000) commented:

> Technology based teaching may not be essential in all classes, but it is most generally facilitative. Technology infusion facilitates learning by providing relevant examples and demonstrations, changing the orientation of the classroom, preparing students for employment, increasing flexibility of delivery, increasing access, and satisfying public demands for efficiency. (p. 229)

Surveys of faculty members across the country have indicated that instructors generally believe in the value of technology and regularly use technology to supplement their instruction in some manner (Jacobsen, 1998; Vodanovich & Piotrowski, 2001). For example, researchers who have surveyed faculty members have found that instructors often make use of word processing, email, and Internet searching and browsing (Jacobsen, 1998). Others have found that they also make frequent use of presentation and spreadsheet software in their classes (Groves & Zemel, 2000). Still others have commented that the "most prevalent uses of instructional technology are enhanced presentations, simulations, computer tools, collaboration and communications, access to support materials, research tools, and evaluation and testing" (Miller et al., 2000, p. 235). In addition, instructors and universities often create web pages for every course where class notes, syllabi, schedules, and assignments can be posted. Furthermore, software programs are often used to simulate experiments that can be costly or dangerous when using real materials.

However, despite the amount of money spent on technology throughout colleges and universities, technological innovations for teaching and learning have not been widely integrated into the curriculum (Geoghegan, 1994), and many instructors only use such technology for fairly low-level purposes (Jacobsen, 1998; Vodanovich & Piotrowski, 2001). Surry and Land (2000) commented that the use of technology for administrative and data management purposes is common on college campuses, but technology has not been fully integrated into classrooms. Likewise, Miller et al. (2000) indicated that whereas technological advancements are present throughout universities, they are least apparent "in the core of the academy-the curriculum" (p. 228), and that the classrooms of today do not differ largely from classrooms of 50 or 100 years ago. Furthermore, Massy and Wilger (1998) presented three levels of technology adoption in higher education, and commented that instructors have yet to reach the highest level of integration, which involves reconfiguring instruction to take full advantage of the technology that is available. Van Dusen (1998) lamented that, "Fewer than half of U.S. colleges and universities have a strategic plan for technology" (p. 64). Massy and Wilger added that a "serious commitment to technology innovation encourages and may even require closer attention to the fundamental principles of pedagogy and quality" (p. 52). Others, however, have argued that instructors are more often using technology for delivery of instruction rather than simply for classroom management purposes (Duhaney, 2005).

Instructors, however, often face several obstacles or barriers, both personal and organizational, to using technology in their classrooms. For example, at the organizational level, some instructors might face a lack of computer resources, a lack of services provided by information technology, highly structured course plans, and limited time for learning and preparing lessons (Weston, 2005). In addition, organizational barriers include a lack of leadership and an inflexible organizational culture (Miller et al., 2000). Furthermore, universities often fail to reward innovative teaching (Jacobsen, 1998; Miller et al., 2000; D. L. Rogers, 2000), and some faculty members might lack technological competence and may not have received adequate training or had access to such training (Miller et al.; Rogers; Van Dusen, 1998; Vodanovich & Piotrowski, 2001). In addition, the cost of implementing the infrastructure for technology is a challenge for many brick-and-mortar institutions (Van Dusen). Furthermore, personal barriers to technology integration include lack of confidence, procrastination, lack of motivation (Weston, 2005), fear of change, and inertia (Miller et al.). Groves and Zemel (2000) found, after surveying higher education faculty members, that the factors most commonly cited that influenced instructional technology use included equipment availability, improved student learning, increased student interest, advantage of traditional teaching methods, and ease of use (Groves & Zemel). Finally, Surry and Land (2000) suggested that instructors not only need to be made aware of the actual benefits of using technology, but also need to know how technology will help or support him or her.

Other authors, however, have been less optimistic about the promise of using technology to enhance instruction, or believe that such use has been uncritically accepted as necessary even though its effectiveness has not been fully established (Grineski, 1999). For example, in an early critique of the effectiveness of instructional media, Clark (1983) argued that the studies and meta-analyses of media effectiveness at the time clearly suggested that media did not affect learning. According to Clark, media is simply a vehicle for the delivery of instruction, and any benefit of a new form of instructional media was a result of either confounding media with instructional method, or due to a novelty effect. Clark further argued that the same methods could be used without the media to obtain the same level of effectiveness, and that the attributes present in a particular media are not unique, and can be shared by several forms of media. For example, radio and television share several attributes. Others such as Kozma (1994), however, have commented that the media must be confounded with the method in order to be effective, and that individual attributes are not what make a particular medium unique, but its combination of attributes. Furthermore, from Kozma’s point of view, media and technological devices are designed by human beings, and even if there were presently no relationship between media and learning, that does not preclude one from being created through enhancing the design of media and technology.

Additionally, several authors have expressed skepticism concerning the educational utility of online learning (Feenberg, 2001; Garson, 1999). Online courses can be demanding of instructors’ time, are often offered by lower-prestige universities, teach students facts rather than critical thinking, are often associated with less writing rather than more, and force instruction to be programmed around concrete learning objectives rather than abstract concepts (Garson). Furthermore, instructors often incorporate technology in a manner that does not take full advantage of its capabilities, such as simply using it as high-tech substitutes for blackboards, overhead projectors, or handouts

(Reeves, 1991, as cited in Jacobsen, 1998). Additional authors have commented that learning requires contact between human beings (Botstein, 2001), and "educational technologies that lack an interactive component, such as televised courses and computer-aided instruction, have never succeeded in displacing teachers from the front of the classroom" (Feenberg, 2001, p. 83). Garson (1999) argued that some of the best uses of technology involve using it to supplement rather than supplant traditional instruction and that using technology to supplant traditional instruction is a threat to liberal education. Also, Feenberg commented that technology should provide instructors with different options to enhance instruction, not provide a single method for all instructors, and stated, "The quality of college education is at stake not in whether we use computers but in how we use them. This is the real problem distorted in the current debates for and against technology" (p. 90). Finally, Botstein argued that the focus of instruction should be on teaching students the skills of discrimination, analysis, and interpretation, presumably to be able to evaluate, select, and integrate the plethora of information available on the Internet and elsewhere.

Benefits of Enhancing Instruction With Technology in Higher Education

Although the concerns of the skeptics do have some merit, the use of technology for instruction in higher education does possess several advantages. One advantage of increasing the use of technology in higher education is to improve access to education for students and provide flexible delivery of content (Miller et al., 2000). More traditional and nontraditional students, who may be geographically isolated or have other obligations that prevent them from physically attending a university, can now enroll in higher education through distance learning programs (Duhaney, 2005; Grant & Spencer, 2003; Levine, 2001), and colleges and universities may be forced to compete for these students (D. L. Rogers, 2000). In addition, some authors have suggested that the use of multimedia technology in classrooms can result in greater student retention at the institutional level (Lau, 2003). Furthermore, students have more options to learn and study at times, locations, forms, and speed that are most suitable to them (Duhaney; Miller et al.). Levine commented that "the education which is going to be asked for is both traditional, campus based courses and nontraditional, any time, any length education" (p. 256). Duhaney stated that technology provides a wider array of methods to present and represent information. For example, faculty members are often using web pages to supplement their classes, as well as establishing discussion boards to continue class discussions.

Other authors have argued that the increased use of technology in higher education can improve cost-effectiveness. This can be accomplished through the use of technology by either increasing enrollment or increasing student or administrative productivity (Grant & Spencer, 2003; Massy & Wilger, 1998; Van Dusen, 1998). For example, Van Dusen stated that technology used for administrative services (scheduling, financial aid, etc.) can easily prove cost effective, and Massy and Wilger argued that productivity software can save individual students' time. However, several authors are skeptical that using technology to enhance classroom instruction can increase cost effectiveness. For example, Balkovich et al. (1985) predicted that the widespread use of computers would be more likely to increase the cost of education than decrease it. In addition, Van Dusen argued that the cost of implementing the infrastructure for technology is a challenge for many brick-and-mortar institutions, and that technology can reduce some student expenses, but may provide little in savings for institutions, making it difficult to fully offset the costs. On the other hand, authors such as Massy and Wilger (1998) do believe that technology has the potential to increase higher education productivity provided faculty members reconfigure instruction to take full advantage of new technology. They stated that, “Technology’s long-term economic advantage is that it opens up more options, more ways to adapt teaching and learning processes to whatever financial conditions may ensure” (p. 52). Furthermore, Balkovich et al. argued that, in order for technology to be cost-effective, universities and faculty members must limit the amount of training time so that it does not interfere with instructional time. Finally, D. L. Rogers (2000) argued that faculty members have had little training on how to make effective instructional use of technology. Universities must invest in faculty development and support services in order to fully take advantage of the cost savings potential of technology (Van Dusen).

Also, a number of researchers have found that enhancing instruction with technology leads to greater levels of learning and achievement among students than traditional instruction. For example, Kulik and Kulik (1987) conducted a meta-analysis including 199 studies across elementary, secondary, post-secondary, and adult education settings, comparing computer-based instruction to traditional instruction on final examination scores, and found an effect size of 0.31 standard deviations in favor of computer-based instruction. Kulik and Kulik published an updated version of this meta-analysis in 1989 and found an effect size of 0.30 from 254 studies. When examining only studies conducted at the higher education level, J. A. Kulik, C. C. Kulik, and Cohen (1980) found an effect size of 0.25 from 59 studies. In a later meta-analysis using 99 studies conducted in higher education, Kulik and Kulik (1986) found an effect size of 0.26 favoring computer-based instruction.

Researchers have also found that students are more satisfied and prefer classes that incorporate technology into instruction. For example, Johnson and Dasgupta (2005) found that introductory statistics students preferred non-traditional classes to traditional classes. In addition, Mitra and Steffensmeier (2000) found that students were more comfortable with computers, had less apprehension toward them, believed they made the process of learning easier, and expressed a desire to use computers throughout college after a university implemented an extensive computerization project campus-wide. Researchers have also found, through the use of meta-analysis, small effect sizes in favor of computer-based education on student attitudes. For example, Kulik and Kulik (1987) found that computer-based instruction resulted in a 0.28 effect size for attitude toward instruction and 0.33 for attitude toward computers when compared to traditional instruction. Computer-based instruction did not, however, have an effect on attitude toward subject matter. Kulik and Kulik found similar results in 1989. When using studies conducted only at the higher education level, J. A. Kulik, C. C. Kulik, and Cohen (1980) found an effect size of 0.24 when student attitudes were used as an outcome measure in 11 of the 59 studies. Finally, Kulik and Kulik (1986) also examined student attitudes at the higher education level and found that computer-based instruction resulted in an effect size of 0.27 for attitude toward computers, and 0.31 for attitude toward instruction. However, they found a near zero effect size for attitude toward subject matter.

Beyond improved achievement and attitudes, researchers have discussed several additional benefits for students that result from using technology in education. For example, Balkovich et al. (1985) suggested that the use of technology for instruction could be beneficial because data can be obtained and analyzed more accurately, and technology can be used to simulate dangerous or costly experiments. In addition, others have suggested that the use of technology can help save students' time and make them more productive (Massy & Wilger, 1998; Van Dusen, 1998). The use of technology can also lead to increased productivity by allowing students to have more convenient access to educational resources, such as library databases and other online sources of information (D. L. Rogers, 2000). Others have added that the use of technology for instruction can help students obtain skills that are highly relevant for the workplaces of today (Forte, 1995; Levine, 2001; Miller et al., 2000; Raymondo, 1996). Finally, including technology for instruction can lead to a reduction in instructional time (Kulik & Kulik, 1986; Kulik & Kulik, 1987, 1989; J. A. Kulik, C. C. Kulik, & Cohen, 1980).

Although numerous researchers have found technology effective in improving student achievement and attitudes, several additional researchers have gone further by providing explanations as to why technology is effective. For example, D. L. Rogers (2000) posited seven reasons that educational technologies can prove advantageous for instruction: (a) it provides more frequent and timely interactions among students and between students and faculty, (b) it encourages coordination and collaboration between students, (c) it promotes active learning techniques, (d) it provides immediate feedback, (e) it makes studying more efficient by providing immediate access to online resources, (f) it provides forums for self and peer evaluation, and (g) it allows for self-paced learning. In addition to Rogers, other authors have commented on the benefits of technology in improving communication among students and between students and faculty members, as well as promoting cooperation and collaboration. For example, Miller et al. (2000) commented that technology has the potential to improve student cooperation, collaboration, and communication with the professor as well as other students. In addition, Lau (2003) stated that faculty members can develop a positive learning environment by using instructional technology to enhance cooperative and collaborative learning in the classroom. Furthermore, researchers have found that students believed that computers affected the ways in which students and teachers interact (Mitra & Steffensmeier, 2000). Brett and Nagra (2005) found, through observations and surveys, that a computer-based social learning space promoted collaboration between students. Van Dusen (1998) commented that technology affects how students learn, with instructors recognizing that students can become more collaborative in their learning processes, and discussions online can often have a higher level of participation and be of higher quality than those in the classroom.

Like Van Dusen (1998), other researchers have commented on the advantages of technology in promoting participation and active learning. Traditional classrooms are often designed to be teacher-oriented rather than student-oriented (Miller et al., 2000; D. L. Rogers, 2000). Erwin and Rieppi (1999) argued that multimedia-enhanced classrooms improve students' participation, enhance the presentation of information, and facilitate student feedback. In addition, Eisenberg (1986) found that students appreciated the use of computer programs for administering course material because they allowed for all students to participate rather than a few dominating ones. Also, the active learning approach that can be implemented in multimedia-enhanced classrooms enhances students' motivation, interest, and enjoyment (Erwin & Rieppi). Forsyth and Archer (1997) found that the introduction of technology into instruction was a positive learning experience for students and enhanced their motivation, and optional use of technology was related to attendance rate and student retention at the university. Furthermore, Mitra and Steffensmeier (2000) found that students' attitudes toward computers improved as the university became more computer-enriched, and they expressed the desire to use computers in college and believed they made the process of learning easier. However, whereas students tend to view the use of computers and technology in college instruction positively and expect some technology use in every class, they have been shown to prefer that less than half of classroom activities were devoted to incorporating technology (Rickman & Grudzinski, 2000). The students felt that instructional technology did not always enhance the instructional process, and often instructors came to class unprepared and resorted to reading the information that was being displayed.

One of the most commonly cited reasons used to justify the use of technology to enhance instructional effectiveness is that it often provides immediate feedback for the student. Lovett and Greenhouse (2000) commented that learning is more efficient when students receive real-time feedback on errors, thus preventing students from repeating errors, and "strengthening incorrect knowledge, acquiring invalid procedures, or strengthening inappropriate connections" (p. 5). Immediate feedback was a much-lauded feature of the early teaching machine and programmed instruction movement (Benjamin, 1988; Reiser, 2001). For example, Garson (1999) commented that programmed instruction was based on the principles of "clearly stated behavioral objectives, small frames of instruction, self-pacing, active learner response to frequent prompts and questions, and immediate individualized feedback to responses" (p. 5). A variety of researchers have speculated that the effectiveness of technology when used for instruction is due, in part, to its ability to provide immediate feedback (Kulik & Kulik, 1989). For example, Aberson, Berger, Healy, and Romero (2003) concluded that the tutorial software they used in their study might have been more effective because it provides immediate feedback to students if they make mistakes, and can address specific misunderstandings. In addition, authors have found that immediate feedback is one of the features most appreciated by students using computer programs for instruction (Eisenberg, 1986). However, Timmerman and Kruepke (2006) conducted a recent meta-analysis on the effectiveness of computer-assisted instruction, and found no significant difference between computer-based interventions that provided feedback and those that did not. Likewise, Buzhardt and Semb (2002) found no difference between the performances of those who received item-by-item feedback versus end-of-test feedback.

The use of technology in instruction also has the potential to improve upon the effectiveness of traditional methods because it often allows for self-paced learning (D. L. Rogers, 2000). Authors such as Lovett and Greenhouse (2000) have suggested that students learn best what they practice and perform on their own. Self-paced learning allows students to proceed at their own individual speeds, where slower students are not penalized for taking longer than others, but also do not hold back the faster students (Grant & Spencer, 2003). Like providing immediate feedback, a variety of attempts at improving instructional effectiveness throughout the years have involved allowing for increased self-paced instruction, including the early programmed instruction and personalized system of instruction (PSI) movements of the 1950s through 1970s (Garson, 1999; Keller, 1968; Reiser, 2001). For example, the PLATO system, a computer-based programmed instructional system, was designed to provide interactive, self-paced instruction to large numbers of students, and reached its zenith in the 1970s (S. G. Smith & Sherwood, 1976). However, self-paced instruction often remains a feature on various computer-based instructional programs of today, and studies have shown that students often appreciate this feature (Shuell & Farber, 2001). In conclusion, the use of technology to enhance instruction in higher education has long held promise, and many of its advantages are beginning to become evident.

Benefits of Enhancing Statistics Instruction With Technology in Higher Education

Like the use of technology to enhance instruction in other fields, a number of authors have provided many similar rationales for the use of technology in statistics instruction, specifically. For example, researchers have argued that the use of technology for instruction can allow students to advance at their own pace (Forte, 1995) and help them develop their own understanding of statistical concepts (Mills, 2003). In addition, others have suggested that the use of technology in statistics instruction can help students develop computer literacy (Fernandez & Liu, 1999; Raymondo, 1996; R. L. Rogers, 1987) and provide them with skills, such as those related to statistical software use, that could have value in future employment settings (Tromater, 1985). Furthermore, students who are being instructed in statistics, like other academic fields, can benefit from the use of technology in that many software programs provide instant feedback (Aberson et al., 2003; Lovett & Greenhouse, 2000). The use of technology in instruction can also increase efficiency in other ways. For example, Raymondo commented that computer technology allowed him to cover more material in a shorter time. Likewise, Fernandez and Liu stated that the use of technology (MS Word presentations, spreadsheets, PowerPoint® slides, SAS macros, etc.) allowed them to present, in their estimation, three times more information than with the traditional lecture method. In addition, Yilmaz (1996) suggested that technology could be incorporated into the curriculum to free up classroom time that is often spent on technical details. For example, Tromater developed a course in computer-assisted statistical analysis in order to reduce the amount of lab time that was lost completing statistical computations. Furthermore, Erwin and Rieppi (1999) found that they could more effectively teach a larger number of students in a multimedia classroom environment than a smaller number of students in a traditional classroom environment. Finally, many statistics professors enjoy using technology to assist in instruction and would like to incorporate technology in instruction more often, such as using computers to have students analyze real, relevant data sets (Strasser & Ozgur, 1995).

Also, like technology use in other fields, the use of technology in statistics instruction promotes active learning by allowing instructors to present concepts and procedures in an active and engaging manner (Forte, 1995). Authors have suggested that students learn best by their own activity; when they practice and perform on their own (Lovett & Greenhouse, 2000; Velleman & Moore, 1996). Likewise, Romero, Berger, Healy, and Aberson (2000) commented, “Learning is enhanced by motor activity, even when the skill being learned is not a motor task” (p. 248). Ben-Zvi (2000) and Moore (1997) suggested that statistics courses be taught using more activities that serve as alternatives to lectures, such as group problem-solving and discussion, laboratory exercises, demonstrations, written and oral presentations, and group and individual projects. Such activities can help students become more active in their learning over lectures that focus on derivations and computations. Technology allows instructors to provide more of these hands-on activities instead of lecturing and providing formulas to be memorized. Marasinghe, Meeker, Cook, and Shin (1996) suggested using simulation programs so that students can actively manipulate the data in order to see the effect of any changes that are made. This method could be more effective than simple data analysis software because students may not grasp the underlying concepts simply by conducting analyses. Velleman and Moore (1996) also suggested that the use of a multimedia platform that allows students to work with real data could improve their understanding of statistics. As Ben-Zvi (2000) stated, technological tools allow for the development of a “richer, powerful, and flexible learning environment in which students are active learners of statistics" (p. 149). Furthermore, teachers should be encouraged to view these tools as extensions of human cognitive abilities.

Others have suggested that the use of technology in statistics instruction can help to reduce anxiety in students (Forte, 1995) and improve their attitudes toward statistics. For example, Fernandez and Liu (1999) integrated a variety of technologies into their statistics instruction, including MS Word presentations, spreadsheets, PowerPoint® slides, SAS macros, java applets, real world data sets, and so forth, and found that students experienced less anxiety, and were more motivated and engaged as a result. Johnson and Dasgupta (2005) found that introductory statistics students preferred non-traditional classes to traditional classes, and Maltby (2001) found that the majority of students surveyed, especially younger students, preferred learning statistics with the assistance of a computer. Furthermore, Lesser (1998) incorporated a greater use of computers and real-world data sets, and found that students significantly preferred various aspects of the technology-enhanced course when compared to a control course. In addition, 92.9% of the students in the treatment section said that they would not have enjoyed the traditional approach more. Some authors have commented that the use of technology can make statistics instruction more interesting (Lovett & Greenhouse, 2000), and that multimedia classrooms improved students' motivation, interest, and enjoyment (Erwin & Rieppi, 1999). Also, Aberson et al. (2003) found that students who were exposed to an interactive computer statistics tutorial expressed interest in using similar tutorials for additional assignments. Finally, Velleman and Moore (1996) stated, "Another goal of a first statistics course, often unstated but nonetheless important, is to motivate students to change their attitude about statistics" (p. 219).

In addition, the use of technology in statistics can have positive effects on student achievement. Lovett and Greenhouse (2000) stated that technology assists in improving students' understanding and problem-solving skills in relation to statistics, "most likely because the technology give students more opportunities to consider conceptual implications and work through problems on their own" (p. 9). In addition, Lesser (1998) found that students who were instructed in statistics in a technology-enhanced environment performed better on questions that involved critical thinking than those who were not. Erwin and Rieppi (1999) found that students instructed in multimedia-enhanced classrooms obtain higher statistics final examination scores than those who were taught in traditional classrooms. Furthermore, Christmann and Badgett (1999) found an effect size of 0.256 in favor of various computer software packages over traditional instruction. Wang and Newlin (2000), however, found that students who enrolled in an online statistics course performed significantly worse on the final examination than students taught in a traditional classroom. Overall, studies indicate that using technology can improve student achievement in statistics courses, although this result might not extend to fully online courses.

Additional authors have suggested that statistics be taught using more analysis of real data using computers to complete computations rather than spending excessive amounts of time lecturing and conducting derivations and computations (Moore, 1997). As Forte (1995) commented, computer-assisted instruction allows students to "spend more time analyzing the implications of generated statistics and less time learning the statistical formulas needed to manually calculate them" (p. 5). This allows instructors to provide more hands-on activities rather than lecturing and providing formulas. In addition, Lovett and Greenhouse (2000) suggested emphasizing students' practical use of statistical reasoning over the memorization of statistical formulas and procedures, whereas R. L. Rogers (1987) commented that the use of computers permits students to spend less time doing calculations and allowing more time for conceptual understanding. Likewise, Fernandez and Liu (1999) commented:

> The traditional mixture of lecture and arithmetic problems tends to be too abstract, forcing students to devote too much time in the actual calculating process and therefore to lack an understanding of the overall structure about what they are learning or what a specific calculation is for. (p. 174)

Furthermore, Moore (1997) commented that although older statistics courses may have focused more strongly on probability-based inference, a more modern course might focus more heavily on data analysis, data production, inference, data exploration, using diagnostic tools, and so forth, that are more ambiguous than the straightforward nature of traditional statistical calculations. Technology assisted in this change. Moore also commented that instructors should focus on topics that cannot be automated (unlike calculations) such as interpretation of graphs, strategies for effective data exploration, basic diagnostics as preliminaries to inference, and the conceptual meaning of various statistical terms. Calculations and derivations should only be done if they help students to understand the underlying concept. He stated, "We should encourage students to use software to explore, visualize, and interact with data and simulations" (p. 10), and "emphasis on visualization and problem-solving are hardly possible if graphics and calculations must be done by hand" (p. 8). Ben-Zvi (2000) concurred by stating that technology has

> Led to a shift in introductory courses from traditional views of teaching statistics as a mathematical topic (with an emphasis on computations, formulas, and procedures) to the current emphasis on statistical reasoning and the ability to interpret, evaluate, and flexibly apply statistical ideas. (p. 130)

Ben-Zvi suggested that statistics instructors allow students to automate computations using technological tools, as they can help students do what they can already do, but do it faster, more often, more accurately, and with fewer errors.

Perhaps the most common rationale provided for using technology to enhance statistics instruction is that technology allows students to easily access, store, and analyze real-world data sets so that students can observe the practical value of statistical concepts. For example, Lovett and Greenhouse (2000) indicated that knowledge tends to be specific to the context in which it is learned, and they suggested that students work on real-world data that cover a variety of contexts. Likewise, Yilmaz (1996) stated that statistics courses should develop several competencies in students, including the ability to link statistics to real-world situations, knowledge of basic statistical concepts, and the ability to understand and communicate the results of a statistical study. In addition, Strasser and Ozgur (1995) suggested that instructors spend less time lecturing on probability and classical hypothesis testing, and spend more time using computers to analyze real data. Furthermore, Moore (1997) commented that statistic classes should be taught using more analysis of real data with computations done by computer, rather than spending time conducting derivations and computations. He stated that, "computing allows realistic problems, serious statistical methods, and emulation of actual statistical practice" (p. 10). Ben-Zvi (2000) also suggested that a first course in statistics incorporate more data and concepts and treat formal derivations as secondary in importance. Technology could be used in statistical education through the incorporation of exploratory data analysis, which includes a heavy reliance on visual displays as analytical tools. Technological tools can assist in developing these visual displays and creating simulations that help students to understand statistical concepts and methods. In addition, a multimedia platform allows students to work with real data, so that students can draw conclusions about the real-world rather than simply complete calculations (Velleman & Moore, 1996). Another advantage of technology is that students can obtain real data sets from the Internet that they can use for data analysis practice using statistical analysis software (Hunt & Tyrrell, 2000; Rowell, 2004). Some instructors have altered their statistics courses according to these principles. For example, Lesser (1998) implemented new content standards for his statistics courses which involved instructing students to be able to critically evaluate statistics in the media and in their major field of interest, and to plan, implement, and communicate the results of a real-world research project, and incorporated a greater use of computers and realistic data sets.

Several professional organizations have also called for changes to be made in statistics instruction from lectures and derivations of formulas to greater use of software to analyze real-world data sets. For example, the American Statistical Association/ Mathematical Association of America Joint Curriculum Committee for the teaching of statistics recommended a greater focus on using data, an emphasis on statistical concepts, and fewer calculations (Moore, 1997). Statistical classes should focus more on statistical ideas and concepts that are not mathematical in nature, and use more analysis of real data with computations done by computer, rather than spending time conducting derivations and computations. In addition, Strasser and Ozgur (1995) commented on the recommendations put forth during the annual conference entitled “Making statistics more effective in schools of business” or MSMESB. The recommendations of conference attendants included a call for greater focus on graphical data analysis, more descriptive statistics, more quality control, and a lesser focus on probability and classical hypothesis testing. In addition, attendees further recommended that instructors spend less class time lecturing and more time using computers to analyze real data. Furthermore, Lesser (1998) altered his statistics courses to be more in line with the National Council of Teachers of Mathematics’ standards, which included instructing students to be able to critically evaluate statistics in the media and in their major field of interest, planning, implementing, and communicating the results of a real-world research project, and incorporating a greater use of computers and realistic data sets.

Statement of the Problem

Instructors of statistics courses have been among the earliest and most common to incorporate technology for the purpose of enhancing instruction. In 1987, R. L. Rogers commented that, among psychology courses, computers were most often used in statistics courses, permitting students to spend less time doing calculations and allowing for more time for conceptual understanding. As early as 1991, a survey indicated that 70% of introductory statistics courses used computer software (Khamis, 1991). However, Fernandez and Liu (1999) commented that the question still remained concerning what and how specific software should be integrated into the teaching and learning process. This study endeavored to provide an answer to that question.

In addition, a great amount of scholarly writing has been devoted to promoting, or defending, the use of various types of computer and technology use for enhancing statistics instruction. For example, a number of authors have commented on the value of using statistical analysis software programs to automate calculations and analyze real-world data sets (Ben-Zvi, 2000; Fernandez & Liu, 1999; Forte, 1995; Lovett & Greenhouse, 2000; Yilmaz, 1996). However, statistical analysis software programs may be one of the least effective uses of technology in statistics instruction (Christmann & Badgett, 1999; Hsu, 2003). An additional integration of the research literature might help to clarify this issue. Furthermore, empirical studies of technology use are quite common in the professional literature on statistics instruction. For example, Becker (1996) found that computer use was the second most frequent topic of empirical writing on statistics instruction. Despite this, few meta-analyses have been conducted to integrate the wide range of findings concerning technology use in statistics instruction.

Purpose/Research Questions

The purpose of the present research study was to examine the effectiveness of using technology to enhance statistics instruction, using meta-analytic techniques, with data analyzed using hierarchical linear modeling (Raudenbush & Bryk, 2002). The term "technology," for the purpose of this study, referred to the use of computers and other similar digital and electronic devices that can be used to assist instruction. Specifically, the major focus was to determine whether or not various uses of technology differentially affected statistics achievement and attitudes among students enrolled in undergraduate or graduate statistics courses. In addition, this study tentatively compared the effects of several additional program/intervention and methodological/study characteristics based on the availability of studies that exhibited such characteristics. The program/intervention characteristics included function of technology use (supplemental versus substitutive uses of technology), duration of technology use (e.g., one-time use, repeated uses, entire semester, etc.), academic discipline (e.g., education, psychology, business, mathematics, sociology, etc.), level of statistics course (e.g., introductory, advanced, quantitative methods, statistical computer applications, etc.), and student academic standing (i.e., undergraduate versus graduate). The methodological/study characteristics included instructor bias (i.e., same professor versus different professor teaching experimental and control conditions), research design (i.e., experimental versus quasi-experimental designs), and source (e.g., refereed journal, ERIC document, doctoral dissertation or master's thesis, etc.). The following research questions guided this study.

Research Question 1

Are student achievement and attitude (e.g., satisfaction with the course, instructor, or content) affected by the use of technology to enhance statistics instruction?

Null Hypothesis

H_0: $\delta_i = 0$, for all $i = 1,\ldots, k$, where δ refers to the true effect size and k equals the number of studies included in the meta-analysis.

Explanation

Most meta-analyses of the effectiveness of technology use in instruction suggest that it has a small or moderate affect on achievement and a small effect on attitudes. For example, J. A. Kulik, C. C. Kulik, and Cohen (1980) calculated an effect size of 0.25 across 54 studies that included examination performance as outcome measures, and an effect size of 0.24 across 11 studies that used student attitudes as outcome measures. Kulik and Kulik (1986) found a mean effect size of 0.26 across 99 studies of various types of technology use in higher education classes, and effect sizes of 0.27 and 0.31 for attitudes toward computers and attitudes toward instruction, respectively. More recently, Timmerman and Kruepke (2006) found an effect size of 0.24 on achievement when instruction was enhanced with technology across 188 studies. Furthermore, Christmann and Badgett (1999) found a mean effect size of 0.26 on achievement across 9 studies that used computer software to enhance statistics instruction. Finally, Lee (1999) found an effect size for computer-based instructional simulations of 0.41 on achievement and –0.04 on attitudes. Although technology used to enhance instruction does appear to have at least a modest effect on student achievement, its effects on attitudes are less clear.

Research Question 2

Does significant between-study variance in effect sizes exist for both achievement and attitudes?

Null Hypothesis

H_0: $\tau^2 = 0$, where τ^2 is the variance of the effect size parameters.

Explanation

If significant variation exists in the effect sizes, this suggests that study and intervention characteristics may moderate the effect sizes (H. A. Kalaian, Mullan, & Kasim, 1999). According to Kalaian et al., “In HLM, conditional analyses are conducted if initial unconditional analyses leave significant unexplained variance” (p. 212). Should the null hypothesis be rejected, further analyses could be conducted to determine the moderating effects of several study and intervention characteristics on the variation found in the effect sizes between studies.

Research Question 3

If significant between-study variance in effect sizes exists, can several study and treatment characteristics explain this between-study variance in effect sizes, including type of technology (i.e., drill and practice, tutorial, simulations, statistical analysis, computer laboratories, enhanced lectures, web-based, programmed or personalized systems of instruction, and online), function of technology (supplemental vs. substitutive), treatment duration, academic discipline, level of statistics class (i.e., introductory, intermediate, or advanced), academic standing of students (i.e., undergraduate or graduate), instructor bias, research design (i.e., experimental vs. quasi-experimental studies), and publication status (i.e., published vs. unpublished)?

Null Hypothesis

H_0: $\gamma_a = 0$, for all $a = 1,\ldots, l$, where γ_a are the regression coefficients associated with each study characteristic, and l is the number of coded study characteristics.

Explanation for the Examination of Study and Treatment Characteristics

Uses of technology. Meta-analyses of computer or technology use in higher education have exhibited conflicting results when examining different types of technology use. For example, an early study by C. C. Kulik, J. A. Kulik, and Cohen (1980) found an overall effect size across all technologies of 0.28 on student achievement, but an effect size of 0.55 when personalized systems of instruction (PSI) were used. However, other meta-analyses found no differences between different uses of technology on student outcomes (C. C. Kulik et al., 1980; Kulik & Kulik, 1987; Kulik & Kulik, 1989). However, several more recent meta-analyses have exhibited differences between types of technology use, often with simulation or tutorial software programs showing larger effect sizes than other types (Bayraktar, 2002; Christmann & Badgett, 1999; Khalili & Shashaani, 1994; Liao, 1998, 1999). Overall, a lack of consistency appears to exist across previous meta-analyses of technology use in higher education when comparing different types of use.

Function of technology. Some authors have argued that technology uses that supplement instruction are more effective than technology uses that serve as substitutes for instruction (Benjamin, 1988; Garson, 1999). However, researchers who have examined substitutive versus supplemental uses of technology have found conflicting results. For example, Khalili and Shashaani (1994) compared courses that used computer-based education as a supplement to instruction to those where computer-based education was used as a replacement for teacher instruction and found similar effect sizes of 0.38 and 0.34, respectively. However, Olson and Wisher (2002) found, through using meta-analytic techniques, a higher effect size on student achievement for classes that blended web-based instruction with face-to-face instruction (0.48) than for fully online classes (0.08). Furthermore, Bayraktar (2002) found similar results, with an effect size of 0.288 when the computer was used in a supplementary role and 0.178 when used as a substitute in a meta-analysis of computer-assisted instruction in science education. Hsu (2003) also found a significantly larger mean effect size for supplemental forms of CAI in statistics instruction than for substitutive forms.

Treatment duration. Previous meta-analyses have found conflicting results when examining the effectiveness of the duration of the technological intervention. For example, several authors have found that instructional uses of technology of shorter duration were more effective than those of longer durations (e.g., Bayraktar, 2002; Kulik, 1983; Kulik & Kulik, 1987; Kulik & Kulik, 1989), suggesting that a novelty, or Hawthorne, effect might be present, whereby students perform better because the technology condition is new and unusual rather than due to an inherent superiority to using the technology itself (Kulik & Kulik, 1989). However, not all researchers have found that higher effect sizes correspond to shorter interventions. For example, Khalili and Shashaani (1994) found a much larger effect size for interventions lasting one to two months (0.94), than for those lasting less than one month (0.14). Timmerman and Kruepke (2006) also found a larger mean effect size when CAI was used more than once than when CAI was used only once. In addition, other researchers have found no differences based on the duration of the intervention (J. A. Kulik, C. C. Kulik, et al., 1980).

Academic discipline. Several authors have examined the effectiveness of the use of technology to enhance instruction across different academic areas. For example, Christmann, Badgett, and Lucking (1997b) conducted a meta-analysis, and found that computer-assisted instruction was most effective in science courses (0.639) and lowest for vocational education (-0.080). Studies that used social studies classes had an average affect size of 0.205. However, Kulik and Kulik (1986) found a larger effect size for the soft sciences (0.35) than the hard sciences (0.15). Likewise, Timmerman and Kruepke (2006) found a larger effect size for the social sciences over the physical sciences and life sciences. Overall, some inconsistencies exist when examining the effectiveness of technology use for instruction across academic disciplines.

Level of statistics class. While researchers have conducted studies of the effectiveness of the use of technology to enhance statistics instruction at both introductory (e.g., Basturk, 2005; Lesser, 1998; Stockburger, 1982; Summers, Waigandt, & Whittaker, 2005) and upper-level statistics courses (e.g., Morris, Joiner, & Scanlon, 2002; Varnhagen, Drake, & Finley, 1997), no meta-analytic studies were found that compared the effectiveness of technology use across levels of statistics course.

Academic standing of students. No studies were found that compared the effectiveness of technology use in statistics course in undergraduate and graduate level courses.

Instructor bias. Instructor bias can occur when instructors teach both the control and experimental conditions, and teach in a manner than they unconsciously provide an advantage to the experimental group. However, several meta-analyses of technology use in instruction have actually found higher effect sizes when different instructors teach each condition. For example, J. A. Kulik, C. C. Kulik, and Cohen (1980) found larger effect sizes when different instructors were used for each condition. Likewise, Bayraktar (2002) found smaller effect sizes (0.218) when the same instructor taught both the computer-assisted and control conditions than when different instructors taught each section (0.328). Khalili and Shashaani (1994) also sizes when different instructors taught the experimental and control groups than when the same instructor taught both. However, Liao (1998, 1999) found larger effect sizes when the same instructor taught both sections, indicating that instructor bias may have been present.

Research design. Authors who conduct meta-analyses regularly code for various study features, including the research design. However, authors have typically found no significant differences in effect sizes between quasi-experimental and experimental designs (Kulik, Kulik, & Shwalb, 1986; Liao, 1999). Khalili and Shashaani (1994), on the other hand, did find larger effect sizes for those studies that used nonrandom assignment over those that used random assignment.

Publication status. A number of meta-analyses have found larger effect sizes for published studies than those that have not been published (Kulik & Kulik, 1987; Kulik et al., 1986). For example, Kulik and Kulik (1986) found that studies obtained from professional journals had significantly higher effect sizes (0.42) than dissertation studies

(0.11). Likewise, Kulik and Kulik (1989) obtained an effect size of 0.44 for published studies and an effect size of 0.24 for unpublished studies.

Significance and Rationale

Colleges and universities spend considerable funds building the infrastructure for technology, making technology available to students and faculty members. However, technology use to enhance or improve instruction has not always proven to be cost-effective (Van Dusen, 1998). According to Van Dusen, the cost-effectiveness of technology use depends on the specific pedagogical applications that are implemented. In other words, enhancing instruction with technology can prove cost-effective if it is done appropriately. As Massy and Wilger (1998) stated, "A serious commitment to technology innovation encourages and may even require closer attention to the fundamental principles of pedagogy and quality" (p. 52). The current study assists in providing instructors with results that may inform them on what uses of technology are most effective and appropriate.

In addition, few meta-analyses have been conducted that have examined the effectiveness of technology use specifically in statistics courses. For example, whereas Christmann and Badgett (1999) conducted a meta-analysis of the effectiveness of several computer-based software packages, they only included software that they labeled as computer-assisted instruction (CAI), problem-solving, and statistical analysis software. Thus, a variety of uses of technology, such as student response systems, presentation software, various Internet-bases uses of technology, and so forth, were not included in the meta-analysis. In fact, Christmann and Badgett only included 14 effect sizes across 9 studies, and 10 of 14 of the effect sizes were from studies that used statistical analysis software, leaving only two effect sizes for each of the other two types of software examined. Hsu (2003) also conducted a meta-analysis of the effectiveness of computer-assisted instruction (CAI) in statistics courses, including drill and practice, tutorials, computational, simulations and games, web-based, expert systems, and multimedia. Like Christmann and Badgett, Hsu also found a great number of studies that used computational software, but relatively few that were assigned to several of the additional categories. Also, while Hsu included four effect sizes from courses that were enhanced using web-based technologies, she did not specifically compare fully online courses to those that make more supplemental uses of Internet technology, such as course web pages, email, discussion forums, chat rooms, and so forth, as an addition to classroom lectures. The current study included results obtained from fully online statistics courses that were then compared to studies of statistics courses that use more supplemental forms of Internet technology. In addition, the previous meta-analyses did not examine the effectiveness of the use of technology to enhance statistics instruction across different academic disciplines, as was one of the purposes of this study. Furthermore, this study provided an update to the earlier meta-analyses by including more recently conducted or published studies.

Furthermore, this study employed hierarchical linear modeling (HLM) techniques to conduct the meta-analysis (Raudenbush & Bryk, 2002). HLM is an appropriate data analysis method for meta-analytic studies because the data, or individual studies, are already hierarchical in nature. In other words, participants are nested within a study. As

Osborne (2000) noted, “people or creatures that exist within hierarchies tend to be more similar to each other than people randomly sampled from the entire population” (p. 1). HLM is often used to analyze data from multi-site studies, where study participants are nested within a particular site. In some respects, meta-analyses can be viewed as multi-site studies, where multiple, similar studies are conducted at a variety of locations by a number of different researchers (S. A. Kalaian, 2003). In such cases, the participants that are available at each site, in this case undergraduate and graduate students enrolled in statistic courses, may be more similar within each site than across the entire population. Also, several studies are often nested within an individual article or paper. Furthermore, HLM allows for the analysis of fixed and random effects, whereas traditional meta-analytic techniques only allow for the analysis of fixed effects (H. A. Kalaian et al., 1999). The use of HLM allows the researcher to treat the individual studies as a random sample from a population of similar studies (S. A. Kalaian, 2003), and models variation among effect sizes as a function of defined study characteristics and random error, whereas traditional meta-analysis approaches assume all variation in effect sizes is explained by known study characteristics (H. A. Kalaian et al., 1999). Researchers have argued that the use of a mixed effects design is more appropriate when the studies examined in the meta-analysis are relatively heterogeneous, rather than homogeneous, in nature (Pearson & Lipton, 1999), when the studies examined represent a sample from a universe of similar studies that have either been conducted or could potentially be conducted (Raudenbush, 1994), and when there is reason to believe that not all moderator variables can be identified and coded (Pearson & Lipton; Raudenbush). Furthermore, the use of the mixed effects design allows the researcher to generalize to a larger population of studies, whereas the fixed effects approach only allows generalization to similar participants (Pearson & Lipton; Raudenbush). Additionally, the use of HLM to analyze mixed-effects models in meta-analyses is appropriate because the data in meta-analyses are almost always unbalanced, leading to unequal precision in estimating effect sizes across studies (Raudenbush). For this reason, ANOVA procedures, which assume balanced designs, are not optimal. Finally, few meta-analytic studies have been conducted using HLM to analyze the data thus far (e.g., S. A. Kalaian, 2003; H. A. Kalaian et al., 1999; Sliwinski & Hall, 1998; Tengs & Lin, 2003), and none have been conducted on the effectiveness of using technology to enhance statistics instruction.

Limitations

Several potential limitations existed in the present study. First, the term "technology" is not easily defined, and can vary in its breadth and inclusiveness. For example, Pearson and Young (2002) defined technology as:

> The process by which humans modify nature to meet their needs and wants. However, most people think of technology only in terms of its artifacts: computers and software, aircraft, pesticides, water-treatment plants, birth-control pills, and microwave ovens, to name a few. But technology is more than its tangible products. Equally important aspects of technology are the knowledge and processes necessary to create and operate those products, such as engineering know-how and design, manufacturing expertise, various technical skills, and so on. Technology also includes all of the infrastructure necessary for the design,

> manufacture, operation, and repair of technological artifacts, from corporate headquarters and engineering schools to manufacturing plants and maintenance facilities. (p. 2)

However, McOmber (1999) presented three different definitions of technology that are in common circulation. First, the "technology as instrumentality" definition refers to technology as any tool created by humans (p. 141). Under this broad definition, a pencil is just as much a form of technology as a computer. According to a second definition, "technology as industrialization," the term technology coincides with the industrialization of the West or the "technological age" (p. 142). Under this definition, technology is "as much an event as a set of practices or objects" (p. 143). Finally, the "technology as novelty" definition refers to objects that represent the most recent human instrumental developments (p. 143). Under this definition, technology is displaced so that what was once considered technology will one day no longer receive such a designation. For example, whereas an overhead projector would have been considered a form of technology at one historical point, especially when compared to a chalkboard, few scholars in the field of instructional technology would include an overhead projector in their definition of technology today. The reader can find elements of all three definitions of technology in the single definition proposed by Pearson and Young.

Whereas all three definitions of technology have merit for certain purposes, the definition of technology used for this study most closely adheres to the "technology as novelty" conceptualization. Under this definition, an overhead projector or television set would not be considered technology, whereas a computerized projection system would, for example. However, none of the three definitions, used independently, perfectly encapsulates the concept of technology, so adhering to any single definition will have its limitations.

Likewise, although no single, all-encompassing definition of technology exists, nor do specific categories of types of technologies. Whereas one of the purposes of this study was to examine the effectiveness of various types of technologies, there are no clear cut boundaries between what is and is not a technology, not to mention what constitutes a certain type of technology. As a result, any categorization scheme used to sort technologies is imperfect. However, attempts were made to overcome this limitation by sorting technologies according to common features, as well as obtaining consensus among more than one rater.

Additionally, several of the categories used to organize types of technologies may appear overly broad. For example, courses conducted entirely online through the Internet can still be conducted through vastly different instructional techniques or methodologies. Some may include only asynchronous communication, whereas others include both synchronous and asynchronous communication. Likewise, some may involve little contact among students and only focus on interactions between the student and the instructor, whereas others may involve a great deal of contact among fellow students. Furthermore, some may involve little more than placing a form of programmed instruction software online for students to access. However, this study attempted to account for some of the differences in different forms of online learning by allowing for each study to be coded according to multiple categories. For example, an online programmed instruction course would have been categorized as both “online learning” and “programmed instruction.” Likewise, the category, “computer laboratory,” was rather broad, but was included because several studies involved the use of optional computer laboratories (e.g., Goolkasian, 1985; Goolkasian & Lee, 1988), with a variety of available types of software, as a supplement to class instruction. In these studies, students were allowed to use any of the software available in the computer laboratories, and it was not possible to distinguish the frequency with which various software programs were used. However, few studies were found overall that could be classified only in the “computer laboratory” category, so this category was later eliminated from the analyses.

Furthermore, several limitations exist when using meta-analytic techniques. For example, in what is termed the “file drawer problem” (Rosenthal, 1979, p. 638), published studies are much more easily obtained than unpublished studies, and often have larger effect sizes as a result of possible editorial biases toward studies with statistically significant results (Clark, 1983; Wolf, 1986). As such, studies without significant findings might be filed away rather than published. Second, studies can only be coded based on the amount and clarity of the information provided in the original manuscript (Glass, 1977). If information is vague, incorrect, or not provided, then the coding becomes flawed as a result.

Finally, a lack of agreement can exist in the coding of studies. Previous meta-analyses of instructional uses of technology in higher education have employed a variety of coding schemes, as well as defined coding categories differently. For example, earlier meta-analyses by authors such as Kulik et al. (1986) defined computer-assisted instruction (CAI) as uses of technology that provided drill-and-practice or tutorial instruction. In a more recent meta-analysis, Bayraktar (2002) included simulation software programs, in addition to drill-and-practice and tutorial programs, in his definition of CAI. Furthermore, Christmann et al. (1997b) defined computer-assisted instruction as "programmed learning using microcomputers" (p. 326). Additionally, authors such as Kulik et al. and Bayraktar have chosen to combine drill-and-practice and tutorial software programs into one category, while others, such as Kulik, Kulik, and Bangert-Drowns (1985) have chosen to examine these types separately.

Operational Definitions

The following definitions were used for the purpose of this study:

Computer laboratories. Computer laboratories are rooms filled with computers and potentially other technological devices where students are provided with access to a variety of computer hardware and software that they can generally use at their discretion. Often, computer laboratories include drill and practice, tutorial, simulation, and statistical analysis software programs, as well as access to email, Internet, and research and library databases.

Computer-Assisted Instruction (CAI). The term "computer-assisted instruction" or CAI has been used both generically to refer to a variety of uses of technology to enhance instruction, and as a specific type of computer use, namely programmed instruction using a computer. For example, Pear and Novak (1996) defined CAI as any program where the computer does the teaching directly. Kulik and Kulik (1986) commented that, "The marriage of computer technology and programmed instruction came to be known as computer-assisted instruction (CAI)" (p. 82). Other authors, such as Liao (1998), labeled generally supplemental software programs, such as drill and practice and tutorials, as CAI. In addition, Chambers and Sprecher (1980) referred to supplemental forms of computer-assisted instruction (CAI) as "adjunct CAI" and substitutive forms of CAI as "primary CAI" (p. 332), indicating that the authors viewed CAI as a fairly general term. In addition, CAI has often been used relatively synonymously with various other terms such as computer-assisted learning, computer-based instruction, computer-based learning, computer-managed instruction, computer-enhanced instruction, etc. For the purposes of this study, however, software programs that directly provide instruction and possess most of the qualities of programmed instruction or personalized system of instruction were labeled as programmed instruction rather than CAI.

Drill and practice. Drill and practice software programs are those that provide students with questions and elicit responses. Students often must answer correctly to proceed to the next question, or proceed after having answered incorrectly a set number of times. Drill and practice programs do not administer any course content.

Online Learning. Online learning courses are those where the instruction is predominantly provided in through an online medium, and may include both synchronous and asynchronous forms of communication, such as email, chat rooms, discussion boards, and so forth. In addition, online courses generally provide access to various course materials and provide links to additional resources. Furthermore, online learning courses may also be conducted using programmed or personalized systems of instruction that are provided through the Internet.

Personalized System of Instruction (PSI). A personalized system of instruction (PSI), first introduced by Keller in 1968, has five defining characteristics: PSI focuses on the written word instead of lectures, PSI requires students to master a unit before proceeding to the next one, students are allowed to proceed at their own individual pace, PSI courses make use of proctors or tutors that can help the students learn the material and provide feedback as needed, and lectures and demonstrations are used as motivational tools, not instructional sessions.

Programmed instruction. Programmed instruction, based on the theories of B. F. Skinner, is designed with clearly stated behavioral objectives, small frames of instruction, require overt responses, provide immediate feedback, and allow self-paced learning (Garson, 1999).

Simulations. Simulation software programs allow students to explore relationships among variables in models that replicate real-world phenomena (J. A. Kulik, C. C. Kulik, & Cohen, 1980). Simulations allow greater interactivity than tutorial or drill and practice software programs, and include virtual reality software programs and some educational games.

Statistical Analysis Software. Statistical analysis software programs are those that can be used to automate calculations, provide output for statistical analyses, and develop charts and graphs of data. Statistical analysis software programs include *SPSS*, *SAS*, *Microsoft® Excel* and other spreadsheet programs, *Minitab*, and so forth.

Substitutive. Substitutive uses of technology include those where a computer software program presents the majority of the content, and the instructor's purpose is to set guidelines and policies, provide technological assistance, and to answer students' questions.

Supplemental. Supplemental uses of technology include those that are used as an adjunct or enhancement to classroom lecture. However, the instructor of the course provides the predominant amount of instruction.

Technology. Technology refers to computers and other digital and electronic devices that can be incorporated into instruction. Uses of technology include, but are not limited to, laptop and desktop computers, handheld devices such as *Palm Pilots* or *BlackBerry* devices, computer software, presentation software and hardware, student response systems, Internet access, and so forth. Devices which are not digital or electronic in nature, such as some types of teaching machines, were not be included under the definition technology.

Technology-enhanced lectures. Technology-enhanced lectures are educational content delivery systems that often include multimedia, such as images, sounds, short videos, and so forth. Technology-enhanced lectures may include the use of presentation software such as *Microsoft PowerPoint*® or the use of student response systems, where students are asked questions during the lecture and are provided a small, handheld device in which they can respond.

Tutorials. Tutorial software programs are those that provide students with some instructional material, often accompanied with questions over the material. Like drill and practice software programs, students must often answer the questions correctly to proceed to the next unit. Tutorial programs, unlike programmed instruction, provide instruction over a single unit or several short units, and are often, but not always, used to reinforce material presented in lectures.

Web-based. Web-based technologies are those that use the Internet to allow greater access to resources and instructional materials, as well as greater communication among students and between students and the instructor. Web-based technologies might include email, class web pages with course syllabi and assignments, electronic bulletin boards and chat rooms, access to library databases or catalogs, links to additional resources, teleconferencing, web-page development, newsgroups, and so forth.

CHAPTER II

REVIEW OF THE LITERATURE

History of Technology in Education

Teaching Machines and Programmed Instruction

The use of technology in education in the 20^{th} century can be traced back to the development of teaching machines as early as 1915 (Benjamin, 1988). According to Benjamin, Sidney Pressey created one of the earliest teaching machines, and exhibited a working model in 1924 at a meeting of the American Psychological Association. The machine was designed so that students who worked on it could not proceed to the next question until they answered the current one correctly. Pressey developed a second machine in 1927, which differed from the first in that it would drop questions that were answered correctly twice in succession. Whereas Pressey never completely gave up work on teaching machines, he met with little success in promoting and selling them. Little research was conducted at the time beyond a small group of Ohio State University doctoral students. B. F. Skinner believed that the educational establishment was not ready for them at the time.

The programmed instruction movement began in the mid-1950s and ran through the mid-1960s (Reiser, 2001). Based on the theories of B. F. Skinner, programmed instruction was designed with clearly stated behavioral objectives, with small frames of instruction, to require overt responses, to provide immediate feedback, and to allow self-paced learning (Garson, 1999). Learners must also answer all questions correctly before moving on to the next section (Reiser). Skinner began working on teaching machines in 1953 after visiting his daughter's fourth grade class (Benjamin, 1988). He was critical of two practices he observed: All students had to proceed at the same pace and they often had to wait 24 hours for feedback. Although Skinner was aware of Pressey's teaching machines, he felt they were primarily testing machines rather than machines that would administer new material. Skinner designed his teaching machines so that they would provide new material in small increments.

Skinner's machines also differed from Pressey's in that he avoided the use of multiple-choice questions (Benjamin, 1988). Students were expected to write a response, and then slide a panel over that revealed the correct response. Skinner felt that multiple-choice questions were unnecessarily confusing due to the inclusion of incorrect answers, constructing a response was a behavior that had more utility in life over selecting one, and students should progress beyond simply being able to recognize that a response is correct to being able to emit a response.

The use of teaching machines was well underway in the early 1960s, with several corporations creating machines based on Skinner's theory of learning (Benjamin, 1988). However, as the popularity of teaching machines increased, so did the criticisms that were levied against them. Chief among the complaints was that the machines dehumanized students, foreshadowing a future where students would be taught by robots. Critics believed that students would not receive enough personal attention in the classroom when teaching machines are used. In addition, critics felt that not all subjects (e.g., English Literature) could be taught adequately by machines, and that teachers might not be adequately trained to implement the machines correctly. Furthermore, critics felt that the use of teaching machines would result in a reduction in the overall number of teachers, and an increase in the student-teacher ratio. However, Skinner and other proponents replied that the machines provided a record of students' performance, and that children and adults appeared to enjoy using the machines. For example, Kulik, Cohen, and Ebeling (1980) found an effect size of 0.28 when programmed instruction was compared to traditional instruction on examination performance, and found that programmed instruction led to an average of one less hour per week of time spent on instruction.

By the late 1960s, many teaching machines had been withdrawn from the market, programs on teaching machines became scarce, and articles debating their value had begun to disappear (Benjamin, 1988). This resulted in a shift in interest among some researchers from teaching machines to the newly developing field of computer-assisted instruction (CAI). According to Benjamin, "CAI was an outgrowth of the Teaching Machine Project at the IBM Research Center in the late 1950's" (p. 710). Kulik and Kulik (1986) referred to CAI as the "marriage of computer technology and programmed instruction" (p. 82). CAI was developed in the 1950s and 1960s at Florida State University, where several courses were administered entirely through CAI, including physics and statistics (Chambers & Sprecher, 1980). In addition, in the 1970s Dartmouth was the source of one of the first adjunct (i.e., supplemental) CAI programs, in conjunction with the Universities of Oregon, North Carolina, Iowa, and Texas, who formed a consortium called CONDUIT. The purpose of CONDUIT was to acquire, evaluate, and distribute instructional computing materials nationally, and was supported by the National Science Foundation.

Another early example of the use of computer-assisted instruction in higher education was the PLATO system (Programmed Logic for Automatic-Teaching Operations), developed in 1960 at the University of Illinois (Garson, 1999). Using PLATO, students could be tested repeatedly for understanding and then be prescribed additional materials for remediation or enrichment (Garson). The PLATO computer system was designed to provide interactive, self-paced instruction to large numbers of students (S. G. Smith & Sherwood, 1976). Students would receive instantaneous reinforcement for correct responses and assistance when they are having difficulties. At the time when Smith and Sherwood's article was written, the PLATO system had 950 terminals located in colleges, universities, community colleges, public schools, military training schools, and commercial organizations. Users had access to more than 3500 hours of instructional material in more than 100 subject areas.

S. G. Smith and Sherwood (1976) further described how PLATO operated in science education and research courses. Students in these classes were asked to design experiments and interpret data, whereas more advanced students were asked to design an experiment to solve a problem they developed. Students using PLATO were presented the material in unique ways that encouraged the student to become an active participant in his or her learning. The PLATO system consisted of three main components: instructional lessons, homework, and an on-line gradebook. Furthermore, PLATO automatically graded assignments, rather than having them graded by the instructor. In addition to Smith and Sherwood, Paden, Dalgaard, and Barr (1977) discussed the use of the PLATO system to teach a Principles of Economics course at the University of Illinois. According to the authors, PLATO "presents material to students, asks them questions, judges their answers, responds to their errors, allows them to proceed (or requires them to review), and stores information about their responses" (p. 14). However, Garson (1999) stated that PLATO often made the learning process tedious for students, with long lists of behavioral objectives, linear presentation of material, and long testing sequences. Despite some limited success, the business model was found to be unprofitable, and PLATO was shut down in the early 1990s.

Also occurring during the time of the programmed instruction movement was the development of personalized systems of instruction or the Keller plan (Keller, 1968). Personalized systems of instruction have several distinguishing characteristics. First, they focus on the written word instead of lectures (Grant & Spencer, 2003; Keller, 1968). PSI instructors are expected to create a syllabus that guides students on course objectives, what material is to be learned, what is expected of them, and possibly a supplementary reading list or practice problems. Second, PSI requires students to master a unit before proceeding to the next one. Students who attempt a quiz are not able to move to another unit until they obtain at or above a set score, typically 80 or 90%. Also, grades are not used to rank students but to promote achievement. Anyone who finishes all units should receive a passing grade. Third, students are allowed to proceed at their own individual pace. Slower students are not penalized for taking longer than others, and faster students are not held back by slower ones. Furthermore, PSI courses make use of proctors or tutors who can help the students learn the material and provide feedback as needed. Finally, in PSI, lectures and demonstrations are used as motivational tools, not instructional sessions.

An abundance of research on PSI was conducted throughout the early 1970s (Grant & Spencer, 2003). Researchers, such as Kulik, Kulik, and Carmichael (1974) and Kulik, Kulik, and Cohen (1979), found that students learned at least as much or more when taught using PSI, and they rated PSI courses more favorably than traditional courses in that they found them more enjoyable, more demanding, and higher in quality, and they believed they learned more. Furthermore, Kulik et al. found, in a meta-analysis of PSI, an effect size of 0.49 when final examination performance was used as an outcome measure, and 0.69 when final grades were used. However, interest in PSI development and research diminished in the 1980s. According to Grant and Spencer (2003), educators had several misconceptions about PSI. First, some believed that PSI is ill-suited to teach higher-order skills such as those involved in concept learning and critical thinking. The authors argued, however, that PSI is well-suited to teach any content for which observable assessments of performance can be made. If not, one has difficulty determining if any method is well-suited. To further bolster their argument, the authors cited research that has shown that positive results have been obtained using PSI to teach critical thinking (Ross & Semb, 1981, as cited in Grant & Spencer), as well as other higher-order skills. A second misconception was that PSI is incompatible with philosophical and theoretic viewpoints other than behaviorism. However, Grant and

Spencer claimed, studies of PSI have occurred in a variety of theoretical contexts, and PSI can even be consistent with constructivist learning approaches if, for example, discovery learning experiences are incorporated. Also, PSI is learner-centered in that students take an active role in their learning rather than passively receiving the information as in a lecture format.

However, despite research that supported the effectiveness of PSI, interest in the methodology declined (Grant & Spencer, 2003). According to Grant and Spencer, some reasons for this decline include resistance to change throughout the educational establishment, ineffective teaching methods that were implemented under the label of PSI but did not meet the requirements of PSI, time demanded to set up and maintain a PSI course, misunderstanding of PSI in the academic literature, and outright prohibitions of PSI courses in some educational institutions. Furthermore, the authors argued that perhaps the most important reason for diminished interest was the lack of concern for empirical evidence of student achievement and satisfaction when choosing instructional methods. However, despite the decline in interest, some forms of PSI are still in use today, particularly in conjunction with online or web-based learning (Harrington, 1999; Pear & Novak, 1996).

Throughout the 1970s and early 1980s, computers were expensive and not particularly user-friendly (Benjamin, 1988). However, improvements in hardware and software, combined with reduced costs, made computers commonplace in classrooms in the 1980s. At that point, the "latest generation of teaching machine [continued] to be an adjunct to classroom instruction," according to Benjamin (p. 711). Teachers used them as an aid to teaching rather than an integrated part of the curriculum. In addition, some of the same criticisms (i.e., dehumanization) that were applied to teaching machines were also applied to computers. Furthermore, several technological problems existed, such as a lack of quality educational software, compatibility issues across computer platforms, and problems with recording of student performance (Pagliaro, 1983, as cited by Benjamin, 1988).

In 1983, the Massachusetts Institute of Technology (MIT) instituted Project Athena as an experiment to explore the potential uses of advanced computer technology in the university curriculum (Balkovich et al., 1985). Athena was a campus-wide network of time-shared computers with public and private workstations. All workstations had the same operating system and general-purpose applications and libraries. However, significant training to learn UNIX was required. The authors described several potential uses of computer technology for instruction, including as a simulator, laboratory instrument, virtual laboratory, tutor, electronic textbook, electronic blackboard, special-purpose learning environment, mediator, communications medium, and recreational device. Furthermore, computers were often used for a variety of additional purposes in higher education including computer science and electrical engineering instruction and research, financial management and record keeping, word processing, spreadsheet analysis, and database management, and electronic mail and other forms of asynchronous communication.

Current Uses of Technology for Instruction in Higher Education

Whereas some methods of computer use in higher education, such as PLATO, have been discontinued, several methods have remained or emerged. For example, technology has been implemented increasingly for administrative purposes throughout higher education, and has already proven cost-effective (Van Dusen, 1998). In addition, Van Dusen claimed that statistics show increased use of technology in colleges and universities including increased use of technology in traditional classrooms, increased enrollment in distance education courses, and increased use of two-way interactive video or teleconferencing. Miller et al. (2000) claimed that the most common uses of instructional technology throughout higher education include “enhanced presentations, simulations, computer tools, collaboration and communications, access to support materials, research tools, and evaluation and testing” (p. 235). In addition, some universities create web pages for every course where class notes, syllabi, schedules, and assignments can be posted. Furthermore, instructors are making greater use of simulations to enhance pedagogy (Garson, 1999), as the use of such software programs can be more cost-effective, more practical, or safer than conducting genuine experiments (Miller et al., 2000). Also, some universities have made greater use of wireless Internet access in order to allow students to be able to study and work in less restrictive, more social environments (Brett & Nagra, 2005). Finally, faculty members, such as Hyden (2005), have attempted to build learning communities that surpassed the boundaries of the classroom by incorporating the use of laptop computers into their curriculum.

Despite the amount of money spent on technology across colleges and universities, technological innovations for teaching and learning have not been widely integrated into the curriculum (Geoghegan, 1994), and many instructors only use such technology for fairly "low tech" purposes, such as substitutes for blackboards, overhead projectors, or handouts (Reeves, 1991, as cited in Jacobsen, 1998). Upon being surveyed, higher education faculty members often admit to only using technology for fairly low-level purposes. For example, Groves and Zemel (2000) surveyed faculty and graduate teaching assistants/associates from one university to determine use of, interest in, and attitudes toward technology. The researchers asked about participants' knowledge and use of technology, factors influencing their use of technology, and barriers to technology use. Participants generally expressed that they were familiar with word processing and e-mail programs, and least familiar with electronic bulletin boards and distance learning. In addition, they stated that they commonly used word processing, the Internet, presentation software, e-mail, and spreadsheet software in their classes, whereas they less frequently used statistical computing software, discussion lists, multimedia, computer-aided instruction, and distance learning. Among the factors most commonly cited influencing instructional technology use were equipment availability, improved student learning, increased student interest, advantage of traditional teaching methods, and ease of use. The least commonly cited factor influencing the use of instructional technology was the frequency of use by colleagues.

Peluchette and Rust (2005) administered a more recent and widespread survey of faculty members' preferences for using technology as an instructional tool. They mailed surveys to 500 management faculty members throughout the country, with 124 usable surveys returned. The results indicated that management faculty most often used PowerPoint® presentations combined with whiteboards or blackboards. Faculty members displayed the least preference for fully online courses and no technology use. However, the majority of faculty members did indicate that technology enhanced their teaching effectiveness. In addition, most of the participants indicated that their universities and department heads were supportive of the use of technology in the classroom. Overall, the results indicated that most of the respondents used some technology, felt comfortable with learning new technologies, and felt that technology enhanced their teaching. However, because response was voluntary, those who were more comfortable with technology or used technology regularly could have responded more often. Also, the authors concluded that the technologies most often chosen by faculty members were fairly “low tech.” The participants frequently cited time constraints as a common factor inhibiting their willingness to implement technology.

Furthermore, Vodanovich and Piotrowski (2001) conducted a survey to determine psychology faculty members’ attitudes, usage patterns, and perceived drawbacks to using web-based instruction. The authors found that psychology faculty generally held positive attitudes toward the Internet and the use of online technology for instructional purposes. However, the faculty members did indicate that they did not have access to formal training. Instructors most often used the Internet for email, providing course syllabi, and assessing professional literature. Barriers to using Internet technology included lack of time and technical difficulties. The authors concluded that, despite faculty members’ generally positive attitudes toward Internet technology, they most often used it for fairly low-level purposes. They have yet to begin using it for assessment or interactive teaching. In addition, when Strasser and Ozgur (1995) surveyed business faculty members, they found that, when asked for changes they would like to make to statistics instruction, the greatest number responded that they would like to incorporate greater use of computers and real, relevant data sets. Studies such as these suggest that most faculty members are supportive of the idea of incorporating technology into their instruction, but may experience some barriers to doing so.

Technologies that can be used to enhance instruction play both supplemental and substitutive roles. Chambers and Sprecher (1980) referred to supplemental forms of computer-assisted instruction (CAI) as “adjunct CAI” and substitutive forms of CAI as “primary CAI” (p. 332). Adjunct CAI supplements instruction and includes short programs that can be used as tutorials to reinforce concepts that are also presented in lectures. Primary CAI programs, however, tend to be longer in duration and serve as substitutes for classroom instruction. Primary CAI programs may cover an entire course rather than one or two lessons. However, other authors categorize only those programs that provide substitutive instruction under the label of CAI (Kulik & Kulik, 1986; Pear & Novak, 1996), whereas authors such as Liao (1998) labeled generally supplemental software programs, such as drill and practice and tutorials, as CAI. However, regardless of the labels used, technology can clearly be used to both supplement classroom instruction or as a substitute for classroom instruction.

While various types of software use can be used for either supplemental or substitutive purposes, supplemental uses of technology generally include assessment, drill and practice, tutorial, simulations, computer laboratories, technology-enhanced lectures, and some web-based uses. Substitutive forms of technology use often include programmed and personalized systems of instruction and some forms of fully online learning. Technologies used for assessment are simply those where the instructor uses technology to administer mid-term or final examinations. Drill and practice technologies are similar to assessment technologies in that students are presented questions over the course material. However, students are generally allowed to use the software as often as they wish at their own discretion in order to master the material. Tutorial programs are similar to drill and practice programs, except that they may include course content for review, as well as questions over the material. Simulation software programs are more interactive and allow students to perform tasks such as conduct experiments or manipulate data sets and sampling distributions. Computer laboratories generally offer students access to a variety of software programs and network access that can be used as needed. Technology-enhanced lectures are those that include multimedia or presentation software that can make lectures more interesting and interactive. Supplemental forms of web-based instruction simply include the incorporation of several aspects of online technology into instruction such as email, discussion forums, chat rooms, web pages, videoconferencing, and links to additional resources, and so forth. Programmed or personalized systems of instruction are those software programs that administer entire courses, and generally substitute rather than supplement the instructor’s instruction.

Finally, fully online courses are those that are taught entirely or almost entirely online, often with the assistance of one or more programmed instructional software programs. However, not all uses of technology fall neatly into these categories. For example, some fully online courses may incorporate a variety of drill and practice, tutorial, or simulation software programs.

Supplemental Uses of Technology in Higher Education Instruction

Assessment or drill and practice. One common use of technology in higher education is the inclusion of software programs that assess students or provide them with drill and practice over course material. Occasionally, this software has been included in computer laboratories along with a variety of additional software (Forsyth & Archer, 1997; Goolkasian, 1985). Although few studies exist where instructors used technology for assessment only, Lloyd and Martin (1996) conducted such a study using computer-based testing for an engineering technology course, and Thoennessen and Harrison (1996) conducted a study using computer-assisted technology for homework assignments and examinations. However, the vast majority of researchers who implemented this type of software used it for drill and practice purposes. For example, in an early study using drill and practice software, Suppes and Morningstar (1969) compared experimental groups to control groups on their relative rates of achievement based on their participation in one of two programs in studies at or near Stanford University. Whereas the researchers did not use higher education students as subjects, they administered a mathematics program designed to provide drill and practice in the skills of arithmetic as a supplement to regular classroom instruction. When comparing the achievement results to control groups in the first year, the authors obtained mixed results, with the experimental group not always significantly outperforming the control group on all measures. However, the researchers conducted the study over several years, and obtained a more consistent pattern of results toward the end of the study, with the experimental groups significantly outperforming the control groups in most grades.

More recent studies have found results more supportive of the use of drill and practice software programs. For example, Gretes and Green (2000) provided students with the opportunity to participate in computerized practice testing for one week before each of two mid-term examinations and the final examination as part of an instructional design and evaluation course. The authors found that those who participated in the practice session significantly outperformed those who did not, and generally felt that the sessions were worthwhile and helped them gain confidence. In addition, Luyben, Hipworth, and Pappas (2003) compared the effects of a CAI instructional package that incorporated mastery quizzing and fluency requirements with a traditional, lecture-based instructional method using undergraduate students enrolled in psychology courses. The authors used “a crossover design with repeated measures and counterbalancing of conditions across groups” (p. 155). They found that students obtained higher scores in the CAI condition than the traditional instruction condition, as well as exhibited more favorable attitudes in the CAI condition. However, meta-analyses have not always exhibited positive results for drill and practice software programs. For example, Bayraktar (2002) found an effect size of -0.107 for drill and practice when used for science instruction, whereas effect sizes of 0.391 and 0.369 were found for simulations and tutorials respectively. However, meta-analyses that involved subjects other than science found positive effect sizes for drill and practice (e.g., Niemiec & Walberg, 1985).

Tutorials. Tutorial software programs are similar to drill and practice software programs in that they are used as a supplement to instruction and they provide students with questions over course material and elicit responses (Balkovich et al., 1985). However, tutorial software programs differ in that they also administer some course content, whereas drill and practice programs do not. Chambers and Sprecher (1980) defined tutorials as short programs that can be used to reinforce concepts that are discussed in lectures. For example, Frith, Jaftha, and Prince (2004), who used spreadsheets as tutorials for the presentation of mathematics content, stated that, "a typical tutorial will consist of several electronic Excel worksheets, some containing the interactive presentation of relevant mathematics content and other comprising examples and exercises" (p. 161). Such tutorials allow the students to work at their individual speeds and provide immediate feedback. Furthermore, Ben-Zvi (2000) referred to tutorials as programs designed to teach or tutor students on specific content or to test their knowledge of this content. The goal of tutorials is to take over a portion of the teacher's duties by providing information, setting tasks, evaluation student responses, and providing feedback.

The use of tutorials is often optional, or at least not necessary for success in a given course. For example, tutorial programs are often included in computer laboratories where students are allowed a wide variety of software programs that suit their individual needs (Forsyth & Archer, 1997; Worthington, Welsh, Archer, Mindes, & Forsyth, 1996) or included in websites that provide a variety of additional resources, such as Wilson and Harris' (2002) evaluation of "The Psychology Place," where the authors found that students who were randomly assigned to use the website generally outperformed those who were not. In addition, some researchers have examined student use and attitudes toward optional versus mandatory use of tutorials. For example, Garland and Noyes (2004) found that the students rated an economics tutorial software program useful regardless of whether use was mandatory or optional, but mandatory users provided higher ratings. However, the most common reason for not using the software among the optional group was that they were not provided any instruction in its use nor had its usefulness demonstrated to them. This indicates that optional tutorials may be used more if students are instructed in how to use them. Furthermore, Bayraktar (2002) found an effect size of 0.369 for the instructional use of tutorial software programs, whereas Liao (1998) found an effect size of 0.436.

Simulations. Simulation software programs, unlike drill and practice or tutorial programs, allow for a greater level of interactivity for the user. Simulation software programs allow students to explore relationships among variables in models that replicate real-world phenomena (J. A. Kulik, C. C. Kulik, & Cohen, 1980), or examine physical or social phenomena that may be too complex to understand from mathematical formulations alone (Balkovich et al., 1985). In addition, newer virtual-reality simulation technologies can place a person in a given time or place, rather than have an instructor describe it in a lecture (Levine, 2001). Furthermore, software programs have been used in psychology courses to simulate psychological procedures or classical experiments

(Worthington et al., 1996). In other psychology courses, simulation programs have allowed students to simulate research on animals, and in one case a computer laboratory replaced an animal laboratory (Goolkasian & Lee, 1988).

The results of studies where simulation software programs were examined have generally been positive. For example, Lee (1999) conducted a meta-analysis to determine the effect of the classroom use of computer-based instructional simulations on attitudes and achievement, and found an overall mean effect size of 0.41 for achievement from 51 effect sizes across 19 studies. However, the mean effect size for attitudes was near zero at -0.04 from 5 effect sizes. In addition, Khalili and Shashaani (1994) conducted a meta-analysis of the effectiveness of computer applications on students' academic achievement and found simulations to have a larger effect size (0.79) than either drill and practice (0.11) or problem-solving (0.41) software programs. A similar result was found by Bayraktar (2002), who obtained an effect size of 0.391 for simulations, larger than those obtained for drill and practice and tutorial software programs. Furthermore, Liao (1998) found a rather large effect size for simulation software of 0.974. However, not all studies using simulations have found positive results. For example, Michael (2001) conducted a study to compare a hands-on activity to a simulated activity using a computer. Specifically, the author wanted to compare the product creativity of those using real building blocks compared to those using a building block simulation program, but found no difference between groups on overall creativity, originality, or usefulness of the products created by the study participants.

Computer laboratories. Another common use of technology in higher education is the creation of computer laboratories, often complete with productivity software, Internet browsers, email software, data analysis software, and specialized course software. Such computer laboratories have been in operation for quite some time. For example, Goolkasian (1985) and Goolkasian and Lee (1988) described the implementation of a computer laboratory in a psychology department in the mid-1980s. The laboratory contained simulations, statistical analysis software, and drill and practice instructional programs designed to assist with the instruction of various psychology courses, including introductory psychology and research methods courses. The simulations included classic experiments and software that allowed them to simulate research on animals. In addition, Raymondo (1996) assisted in creating a computer laboratory for undergraduate sociology courses in order to help students develop appropriate skills to function in a technological age and cover more material in a shorter time. Furthermore, Forsyth and Archer (1997) conducted a study where students were expected to attend a one hour session per week in a technology-enhanced classroom that included basic productivity software programs, network tools, and psychology-related tools. Students were also permitted and encouraged to use the classroom voluntarily at their discretion. The authors found that students who made greater optional use of the classroom outperformed those who made lesser use of the classroom in course grades. However, students that made use of the technology-enhanced classroom did not experience greater gains in performance than earlier cohorts of students who did not have access to the classroom. Finally, a more modern example of a computer laboratory was described by Brett and Nagra (2005) who examined a social learning space that contained 24 computers arranged on circular tables, wireless Internet access, and a coffee bar, and had no restrictions on talking, eating, or drinking. The authors found that students often used the learning space for both academic and social purposes, and that it promoted collaborative study.

Enhanced lectures. A further use of technology in higher education is to enhance classroom lectures, often through the use of multimedia. One advantage of technology-enhanced lectures is that computer-generated graphics can be more elaborate and clearer than drawings on a blackboard (Balkovich et al., 1985). In addition, researchers such as Erwin and Rieppi (1999) felt that multimedia-enhanced environment increase student interest and motivation, and allow them to become more active learners. They conducted a study of a classroom that incorporated multimedia in the form of text, graphics, and sounds, such as computer-generated animations and brief video clips. The classroom also contained keypads that allowed students to provide feedback. The authors found that students in the multimedia classroom obtained higher final examination scores than those in traditional classrooms. Furthermore, DeBord, Aruguete, and Muhlig (2004) compared a class where the lecture was supplemented with computer-assisted visuals to a class that received the lecture along with overhead transparencies. Students in the experimental condition more strongly preferred computer-assisted presentations over traditional lecture methods, but they did not perform significantly different from those in the control group on final examination scores or course grades. Whereas surveys have shown that instructors infrequently make use of technology to enhance their lectures (Jacobsen, 1998), Rickman and Grudzinski (2000) found that students view the use of presentation software and document cameras more useful than the use of Internet sites or videocassette tapes in lectures. However, the students also felt that technology must enhance the instructional experience, not overshadow or replace it.

Also, an increasingly fashionable use of technology to enhance lectures involves the use of student response systems (SRS). For example, Fitch (2004) conducted a pilot study on the use of a student response system in a college class covering the topic of communication disorders. The student response system allowed the instructor to present a question on a screen, followed by a set of answers. The students then had limited time to respond, and could do so using individual keypads. After the time expired, the instructor could present tallies of student responses followed by the correct answer. Fitch administered a survey designed to measure student reactions to the student response system and found that students typically viewed the technology positively, although they stated that they did not come to class more prepared knowing that the technology would be used. The author concluded that the use of student response systems promoted focus, provided feedback for both the student and the instructor in that it allowed the instructor to find what areas in which students were deficient, and the students enjoyed it.

Web-based uses of technology. Web-based technologies can be used to either supplement classroom instruction or act as a substitute for it. When used as a supplement, web-based technologies are most commonly used as communication media, such as email, class web pages with course syllabi and assignments, electronic bulletin boards, access to library databases or catalogs, or links to additional information (Balkovich et al., 1985; Vodanovich & Piotrowski, 2001). Research has shown that professors have made increased use of Internet technology in their traditional classrooms, as well as increased use of two-way interactive video or teleconferencing (Van Dusen, 1998). In addition, Jacobsen (1998) conducted a survey of faculty members at two North American universities concerning their technology use, and found that faculty members most often used email and Internet searching and browsing software programs. However, they least often used web page creation software and newsgroups. Furthermore, Groves and Zemel (2000) conducted a survey of faculty and graduate teaching assistants at one university and found that they commonly used the Internet and e-mail in their classes, whereas they less frequently used discussion lists and distance learning. Vodanovich and Piotrowski (2001) found, through a national survey of psychology faculty, that whereas faculty members generally possessed positive attitudes toward the inclusion of Internet technology into their instruction, professors most often used Internet technology for fairly low-level purposes. Few professors used the Internet for more advanced purposes such as assessment or interactive teaching.

Students often expect some use of Internet technology in their classes, and generally find it useful. For example, Forsyth and Archer (1997) found that students viewed a computer-based intervention as a positive experience, and especially liked the enhanced online access to information resources and email feedback. In addition, Mitra and Steffensmeier (2000) conducted a longitudinal (3 year) study of students' attitudes toward computers after a university implemented an extensive computerization project campus-wide. Over the course of the study, the researchers found that students increasingly used the Internet to gather information and email instructors, and found that students became more comfortable with technology and found interacting with teachers easier. Furthermore, other authors commented that students can become more collaborative with the use of Internet technology, as online discussions can often have a higher level of participation and be of a higher quality than those in the classroom (Van Dusen, 1998).

Studies of the effects of supplemental web-based uses of technology for instruction on student achievement have shown these uses to be at least as, or more effective than, traditional forms of instruction. For example, DeBord et al. (2004) compared the final grades of students who made use of an optional supplementary website to those who did not. The website offered class notes, a practice quiz, the class schedule, a copy of the syllabus, and study guides. The authors found that the students in the experimental and control conditions did not differ significantly on final exam scores and final course grades, but those in the experimental group did rate the website more positively than those in the control group. Overall, the results indicated that computer-assisted learning was no worse than traditional instruction approaches. In addition, Wilson and Harris (2002) conducted a study where they incorporated assignments from a psychology website containing tutorials, readings, and links, into an undergraduate psychology class. The authors used a matched-groups, counterbalanced design, where one group acted as the experimental group on assignments 1, 4, and 5, whereas the other acted as the experimental group on assignments 2 and 3. The course content was presented both during lectures and through the assignments on the psychology website. The results from five quizzes over each section indicated that students in the experimental group significantly outperformed the members of the control group on 4 out of 5 quizzes. Furthermore, Olson and Wisher (2002) conducted a meta-analysis of the effectiveness of web-based instruction on student achievement and found a moderate average effect size of 0.48 for studies that blended web-based instruction with face-to-face instruction.

Substitutive Uses of Technology in Higher Education Instruction

Programmed or personalized systems of instruction. Technologies that substitute for classroom lecture include programmed and personalized systems of instruction, as well as similar instructional programs. Some authors generically refer to some of these additional programs as computer-assisted instruction (CAI). For example, Pear and Novak (1996) defined CAI as any program where the computer does the teaching directly. Kulik and Kulik (1986) commented, “The marriage of computer technology and programmed instruction came to be known as computer-assisted instruction (CAI)” (p. 82). However, not all authors define CAI in this manner. For example, Chambers and Sprecher (1980) created two categories of CAI: adjunct and primary. Adjunct CAI supplements instruction, and includes software such as tutorials. Primary CAI programs are longer in duration and serve as substitutes for classroom instruction.

For the purposes of this study, software programs that directly provide instruction were included in the programmed or personalized systems of instruction category. Programmed instruction programs, based on the theories of B. F. Skinner, are designed with clearly stated behavioral objectives, small frames of instruction, require overt responses, provide immediate feedback, and allow self-paced learning (Garson, 1999).

Likewise, personalized systems of instruction (PSI), first introduced by Keller in 1968, have several defining characteristics (Grant & Spencer, 2003). First, it focuses on the written word instead of lectures. Second, PSI requires students to master a unit before proceeding to the next one. Third, students are allowed to proceed at their own individual pace. In addition, PSI courses make use of proctors or tutors that can help the students learn the material and provide feedback as needed. Finally, in PSI, lectures and demonstrations are used as motivational tools, not instructional sessions. Several instructional programs labeled as CAI share most, if not all, of the qualities of either programmed instruction or personalized systems of instruction.

For example, an early CAI program at Stanford University was highly influenced by the programmed instruction and PSI movements (Suppes & Morningstar, 1969). The CAI program was designed to instruct first and second year college students in the Russian language. The program required students to spend about 50 minutes a day, 5 days a week, in front of a computer terminal, for a total of 135 lessons. In other words, this program was a substitute for lecture rather than a supplement to it. In a later study by Eisenberg (1986), students in a teacher training college enrolled in a Jewish studies course completed the course using computer programs that, like PSI and programmed instruction, administered the lessons, quizzed the students, and provided feedback. Furthermore, Welsh and Null (1991) defined computer-based instruction as computer programs that instruct students without aid provided or discussion initiated by an instructor. Several forms of programmed and personalized systems of instruction have been implemented in recent years, such as combining PSI with online learning (Grant &

Spencer, 2003; Pear & Novak, 1996), and including multimedia with textual delivery, followed by multiple-choice and short-answer questions (Lynch, Steele, Palensky, Lacy, & Duffy, 2001).

Studies of programmed or personalized systems of instruction have yielded mixed results. Most studies have shown that students have positive attitudes toward CAI. For example, Eisenberg (1986) found that the vast majority of the students enrolled in the CAI program responded that they found the course interesting or very interesting (96%), and indicated that they would like to take another course by computer (92%). The characteristics of CAI that students found most beneficial were that they could learn at any time, they must master the material before proceeding, the computer provided immediate feedback, responses were confidential, everyone could participate rather than a few dominating students, the pace was fair and reasonable, and they generally found it enjoyable. In a second study, Rainbow and Sadler-Smith (2003) administered a survey to assess business students' attitudes toward CAI. They found that the majority of students found the CAI system easy to use and allowed them to organize their learning experiences. However, they found that the students were much more ambivalent about CAI in terms of enjoyment. Furthermore, Pear and Novak (1996) conducted a study of a CAI version of PSI administered both online and on-campus for two undergraduate psychology courses, and found that, overall, students liked the class because of the lack of regular class sessions, the ability to work at their own pace, and the convenience of being able to work where and when they chose. However, students often disliked the lack of interaction with other classmates and/or the instruction. A majority of the students were satisfied with the course and a large minority thought it was better than a traditional lecture course. Students also reported a higher level of comfort with computers at the end of the course than at the beginning.

Although the results of surveys of students' attitudes toward CAI have generally exhibited positive results, its effects on achievement are much less clear. For example, Suppes and Morningstar (1969) examined the effects of a CAI program that provided instruction in the Russian language on student achievement, and found that students in the CAI course had significantly fewer errors than students in the control course on the final examination in two of the three quarters under which the program was studied. In addition, students in the second-year Russian course significantly outperformed the control group in course grades. Also, when compared to a control group, the authors found a lower withdrawal rate for those in the CAI course. However, Welsh and Null (1991) conducted a study to examine the effect of computer-based instruction in an advanced laboratory in cognition and thinking and found different results. The computer program used allowed students to participate as subjects of well-known studies, view the results, and learn about the theoretical background. In the first experiment, the participants were assigned randomly to groups, with one group being subject to the experimental condition for two weeks, followed by the other group. The authors, however, found no significant differences between the two groups on a quiz containing questions about the content covered for either experimental intervention. In the second experiment, students enrolled in experimental psychology participated in the study. The procedure was the same as the first study, with manipulation checks added. In this case, the authors found that the control group significantly outperformed the experimental group on the quiz. However, other studies have shown that the use of CAI has resulted in a reduction in instructional time (Lewis, Dalgaard, & Boyer, 1985).

Online learning. Fully online courses are those that are taught entirely at a distance with few or no face-to-face meetings. Often, online courses include forums for posting announcements, assignments, lecture notes, course documents, and links to additional resources, as well as email, discussion boards, and chat rooms for synchronous and asynchronous communications (Beard, Harper, & Riley, 2004). For example, Pear and Novak (1996) created an online version of a PSI course. The characteristics of the course were that the instructor selected the course materials, the study objectives and questions were based on the assigned readings, short essay tests were presented by computer over each unit, students were provided with rapid feedback, and students were provided with remediation and retesting if they failed to meet a mastery criterion on a unit test. Overall, students liked the class because of the lack of regular class sessions, the ability to work at their own pace, and the convenience of being able to work where and when they chose. Likewise, Lawson (2000) created a web-based version of a social psychology course that he attempted to make as consistent with the on-campus version as possible. He used the same textbook, posted lecture notes on the web site, and included many of the same in-class discussions, demonstrations, and exercises. The online course contained sections that included course content, chat rooms, bulletin boards, online quizzes and surveys, course tools, links to related web sites, and a class list. The course content module included lecture notes, interactive exercises, exam study guides, links to practice tests, and instructions. Like Pear and Novak, Lawson found that students liked the convenience and flexibility of the online format.

When compared to traditional lecture formats on student achievement, online courses are generally found to be similarly effective. For example, Warren and Holloman (2005) conducted a study comparing a teacher leadership and communication course taught online to one taught face-to-face. They found no significant difference between the two methods on student performance based on student portfolio ratings, indicating that online courses can be just as effective as face-to-face courses. Likewise, Hacker and Sova (1998) conducted a study where students enrolled in a lesson planning courseware module were assigned randomly to four treatment groups: traditional lecture and seminar, traditional lecture supplemented with a school-based courseware package, lecture and seminar through computer-mediated delivery, and the lecture and school-based courseware package through computer-mediated delivery. The authors found no significant difference in students' achievement gains whether the course was delivered in a traditional or online format.

However, despite results that indicate no differences between student achievement in online and traditional classes, and results that students have similar levels of satisfaction with both (Allen, Bourhis, Burrell, & Mabry, 2002), several authors and researchers have expressed concerns about the value and use of online learning (Grineski, 1999). For example, Garson (1999) commented, "There is great skepticism in academia that online education is pedagogically sound" (p. 15). In addition, some authors have commented that online courses that do not take full advantage of Internet technology simply continue the correspondence school model of education rather than improve upon it (Feenberg, 2001). Furthermore, online learning courses can be very demanding of faculty members' time (Garson, 1999), and some surveys of faculty members have shown that they prefer implementing classes in traditional over fully online formats (Peluchette & Rust, 2005). And although students, when surveyed, have stated that they were satisfied with the course, this could be a consequence, in part, of the high withdrawal rates that some researchers have found in studies of online courses (Pear & Crone-Todd, 1999). Furthermore, Garson argued that the best uses of technology may involve using it to supplement rather than supplant traditional instruction. Garson's view appears to be supported by the results of a meta-analysis conducted by Olson and Wisher (2002), who found an effect size of 0.48 when web-based instruction was blended with face-to-face instruction, compared to an effect size of 0.08 for fully online courses.

Uses of Technology in Statistics Instruction

The use of computers and technology for statistics instruction has been one of the most frequently researched topics (Becker, 1996), and a wide variety of technologies are now available for use in statistics instruction. In addition, students in statistics classes generally prefer non-traditional teaching methods over traditional classes (Johnson & Dasgupta, 2005). According to Rowell (2004), some of the technologies available for statistics instruction include

> Java applets for simulation and visualizing concepts in probability and statistics, free Internet-based analysis tools for calculating test statistics and p-values, multimedia textbooks which show video clips of real world situations that use

> statistics, and easy access to real data sets which can be downloaded and used in one of the many statistical application software packages. (p. 1)

Ben-Zvi (2000) also commented on the variety of technological tools available for inclusion in statistics instruction, including statistical analysis packages, “microworlds,” tutorials, online resources, and teachers’ “metatools” (p. 144). Statistical packages are generally used to conduct data analyses, but can also be used to create charts, graphs, and other visual representations. Microworlds are software programs designed to demonstrate statistical concepts and methods, including interactive experiments, exploratory visualizations, and simulations. Students can manipulate graphs, parameters, and methods (e.g., distributions, sample size, etc.). Tutorials are programs designed to teach or tutor students on specific content or to test their knowledge of this content. The goal of the program is to take over part of the duty of teachers and textbooks by providing information, setting tasks, evaluating student responses, and providing feedback. Resources include online course materials, online texts, JAVA demonstrations, electronic journals, electronic discussion lists, data, and general links. Finally, teachers’ metatools allow teachers to adapt software for their specific audience and educational goals. For example, these might allow a teacher to construct his or her own simulation, such as a spreadsheet program that allows users to create small programs called macros.

Technological tools used for statistics instruction can also be divided into those that play a supplementary role and those that play a substitutive role. Technologies that play a supplementary role are those that do not provide novel content, but simply assist students in learning material already presented by an instruction. Such technologies might include drill and practice, tutorials, simulations, statistical analysis software, enhanced lectures and classrooms, or some supplementary web-based uses of technology. Technologies that serve as substitutes for instruction are those that provide novel content not provided by the instructor. Examples of substitutive technologies include programmed or personalized systems of instruction and fully online courses.

Supplemental Uses of Technology in Statistics Instruction

Drill and practice. Drill and practice programs are those that administer test questions to students and elicit responses. Often, students cannot proceed to the next section until a correct answer is provided, or until they have answered incorrectly upon several attempts. Professors who make use of drill and practice programs usually allow students to use the program as often as they wish until they feel comfortable with the material. For example, Thelwall (2001) created a computerized practice test that randomly generated different test items where students were allowed unlimited practice for an introductory statistics course. After administering a survey to assess students' attitudes toward this practice assessment, the author found that the students believed that the program helped their understanding of statistical concepts and improved their confidence. In addition, students who under-performed on the practice test became motivated to work harder. However, those who performed satisfactorily were discouraged from further study. In an earlier study, Marcoulides (1990) compared students who were randomly assigned to three groups, an expert system program as an adjunct to lectures, computer-assisted drill and practice instruction as an adjunct to lectures, and a control group with lectures only, on a statistics achievement test. The author believed that systems should not only deliver content, but also provide problem-solving capabilities that emulate human decision-making capabilities as developed in the field of artificial intelligence. However, many computer-assisted instructional programs only provide drill and practice. For this reason, the author used an expert system, call ZEERA, that implemented the instructional strategies of self-evaluation, simulation, and tutorial. This program would ask students questions and then tailor its output and future questions to each student's response. The results of the study indicated that students in the expert system group and the drill and practice instructional group both significantly outperformed those in the control group on a statistics achievement test that measured students' ability to select appropriate statistical procedures to analyze particular data sets. However, the two computer-assisted instructional groups did not differ significantly from one another, indicating that students can benefit more from the inclusion of computer software than from simply receiving lecture alone.

Tutorials. Tutorial software programs generally are used as supplements to regular classroom instruction. They differ from programmed instructional software programs in that they generally cover small, individual topics rather than entire courses. Tutorial programs are typically those that provide students with additional practice in learning course material beyond that which is provided in the classroom, and may be included among the software available in computer laboratories that students may use at their discretion (Goolkasian, 1985). Frith et al. (2004) provided an example of a tutorial used in statistics instruction where they implemented spreadsheets as tutorials for the presentation of statistical concepts. The authors commented that "a typical tutorial will consist of several electronic Excel worksheets, some containing the interactive presentation of relevant mathematics content and other comprising examples and exercises" (p. 161). The authors stated that such a tutorial allows the student to work at his or her own pace, and provides immediate feedback. The authors compared three groups of students who had access to the spreadsheet tutorial at different times over a three-day period, and concluded that the computer tutorial provided students with an understanding of the concepts and helped them retain what they had learned, whereas the lecture sessions focused on teaching students how to calculate various statistics rather than the concepts themselves.

However, tutorials programs occasionally can be used as substitutes for classroom instruction for individual topics. For example, Aberson et al. (2003) created an interactive tutorial that covered hypothesis testing concepts, and compared a group that used this tutorial program, and were provided no lecture instruction on hypothesis testing, to a group that received a standard laboratory assignment including several *z*-test problems on a 10-item quiz covering the statement of the hypotheses, calculating *z* values, drawing statistical conclusions, and normal distribution probabilities. The researchers found that the tutorial group significantly outperformed the control group on the quiz. Furthermore, those students who received the tutorial found it easy to use, interesting, and the content presented in a clear manner. The authors concluded that the tutorial might have been effective because it provided immediate feedback to the students. In a second study using a similar tutorial covering the central limit theorem, Aberson, Berger, Healy, Kyle, and Romero (2000) found no difference between an experimental group that completed the tutorial and a control group that received a standard lecture and a demonstration on a short quiz. Although the groups did not differ significantly, both exhibited significant improvement from pretest to posttest. The authors concluded that students could learn just as effectively using a tutorial as they could from a standard lecture.

Others researchers have found similar results. For example, Morris (2001) conducted a study to evaluate the effectiveness of a tutorial program used by psychology students to review their understanding of the concept of correlation. The program used authentic data and was designed to address students' common misconceptions about correlation. The program allowed for direct manipulation and interaction, immediate feedback, multiple representations, and record keeping. To evaluate the program, the author used a pretest-posttest control group design, with an experimental group, a basic control group, and an instructional control group. The instructional control group completed traditional instructional materials that covered the correlation topic, whereas the basic control group received instruction in material not related to statistics. The three groups did not perform significantly different at pretest, but did differ significantly on the posttest. The results indicated that the computer-assisted group and instructional control group both exhibited significantly improved performance from pretest to posttest, and outperformed the basic control group. However, they did not significantly differ from each other, indicating that while students could learn the concept using computer-assisted instructional methods, they were no more effective than more traditional instructional methods. Similarly, Gonzalez and Birch (2000) compared groups assigned to a basic computerized tutorial, a tutorial program that included multimedia, a paper-and-pencil tutorial, and non-intervention control group, and found that students in the three intervention conditions did not score significantly different from each other on a test of statistical comprehension or on the performance on the tutorial modules. However, all intervention groups did significantly outperform the non-intervention group on the test of statistical comprehension, again suggesting that tutorial programs can be as effective as traditional methods, but may not be more effective.

Whereas the research generally has shown that statistics tutorials exhibit effectiveness similar to that of more traditional methods, researchers have provided some advice concerning how to create effective statistics tutorials. For example, Romero et al. (2000) suggested that an effective statistics tutorial should support students' self-regulation and awareness of their learning, elicit elaborative thinking, compel students to assess their own learning, force students to confront their misconceptions, and help students associate new material with familiar concepts. The authors proposed that information presented through the use of more than one media enhances learning. Furthermore, the authors commented that tutorials can be effective because "information that is presented simultaneously in two different media enhances learning and problem solving" (p. 248).

Simulations. Simulation programs are those that allow the student to manipulate data sets in order to examine the effects that various changes in parameters have on the nature of the underlying distribution. According to Mills (2003), the "application of learning statistics using computer simulation methods, grounded in the theory of constructivism, may benefit students by empowering them to develop their own understanding of statistics concepts" (p. 58). Mills further argued that the advantage of using computer simulation methods to teach statistics was that they allow students to develop their own understanding of statistical concepts and they can assist students in overcoming any misconceptions they have about these concepts. Other researchers (Marasinghe et al., 1996) stressed the importance of active, rather than passive, learning that is promoted by the use of simulations. Marasinghe et al. argued that using computers for simple data analysis has limited instructional effectiveness as students might not grasp the statistical concepts themselves. However, through the use of simulations and dynamic graphics, students can manipulate the data in such a way that they can see the effect of any changes that are made. The authors further stated, "Using a judicious combination of traditional data analysis exercises and new exercises based on carefully designed and implemented simulation-based experiences has the potential of improving the students' level of understanding of statistical methods" (p. 351). Moore (1997) concurred by commenting that modern courses in statistics should use technology to encourage more active participation by focusing more strongly on data analysis, data production, inference, data exploration, using diagnostic tools, and so forth, rather than the more straightforward nature of traditional statistical calculations.

Several examples exist where simulations were used to enhance statistics instruction. For example, delMas, Garfield, and Chance (1999) used a sampling distributions simulation program, which allowed students to specify and change the shape of a population, choose different sample sizes, and simulate sampling distributions by randomly choosing large numbers of samples. Additional examples of simulations include those that illustrate the central limit theorem, where students can observe the changes in the standard deviations that come with changes in sample size, as well as observe that as the sample size increases, the distribution of the sample means appears more and more like a normal distribution (Mills, 2002). Other computer simulation programs cited by Mills include those that illustrated the *t*-distribution, confidence intervals, regression analysis, sampling distributions, hypothesis testing, and survey sampling.

An example of an early study using simulations was conducted by Stockburger (1982), who compared the performance of students in an introductory statistics class required to complete three simulation activities covering means, normal curve, and correlation coefficient estimation to students not required to do so on tests designed to measure their ability to quickly estimate statistics. The results indicated that students in the treatment group attempted significantly more estimations and obtained more correct estimations on both a means estimation test and on a normal curve estimation test. In addition, survey data indicated that most students believed that the computer exercises helped in their understanding of course material and were more comfortable using the computer after having participated in the exercises. A second study of simulations was conducted by Sterling and Gray (1991), who created several statistics simulation programs that allowed users to manipulate several small data sets, such as altering the shape of sampling distributions and examining the effects of the presence of outliers. The results indicated that the experimental group who used these simulations outperformed the control group on a statistics test. However, whereas the experimental group believed that the use of computers did improve their performance, they resented having to spend extra time on the computers, as they became aware that other sections did not have the computer-related assignments that they were required to complete. As a result, the control group actually provided a significantly higher rating concerning their attitude toward computer usage as an aid to instruction.

More recently, researchers have been increasingly making use of Java applets in order to create simulations intended for use in statistics instruction (Bertie & Farrington, 2003). For example, Aberson, Berger, Healy, and Romero (2002) used the WISE power applet tutorial, which allowed the student to manipulate population means, standard deviations, sample sizes, Type I and Type II error, power, and effect size, and to simulate drawing samples, as an extra credit assignment for those students who wished to complete it. Eighteen students chose to complete the assignment whereas seven did not. Those who chose to use the tutorial were found significantly to outperform those who did not on a final examination question related to statistical power. Those who used the tutorial also generally found it easy to use, and felt more comfortable with the topic of statistical power after use. However, because the sample was self-selecting, the possibility exists that those who chose to use the tutorial were already the higher performing or more motivated students. In addition to Java applets, another common recent use of simulation methods include Monte Carlo simulations using the MC2G program, which can be used to demonstrate several statistical concepts including robustness analysis, power analysis, and sample size analysis for one or two groups (Brooks, 2003).

Statistical analysis software. Another common use of technology in statistics instruction involves the incorporation of statistical analysis software programs. Generally, students are either given access to statistical analysis software in computer laboratories (Goolkasian, 1985; Raymondo, 1996), or the use of such software may be incorporated into regular class time. In some cases, students are offered classes specifically designed to instruct them in the use of statistical analysis software programs (Tromater, 1985). Often, use of such programs is combined with the inclusion of large, real-world data sets so that students can observe the value of various statistical procedures (Stork, 2003). According to some researchers, such as R. L. Rogers (1987) and Tromater, statistical analysis software programs can permit students to spend less time doing calculations and allow them more time for conceptual understanding. Moore (1997) agreed by stating that statistics instructors should spend less instructional time on topics that can be automated, such as calculations, and more time on the interpretation of graphs, strategies for effective data exploration, basic diagnostics, and the conceptual meaning of various statistical terms.

In addition, Rogers believed that by using computers, students could develop some level of computer literacy. Furthermore, proponents such as Ben-Zvi (2000) argued that the use of statistical analysis software can assist in exploratory data analysis, which includes a heavy reliance on visual displays as analytical tools. In some fields, such as business, professional organizations have made recommendations for improving statistics instruction, including reducing lecture time and making greater use of statistical analysis software for analyzing real data sets (Strasser & Ozgur, 1995). In addition, Lesser (1998) discussed new content standards developed for his courses, which involved instructing students to be able to evaluate statistics critically in the media and in their major field of interest, and to plan, implement, and communicate the results of a real-world research project. The standards, Lesser stated, were similar to those of the National Council of Teachers of Mathematics. The author believed that the status quo for teaching statistics was too often either rule-bound, structured courses for calculating statistics, or overly mathematical introductions to statistical probability. In order to counter this, the author incorporated a greater use of statistical analysis software and realistic data sets, as well as a course web site complete with class announcements, course policies and rubrics, data sets and exercises, and discussion forums.

After altering his courses, Lesser (1998) then compared students in more traditional classes to students in the computer-enhanced classes. He found that those in the computer-enhanced classes were significantly less likely to believe that statistics published in the media are always accurate and unbiased, and more likely to agree that statistics classes should provide opportunities to work in teams. In addition, he calculated effect sizes for various satisfaction-related survey items, and found that students significantly preferred various aspects of the treatment courses over the traditional courses, with effect sizes ranging from 0.25 to 0.64, on questions covering group work, lab and project approach, criterion-referenced over norm-referenced assessment, and the use of real-life examples. In addition, 93% of the students in the treatment section said that they would not have enjoyed the traditional approach more. After assessing student performance, Lesser found that the students in the treatment condition significantly outscored the tradition sections on a critical thinking problem, but did not differ significantly on two traditional problems (i.e., a probability problem, and a two-population hypothesis testing problem). This indicated that the students in the treatment section exhibited greater critical thinking, but did not suffer on the more traditional statistical problems as a result of having been assigned to a treatment section.

In addition to Lesser, Basturk (2005) also conducted a study of the effectiveness of statistical analysis software programs. The author conducted a quasi-experimental study, comparing a lecture-only introductory statistics class to a computer-assisted class. Students in the computer-assisted class used SPSS to practice data analysis on real data sets. The author believed that the use of SPSS may help students develop their own understanding of statistics concepts. The author found statistically significant differences in favor of the computer-assisted learning class over the lecture-only class on both the midterm and final examinations. In addition, students in the computer-assisted learning class improved their examination scores from the midterm to final, whereas students in the lecture-only class declined in performance. In other words, the difference between the two groups increased as the subject matter moved from descriptive to inferential statistics. Basturk believed that traditional methods of statistics instruction do not establish a clear link between the statistics themselves and their use in the real world.

Although several authors examined the usefulness of statistical analysis software programs such as SPSS or SAS, other researchers promoted the use of spreadsheet programs for use in statistics instruction (Warner & Meehan, 2001). Warner and Meehan argued in support of the use of a spreadsheet program, such as Microsoft® Excel, for the instruction of statistics rather than a software program designed specifically for statistical analysis because they felt that spreadsheet programs were more popular, spreadsheet skills were in more demand among employers, they are easier to use, they are often readily available to students, they perform most basic statistical analyses, custom graphs and charts can easily be created, and in the case of Excel, an Analysis ToolPak add-on is available. After students in the authors' classes completed five required assignments using Excel, Warner and Meehan administered a survey to obtain their opinions about the homework assignments and course manual. Overall, the students felt that the assignments were useful, did not require too much work, and required some creativity and critical thinking. However, they rated the assignments higher in improving their computer skills than their knowledge of statistical concepts.

While researchers such as Lesser (1989) and Basturk (2005) found differences favoring groups of students instructed with the assistance of statistical analysis software over those who were not, other researchers have found little difference between groups. For example, Raymondo and Garrett (1998) implemented a quasi-experimental design to determine if students who participated in four computer laboratory experiences that consisted of using SPSS to conduct various statistical analyses outperformed students who did not on a test of comprehension of statistical concepts, and found no significant difference between the two groups. In addition, Ware and Chastain (1989) conducted a study to compare college students' skills and attitudes associated with the use of computers or the use of calculators and paper/pencils to complete calculations. Both groups were enrolled in an introductory statistics course, with the experimental group using a mainframe to conduct statistical analyses and the control group completing statistical analyses using calculators and/or paper and pencil. A third group was assembled consisting of volunteers who previously had not taken a statistics course. The three groups did not significantly differ on pretest scores for selection of appropriate statistical procedures, interpretation of statistical results, attitudes toward statistics, and attitudes toward computers on the pretest. The results indicated that, although the computer-assisted group and the traditional instruction group performed significantly higher on the selection and interpretation scales than students who had no statistics background, they did not significantly differ from each other. However, the computer-assisted group scored significantly higher on the attitudes toward statistics scale than the traditional instruction and no statistics groups. Finally, Christmann and Badgett (1999) conducted a meta-analysis of computer-based software packages and found, from 10 effect sizes, a small average effect size of 0.043 for studies that examined the effect of statistical analysis software packages on student performance. Although the results of such studies do not favor the use of statistical analyses programs, they do indicate that computer-assisted instructional approaches can result in similar performance to traditional approaches.

However, as some authors such as R. L. Rogers (1987) and Tromater (1985) believed that the use of statistical analysis software programs would allow students to spend less time conducting calculations and allow more time for conceptual understanding, other authors have cautioned against the use or overuse of statistical analysis software packages in statistics classes. For example, Mills (2002) argued that most professors simply used statistics software to perform routine data analysis, which does not allow students to obtain an understanding of abstract statistical concepts. The author suggested implementing computer software programs designed for simulation purposes instead. In addition, Searle (1989) argued that the use of statistical computing programs can result in students calculating statistics they should not be calculating or do not need rather than using them to calculate the appropriate statistics. The author believed that students may be tempted to choose more sophisticated statistical analyses, simply because they are easy to perform, rather than choosing the appropriate statistical analysis for the given situation. Furthermore, Searle suggested that inexperienced or novice researchers get assistance from experience statisticians when interpreting output from a statistical computing package. Finally, the author argued that “that easy-to-use computing packages can be a substitute for (rather than supplement to) a proper knowledge of statistical methodology” (p. 190). Other researchers, however, have disagreed. For example, Dallal (1990) responded to Searle’s suggestions by commenting that statistical analysis software packages can help students to understand the underlying meaning of statistical methods, not just how to conduct them. He indicated that many students do not really understand some statistical analyses until they use them on a particular set of data, with the assistance of statistical analysis software packages.

Technology-enhanced lectures and classrooms. A number of instructors of statistical methods classes have made use of technology and multimedia to enhance their lectures or classroom environments. Velleman and Moore (1996) defined multimedia as a system that combines text, sound, images, video, animation, and computer graphics. They suggested that instructors use multimedia that combines the full range of appropriate technology with sound principles of instruction, and indicated that multimedia can encourage students to become more active participants in their learning and students learn best by their own activity, not by passively receiving information. They further stated that "another goal of a first statistics course, often unstated but nonetheless important, is to motivate students to change their attitude about statistics" (p. 219). A technology-enhanced lecture or classroom environment can help improve students' interest and assist in achieving this goal. For example, Erwin and Rieppi (1999) felt that multimedia classrooms increase students' motivation, increase students' participation, enhance the presentation of information, and facilitate student feedback. They compared the effectiveness of multimedia and traditional instructional approaches on students' academic performance in several psychology courses, including statistics. The results indicated that students in the multimedia classroom obtained higher scores on the final examination in all courses than those in the traditional condition. The authors also placed more students in the multimedia classrooms, so the results suggested that a larger number of students can learn more effectively in a multimedia classroom environment than a smaller number of students in a traditional classroom environment. The authors concluded that the active learning approach in the multimedia environment enhanced students' motivation, interest, and enjoyment. Additionally, some instructors of statistical methods courses have begun to incorporate the use of student response systems into their courses in order to improve attendance and interaction, and reduce passivity (Wit, 2003). The system described by Wit encourages students who attend class to express their knowledge or beliefs without the fear of ridicule. In addition, the student response system provided useful feedback to the instructor so that he could modify his lessons as needed. Overall, the author concluded that the use of the student response system in class breaks up the monotony of lectures, makes lectures more interactive, involves the whole class, allows students to contribute without the fear that they might make a mistake, and provides the instructor with feedback to observe how well students are doing in class. Furthermore, Hyden (2005) incorporated the use of laptops in his statistics class, which assisted in the development of classroom presentations that allowed the instructor more graphical options than using a chalkboard and allowed him to demonstrate a variety of statistical concepts that students often have difficult with. Finally, Liao (1999) conducted a meta-analysis to compare classes that incorporated the use of hypermedia (i.e., interactive videodiscs, computer simulators, or interactive multimedia) to classes that used other forms of computer-assisted instruction and traditional instruction, and found an effect size of 0.41 favoring the use of hypermedia.

Web-based uses of technology. A variety of instructors of statistical methods have begun to incorporate Internet technology into their lectures and curriculum. A common use of the Internet for instruction is to create course web pages complete with class announcements, course policies and rubrics, data sets and exercises, discussion forums, and links to additional resources, such as journal articles, real data sets, and interactive tutorials (Aberson, Berger, Healy, & Romero, 2001; Ben-Zvi, 2000; Hyden, 2005; Lesser, 1998). According to Aberson et al., traditional approaches to teaching statistics focus on computations and not on the relevance of statistics as a tool for summarizing data. Supplemental online resources provide students with greater flexibility and increase their interest in the material. In addition, Hunt and Tyrrell (2000) discussed the advantages of using the Internet for instruction, and stated that the web is accessible, adaptable, variable, resource-full, and interactive. Students can take charge of their learning and teachers can use a variety of instructional media. Students can also obtain real data sets from the Internet that they can use for data analysis practice. Furthermore, the Internet also contains a variety of Java applets that can be downloaded and used to teach and reinforce various statistical concepts. Finally, the Internet can also be used to communicate and collect data, which can then be used as an example upon which statistical analyses can be demonstrated (Sanders, 1996).

Varnhagen et al. (1997) provided an example of one use of the Internet as a supplement to traditional lecture for statistics instruction. The authors created an Internet-based statistical laboratory to help students learn how to use statistical analysis software, learn to interpret output, to practice writing in American Psychological Association (APA) format, and to familiarize students with the Internet as a research and dissemination tool. On their site they included a syllabus and course information, project descriptions and data sets, online help for describing data, links to other sites, email addresses for the instructor and graduate assistants, a discussion forum, and an electronic assignment submission form. When surveyed, Varnhagen et al. reported that the "students perceived the computer system as extremely good, stimulating, productive, moderately friendly, fun, and moderately timesaving" (p. 277). The authors also noted that the quality of students' discussions appeared to improve throughout the semester.

Another use of the Internet as a supplement to instruction involves using it to enhance communication between students and the instructor. For example, Hyden (2005) incorporated the use of laptops in his statistics class in order to create a community of learning among the students and himself, as well as allow students to drive the direction of the course within the boundaries of the syllabus. He included discussion boards, which allowed students to submit and respond to questions and helped students realize that others experience similar struggles and difficulties, as well as daily online quizzes, which allowed the instructor to use class time more productively by obtaining feedback concerning which topics students were struggling with prior to the next class. Similarly, Benedict and Anderton (2004) applied an instructional technique they referred to as "Just-in-Time Teaching" (Novak, Patterson, Gavrin, & Christian, 1999, as cited in Benedict & Anderton) to their courses that allowed students to inform the instructor of their progress with the course material just in time for class. The technique involved posting several questions on the Internet to gauge student knowledge, allowing the instructor to adjust how class time is used. The authors compared a group that received this approach in a psychological statistics class to a control group that did not. The control group received a five-item multiple-choice quiz weekly in class prior to the discussion of material. The experimental group responded to a weekly set of questions posted on an Internet site prior to class. The results indicated that the experimental group significantly outperformed the control group on final exam scores. In addition, the experimental group felt positive overall about this approach in terms of facilitating their problem solving skills, their understanding of statistical concepts, and their overall learning of statistics, and they felt that this approach should be used in future classes.

Substitutive Uses of Technology in Statistics Instruction

Programmed or personalized systems of instruction. Not only have personalized systems of instruction (PSI) been used widely throughout several subject areas in higher education, some researchers have implemented the methodology for statistics instruction. For example, Wagner and Motazed (1972) created a modified, computerized version of Keller's (1968) Personalized System of Instruction (PSI) methodology and used it to administer an applied statistics course in an early study of the effect of PSI on statistics instruction. The characteristics of PSI include (a) individual pacing, (b) mastery before advancement, (c) focus on the written word rather than lectures, (d) use of lectures and demonstrations as sources of motivation rather than instruction, and (e) the use of proctors and tutors. The authors found that students believed that the PSI teaching method was superior to the traditional lecture method, and the professor who taught the course felt that students learned more, were more motivated and enthusiastic, and were more satisfied in the PSI course. In addition, more students received "A" grades in the class than in years before PSI was implemented, but this is largely due to the nature of PSI, which requires mastery on one concept before proceeding to the next. All students who completed the course thereby earned "A" grades. In addition, Malec (1982) implemented PSI methodology in a statistics course in 1973 in order to accommodate different students' rates of learning the material. The author found that the advantages of the PSI course for students were that they could work at a comfortable pace, they had a tutor available, they received prompt feedback, and they exhibited increased mastery of the course content. The author also commented that the students generally liked the system.

In a later study, Tsai and Pohl (1980) conducted three experiments designed to examine the effectiveness of computer-assisted instructional version of a personalized system of instruction program. The first experiment compared a computer-assisted instructional class to a traditional lecture and discussion class on a statistics achievement test as well as a shorter retention test. The authors found no significant differences between the treatment and control groups on either the statistics achievement test or the retention test. The second experiment compared computer-assisted classes to traditional classes, but also combined teacher/student discussion sessions to augment the computer-assisted instructional classes. The authors found that the class taught using computer-assisted instruction outperformed the traditional class on the achievement test, but not the retention test. The third experiment further examined the effects of the teacher/student discussions by comparing five groups: a traditional class, a class taught using a printed version of the PSI materials, a class taught using a printed version of the PSI materials augmented with teacher/student discussions, a computer-assisted instructional group, and a computer-assisted instructional group augmented with teacher/student discussions. Overall, the computer-assisted instructional group augmented with teacher/student discussions outperformed all other groups. However, the PSI group augmented with teacher/student discussions outperformed the group only provided with computer-assisted instruction. Overall, the results provided some support for the use of computer-assisted instruction, but provided greater support for the use of teacher/student discussions to augment whatever personalized system of instruction is implemented, whether on paper or on computer.

Like Tsai and Pohl (1980), Mausner et al. (1983) created an interactive computer-assisted instructional course using PSI methodology. According to the authors, their program taught students statistics from a problem-solving orientation, and was more interactive than an electronic workbook. The authors evaluated the CAI program using 42 students enrolled in a statistics course, and administered several performance and attitudinal measures. The results indicated that the students reacted positively to the CAI method, and felt that it should be used in other classes. Mausner et al. believed that the good students did well with the CAI method, but the marginal students had difficulty pacing themselves and finishing the program by the end of the year.

Likewise, Varnhagen and Zumbo (1990) used the PLATO86 software program to instruct psychology students in introductory statistics, which made use of text, computer-guided problem solving, simulation, graphics, and branching routines, which could alert and correct student errors. The authors also used the older PLATO74 software program, which was much less interactive, and resembled a computerized textbook. However, both programs were self-paced. They compared both versions of the program to a control group in order to determine whether the computer or the method of instruction was the primary factor affecting student learning. The authors collected homework assignments as outcome measures, as well as affective measures of students' statistics anxiety and attitudes toward the computer laboratory. The results indicated that students who used the

PLATO86 program felt more prepared for the midterm examination, than those in the other groups. They also felt that the lab was more worthwhile, more interesting, and more enjoyable. However, no significant differences were found between groups on their homework assignment scores.

More recent studies of PSI have used a modified version that is a multimedia software package add-on to SPSS, called ActivStats (Harrington, 1999; Morris et al., 2002). ActivStats teaches introductory statistics concepts and methods, and includes activities that contain narration and animation, video, simulation, interactive experiments, and self-test quizzes (Mills, 2004). The software package was designed to be a complete introductory statistics course. Harrington (1999) incorporated ActivStats into a distance learning course. The author compared 33 students who took the programmed instruction/distance-learning course to 61 students who took the traditional version of the course. All classes were taught by the same instructor, and included the same content and assignments. Both classes had access to a data analysis package, but were not required to use it. Students in the distance learning class could work on the course material at their individual paces and could obtain additional assistance from the instructor if necessary. The results indicated that students in the traditional classroom significantly outperformed those who enrolled in the distance-learning course on final course grades after controlling for age, sex, race, and GPA. GPA by class type interaction was also significant, with students with high GPAs in the distance learning class outperforming those with low GPAs. GPA was not related to performance in the traditional class. The author also surveyed the students in the distance-learning course and found that they rated 8 of 15 aspects of the course favorably. Specifically students reported that they enjoyed using ActivStats, found it easy to use, found the homework assignments helpful, thought the explanations were clear, felt that the instructor provided assistance as needed, felt that feedback on homework assignments was helpful, felt the course was structured appropriately, and enjoyed the flexibility of the distance learning format. However, because this was a quasi-experimental study and students could choose which section in which to enroll, personal characteristics that relate to their decision to take a particular section may be related to academic performance.

In a second study using ActivStats, Morris et al. (2002) wanted to determine if computer-based activities that provide multiple representations and direct manipulation of data contributed to students' understanding of correlations and measures of central tendency, and to determine if computer-based activities that involve estimation of statistics from graphical displays contribute to students' understanding of correlations and measures of central tendency. The study included 50 psychology students placed in a pretest/posttest control group design. All of the students had completed courses in introductory statistics. The students were divided into three groups: the ActivStats group, the computer-based learning group, and a control group. The ActivStats group was able to directly manipulate the data, whereas the computer-assisted group simply estimated statistics from graphical displays. Members of the control group completed computer-based exercises related to a different topic, namely probability. The results indicated no significant differences between the ActivStats group, the computer-based learning group, and the control group on tests of correlations and measures of central tendency. However, the ActivStats group did show significant improvement from pretest to posttest on the test of measure of central tendency, whereas the other groups did not.

Online instruction. Like other fields and topic areas, instructors have begun to explore the application of online forms of instruction to statistics. For example, Harrington (1999) compared two courses, one taught traditionally and one taught as programmed instruction online. Both courses were taught by the same instructor, and included the same content and assignments. Both classes had access to a data analysis package, but were not required to use it. The distance learning class used a CD-ROM package called ActivStats, which has a structure similar to a textbook, and includes video clips, interactive visualization tools, simulations, exercises, and self-test quizzes. Students in the distance learning class could work on the course material at their own pace and could obtain additional assistance from the instructor if necessary. Harrington found, however, that students in the traditional classroom significantly outperformed those enrolled in the online course on final grades after controlling for age, sex, race, and grade point average. Although students in the online course did not perform as well as those in the traditional course, they did enjoy several aspects of the online course including using ActivStats, receiving assistance and feedback from the instructor, and experiencing the structure and flexibility of the course. In terms of student achievement, other authors have found similar results to those of Harrington when examining the effectiveness of fully online forms of statistics instruction. For example, Wang and Newlin (2000) compared three web-based sections of a statistical methods in psychology course to three traditional sections taught by the same instructor. The web-based was delivered as an online “cyberclass,” with mandatory, synchronous biweekly lectures in online chatrooms. Both class types had the same course content and assignments. The results indicated that the students who were enrolled in the online version of the course performed significantly worse on the final examination and in overall course grades than those who took the traditional version. However, one group of researchers did find no significant difference between the achievement of students in an online statistics course and those in a traditional course, but did find that students in the online course were less satisfied (Summers et al., 2005).

Meta-Analyses of the Effectiveness of Technology Use in Higher Education

Early Meta-Analyses

James Kulik and his colleagues have conducted some of the earliest meta-analyses of the effectiveness of technology use in education. For example, C. C. Kulik et al. (1980) conducted a meta-analysis that examined the effects of five types of instructional technologies: personalized system of instruction (PSI), computer-based instruction, the audio-tutorial approach, programmed instruction, and visual-based instruction. The authors found an average effect size of 0.28 across 278 studies. The type of technology had the strongest effects on study outcomes, with PSI having the largest effect size (0.55) among those examined. In addition, the authors found an effect size of 0.10 in favor of instructional technology in students’ rating of course quality. The effect size also differed by instructional type, with PSI (0.44) exhibiting a larger effect size than other technologies (0.01). Overall, the results indicated that instructional technology had a small, but significant effect on student achievement and student ratings of effectiveness.

However, this meta-analysis included all levels of education, from elementary to postsecondary.

In a second meta-analysis by J. A. Kulik, C. C. Kulik, and Cohen (1980), the authors specifically examined the effectiveness of computer-based teaching at the college level. The authors obtained 59 studies that they divided into four major types of computer applications for instruction: tutorials, computer-managed teaching, simulation, and computer programming for problem solving. Tutoring instruction included those that provided instruction directly to students. In computer-managed instruction, the computer evaluated student performance, diagnosed weaknesses, and guided students to appropriate instructional resources. Simulation studies involved applications where students explored variables in models simulating aspects of social or physical reality. Programming studies were those where the computer was programmed by students to solve problems in the academic field in which in which they were studying. The results indicated that, among 54 studies that included examination performance as outcome criteria, the overall effect size was 0.25. J. A. Kulik, C. C. Kulik, and Cohen found no significant differences in effect size based on the type of computer use. In addition, the authors found an effect size of 0.24 when student attitudes were used as an outcome measure in 11 of the 59 studies. Finally, the authors examined the relationship between instruction time and computer use and found that computer use can significantly reduce instructional time, from 3.5 hours on average to 2.24 hours in eight studies.

In a later meta-analysis of computer-based education in colleges, Kulik and Kulik (1986) found a mean effect size across 99 studies of 0.26. The authors compared three types of computer-based education: CAI (tutorial and drill and practice), computer-managed instruction, and computer-enriched instruction (simulation, programming, and problem-solving). CAI exhibited an effect size of 0.08, computer-managed instruction 0.11, and computer-enriched instruction 0.05. Furthermore, the authors found a small mean effect size of 0.18 across 6 studies that included retention tests as an outcome measure. Furthermore, the authors found that studies obtained from professional journals had significantly higher effect sizes (0.42) than dissertation studies (0.11), and the mean effect size was higher for the soft sciences (0.35) than the hard sciences (0.15). The authors also examined studies that used student attitudes as outcome measures, including attitudes toward computers (0.27), attitudes toward instruction (0.31), and attitudes toward subject (-0.03). Finally, among 15 studies that included instructional time as an outcome variable, the authors found that classes that used computer-based education resulted in a one-third reduction in instructional time. Kulik et al. (1986) found similar results when examining computer-based instruction in adult education. The authors found a significant reduction in instructional time when computer-based instruction was compared to traditional instruction, and found a mean effect size of 0.42 across 24 studies.

In a later study, Kulik and Kulik (1987) again examined the effectiveness of computer-based instruction across all levels of education, including elementary, middle school, high school, college, and adult education. Across 199 studies, the authors found a mean effect size for achievement of 0.31 standard deviations, but found no significant differences between different types of computer-based instruction. In addition, the authors found that computer-based education reduced instruction time by an average of 32%. Furthermore, the authors found that computer-based instruction resulted in a 0.28 effect size for attitudes toward instruction and 0.33 for attitudes toward computers when compared to traditional instruction. Computer-based instruction did not, however, have an effect on attitudes toward subject matter. The authors also found larger effect sizes when the studies were found in published sources rather than unpublished sources, when different teachers taught the experimental and control groups compared to the same teacher instructing both, in more recent studies compared to older ones, and in shorter duration studies compared to longer ones. Kulik and Kulik updated this meta-analysis in 1989, including 254 studies across all levels of education. They found a mean effect size of 0.30 across all studies, and found that computer-based instruction resulted in a 30% reduction in instructional time. Again, Kulik and Kulik found no significant differences in effect sizes based on the type of technology used. Furthermore, the effect size for attitude toward instruction was found to be 0.28, and attitude toward computers was 0.34. Again, however, the effect size for attitude toward subject matter was found to be near zero. The authors also obtained several significant results related to study features. For example, the effect sizes were larger for published studies (0.44) than unpublished studies (0.24), larger when different teachers taught the control and experimental groups (0.42) than when the same teacher taught both (0.25), and larger for shorter interventions (0.42) than for longer ones (0.26), which the authors believed could be contributed to a novelty effect.

Recent Meta-Analyses

Recent meta-analyses of the technology use in education have uncovered similar, modest effect sizes to those found in the early studies. For example, Khalili and Shashaani (1994) conducted a meta-analysis of the effectiveness of computer applications on students' academic achievement using studies that included students from elementary school to college. The average effect size from 151 comparisons obtained from 36 studies was 0.38. Unlike several earlier studies, however, Khalili and Shashaani found significant differences in mean effect sizes by type of technology used. The authors found higher effect sizes in studies that used simulations (0.79) than those that used drill and practice (0.11) or problem solving software (0.41). In addition, they found a higher mean effect size when different teachers taught the experimental and control groups (0.45) than when the same teacher was used (0.35), when the treatment lasted one to two months (0.94) than when treatment lasted less than one month (0.14), and when the participants were high school students (0.62) than when the participants were middle school students (0.11).

Other researchers have conducted meta-analyses examining the effectiveness of specific types of technology uses. For example, Liao (1998) located 35 studies comparing the effects of instruction enhanced with hypermedia (i.e., interactive videodiscs, computer simulators, or interactive multimedia) to traditional instruction on students' achievement. The overall study-weighted mean effect size was found to be 0.48, indicating that instruction including hypermedia has a moderate positive effect on student learning over traditional instruction. Three variables had significant impacts on the effect size: instructor bias, research design, and type of delivery system. The results indicated that studies where the same instructor taught both the control and experimental conditions had a significantly larger mean effect size than those with different instructors, indicating that some instructor bias might be occurring. In addition, studies employing one-group, repeated measures designs exhibited significantly higher effect sizes than studies using other designs. Finally, studies were simulators were used exhibited significantly higher effect sizes than those that used interactive videodiscs or interactive multimedia.

Liao conducted an updated meta-analysis of the effectiveness of hypermedia in 1999, comparing classes that used hypermedia to classes that used traditional instruction, computer-assisted instruction, or videotapes in a meta-analysis. Overall, the author found 46 studies across all grade levels. The overall mean, study-weighted effect size across all studies, including 143 effect sizes, was found to be 0.41. In addition, the author found that one-group repeated measures studies resulted in significantly higher effect sizes than those that included control groups. Furthermore, larger effect sizes were found in studies that used simulators compared to those that used interactive multimedia.

Additional researchers have conducted meta-analyses of specific types of instructional technology use in recent years. For example, Lee (1999) conducted a meta-analysis to determine the effect of the classroom use of computer-based instructional simulations on attitudes and achievement. The author obtained 51 effect sizes from 19 studies, and found an overall mean effect size of 0.41 for achievement. The overall mean effect size for attitudes was -0.04, from 5 effect sizes. In addition, Olson and Wisher (2002) conducted a meta-analysis of the effectiveness of web-based instruction on student achievement. The authors obtained 15 effect sizes, and found a mean effect size of 0.24 for web-based instruction. The mean effect size was higher for those classes that blended web-based instruction with face-to-face instruction (0.48) than for fully online courses (0.08). The authors concluded that web-based instruction seems to be an improvement over conventional classroom instruction, but it may not compare favorably to computer-based instruction. Finally, Allen et al. (2002) compared student satisfaction with distance learning to traditional classrooms in higher education. The authors found, across 25 studies, that students exhibited a slight preference for traditional education over distance education (average $r = 0.031$), but exhibited little difference in satisfaction levels.

In addition to the study of specific technologies, other researchers have conducted meta-analyses of the effectiveness of the use of technology on achievement in specific fields. For example, Bayraktar (2002) conducted a meta-analysis of effectiveness of computer-assisted instruction in science education. The author obtained 42 studies with 108 effect sizes, ranging from -0.69 to 1.295. Overall, Bayraktar found a mean effect size across all studies of 0.273, and found significantly different effect sizes based on the mode of instruction. A mean effect size of -0.107 was found for drill and practice, while mean effects sizes of 0.391 and 0.369 were found for simulations and tutorials, respectively. However, the author noted that previous meta-analyses that involved subjects other than science found positive effect sizes for drill and practice (e.g., Niemiec and Walberg, 1985). Furthermore, an effect size of 0.288 was found when the computer was used in a supplementary role and 0.178 when used as a substitute. Also, Christmann and Badgett (1999) examined the effectiveness of technology use for instruction in a specific topic, namely statistics. The authors found a mean effect size across 14 effect sizes obtained from 9 studies of 0.256. The authors further divided the studies into software type used (i.e., computer-assisted instruction, problem-solving, and statistical software packages) and found the effect sizes to be 0.929, 0.651, and 0.043, respectively. However, 10 of the 14 effect sizes were from studies using statistical software packages, whereas the other two categories had 2 effect sizes each.

Finally, Timmerman and Kruepke (2006) conducted one of the most recent meta-analyses of the use of technology for instruction at the higher education level. The authors first identified five study features that they believed might moderate the effectiveness of CAI: course subject, comparison group type, publication year, student academic standing, and duration. They also categorized the type of media used into five channels: text, text with graphics, audio, video, and apparatus. Furthermore, interventions were categorized as those that provided feedback and those that did not, and interventions were categorized as those created by a course instructor or those created by a general publisher. The authors found 118 studies that met their criteria for inclusion, and found an average effect size (Pearson's *r*) of 0.12. This corresponds to a Cohen's d effect size of 0.24, similar to the effect sizes found in previous studies. When comparing course subjects, the authors found that all categories differed from one another, with social sciences exhibiting the largest mean effect size (0.18), followed by the physical sciences (0.16) and the life sciences (0.09), whereas language/humanities exhibited the smallest effect size (0.03). The authors also found that newer studies had significantly smaller effect sizes than older studies. In addition, differences were found by media channel, with audio having a larger effect size (0.26) than text (0.14), followed by text with graphics (0.12), video (0.07), and physical apparatus (-0.05). Furthermore, no differences were found between interventions that provided feedback and those that did not. However, the authors did find a significant difference in effect sizes based on whether the CAI was created by the course instructor (0.14) or by a general publisher (0.09). Overall, across early and more recent meta-analyses of technology use in education, researchers have obtained small to moderate effect sizes for student achievement, and small or near-zero effect sizes for student attitudes.

Conclusion

The integration of technology into higher education instruction can be traced as far back as the use of teaching machines in the 1950s (Reiser, 2001), and progressed through the programmed instruction and personalized system of instruction movements of the 1960s and 1970s. The computer-assisted instruction movement developed from the combination of techniques associated with the programmed instruction movement and the development of computer technology (Kulik & Kulik, 1986). As a result, a variety of higher education institutions had instituted computer-assisted instructional programs, such as PLATO, throughout the 1970s and 1980s (Garson, 1999).

Currently, a variety of modes of technology are being used to enhance instruction in higher education. For example, supplemental uses of technology include drill and practice, tutorial, and simulation software, computer laboratories, enhanced lectures, and some forms of web-based instruction. Substitutive forms of technology use include programmed and personalized systems of instruction and fully online courses. Statistics courses were some of the first to incorporate technology (R. L. Rogers, 1987), and often make use of statistical analysis software programs in addition to the other forms of technology.

Overall, researchers have found small to modest effect sizes for achievement in favor of the use of technology to enhance instruction in higher education. On the lower end, J. A. Kulik, C. C. Kulik, and Cohen (1980) found an effect size of 0.25 favoring the use of technology to enhance instruction in higher education. Authors of more recent studies, however, have found effect sizes as high as 0.48 (Liao, 1998) and 0.41 (Lee, 1999; Liao, 1999). In addition, technology has been shown to generally have small to near zero effects on student attitudes (Kulik & Kulik, 1986). Furthermore, researchers have found that the size of the effect tends to vary based on the type of technology used (Bayraktar, 2002; Khalili & Shashaani, 1994; C. C. Kulik et al., 1980). Finally, Christmann and Badgett (1999) and Hsu (2003) examined the effectiveness of using technology specifically in statistics classes, and found average effect sizes of 0.256 from 14 effect sizes and 0.43 from 31 effect sizes, respectively.

CHAPTER III

METHODOLOGY

The method used for this study was similar to the meta-analytic techniques suggested by Cooper and Hedges (1994) and Wolf (1986). Meta-analysis is defined as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass, 1976, p. 3). The use of meta-analytic techniques allows for the comparison of the effectiveness of technology use in statistics education on achievement and attitudes across a large number of studies. In addition, meta-analysis allows the comparison of effect sizes in relation to various features of the study. Furthermore, this study incorporated the use of hierarchical linear modeling (HLM) to analyze the data (Raudenbush & Bryk, 2002).

HLM was used because it allowed for the analysis of fixed and random effects. The use of a mixed effects design is appropriate for meta-analysis when the studies included are relatively heterogeneous, when it is assumed that the studies included are a sample from a larger population of studies that have, or could potentially be, conducted, and when there is reason to believe that not all moderator variables can be identified and coded (Pearson & Lipton, 1999; Raudenbush, 1994). The present study met these conditions because the interventions conducted in the studies to be included in the meta-analysis are rather heterogeneous (e.g., tutorials, simulations, data analysis software, programmed instruction, online learning, etc.). Additionally, because the study of the use of technology to enhance instruction is ongoing, it is likely that the studies included in this meta-analysis are a mere sample of the universe of studies that have been, or will be, conducted. Furthermore, although a large number of moderator variables have been identified, it is likely that several have not, or are not identifiable based on the information provided in the written reports of several studies. Also, HLM is appropriate for the analysis of mixed effects designs in meta-analysis because the data is usually unbalanced, rendering ANOVA methods inappropriate (Raudenbush, 1994). Finally, HLM is an appropriate data analysis method for meta-analytic studies because the data, or individual studies, are already hierarchical in nature, as participants are nested within studies (Raudenbush & Bryk, 2002).

Research Questions

The research questions guiding this study include:

Research question 1. Is student achievement and attitude (e.g., satisfaction with the course, instructor, or content) affected by the use of technology to enhance statistics instruction?

Research question 2. Does significant between-study variance in effect sizes exist for both achievement and attitudes?

Research question 3. If significant between-study variance in effect sizes exists, can various study and treatment characteristics explain this between-study variance of effect sizes, including uses of technology (i.e., drill and practice, tutorial, simulations, statistical analysis, computer laboratories, enhanced lectures, web-based, programmed or personalized systems of instruction, and online), function of technology (supplemental vs. substitutive), treatment duration, academic discipline, level of statistics class (i.e., introductory, intermediate, or advanced), academic standing of students (i.e., undergraduate or graduate), instructor bias, randomization of participants (i.e., experimental vs. quasi-experimental studies), and publication status (i.e., published vs. unpublished)?

Literature Search

An extensive search of the literature was conducted in order to obtain as many studies as possible that have examined the effectiveness of technology use in statistics instruction. This study focused on studies of statistics instruction that have been enhanced with some form of computer or technology use. Several databases were used to obtain published and unpublished studies. The databases used to search for published studies were located through *EBSCOhost*, and included *Academic Search Premier*, *Pre-CINAHL*, *CINAHL*, *Education Research Complete*, *Health Source: Nursing/Academic Edition*, *Information Science & Technology Abstracts (ISTA)*, *Mental Measurements Yearbook*, *Psychology and Behavioral Sciences Collection*, *Sociological Collection*, *SocINDEX*, *Professional Development Collection*, and *Communication and Mass Media*. Also, searches of *PsycINFO*, and *Educational Abstracts* were conducted. The keywords used in this search included, “teaching statistics,” “statistics education,” and “statistics instruction.” The search results were then examined and limited to only those that studied the use of technology.

A search for unpublished studies was conducted in order to mitigate the effects of the “file drawer problem” (Rosenthal, 1979, p. 638). The “file drawer problem” refers to the tendency for studies where no significant results were obtained to remain unpublished. In order to address this issue, databases such as *Educational Resources Information Clearinghouse (ERIC)*, *Dissertation Abstracts International*, and *ProQuest Dissertations and Theses* were used to search for unpublished studies. In addition, a search of the World Wide Web, including *Google Scholar*, was conducted in order to locate additional published and unpublished studies. Furthermore, published and unpublished studies were also obtained through the examination of the reference lists of other acquired articles. Finally, the fail safe number was calculated to determine the number of additional studies with no effect size that it would take to overturn the statistical significance of the observed mean effect size in this study (Wolf).

Inclusion and Exclusion of Studies

Studies were included in the sample if they meet several criteria. First, the studies must have been conducted on uses of technology to enhance statistics instruction. Uses of technology included the use of computers and other electronic devices. Also, introductory, intermediate, and advanced statistics courses were included in the meta-analysis, as well as quantitative methods courses that are heavily statistical in focus. Second, studies were included if they were conducted using undergraduate or graduate students as participants. Third, the studies were included if the authors used an outcome variable that measured student achievement, such as classroom tests and quizzes, standardized test scores, or course grades, or student attitudes, such as course satisfaction, attitudes toward computers, attitude toward instruction, or attitudes toward subject matter. Fourth, studies were included in the meta-analysis if a control group was used.

Both experimental and quasi-experimental studies were included, as well as studies incorporating both pretest/posttest designs and posttest only designs. In addition, studies were included in the analysis if effect sizes can be determined from the statistics provided in the written report. Furthermore, both published and unpublished studies were included, provided the unpublished studies were able to be located.

Studies were eliminated from consideration if they failed to meet the inclusion criteria. For example, a study was eliminated if the authors compared teaching machines or other forms of programmed instruction not using computers or electronic technology to lecture courses. In addition, studies were eliminated if they include only primary and secondary students as participants, rather than undergraduate and graduate students. Pre-experimental and correlational designs were also eliminated. Studies have shown that pre-experimental, pretest/posttest designs tend to have significantly larger effect sizes than experimental and quasi-experimental designs (Liao, 1998, 1999). Finally, studies were eliminated from consideration if they possessed any obvious and crippling methodological flaws.

Coding of Studies

A coding sheet was developed to record and organize relevant information from each study examined in the meta-analysis, and is presented in Appendix A. The information to be coded included the following:

Uses of technology. The uses of technology were divided into several categories, including drill and practice, tutorial, simulations, statistical analysis, computer laboratories, technology-enhanced lectures, web-based, programmed or personalized system of instruction, and online learning.

Function of technology. Studies were coded based on whether the technology used was for supplemental or substitutive purposes. Supplemental uses of technology are those that are used to reinforce material presented through more traditional instructional methods, while substitutive uses of technology are those that present new material not presented by the course instructor.

Duration. Studies were coded based on how long they were used throughout a course. Uses of technology may vary considerably based on how often they are used within a class. Some may be used for an entire quarter or semester, while others may only be used for one classroom session. Duration was coded according to whether the technology was used once, several times to half semester/quarter, or entire semester/quarter or longer.

Academic discipline. Statistics courses are taught throughout a variety of academic disciplines in most higher education institutions, including education, psychology, sociology, business, nursing, mathematics, and the natural sciences.

Course level. Studies were coded according to whether the statistics course was taught at the introductory level or a more intermediate or advanced level.

Student academic standing. Studies were coded according to whether the students enrolled in the course were undergraduate or graduate students.

Instructor bias. Studies were coded based on whether the same, or different, instructors taught the experimental or control groups.

Research design. Studies were coded based on whether or not random assignment was used.

Publication source. Studies were coded based on whether they were published in peer-reviewed professional journals or through *ERIC* or *Dissertation Abstracts International*.

Outcome variable. The meta-analysis included studies that used achievement measures, such as classroom tests and quizzes, standardized tests, or course grades, and attitudinal measures, such as course satisfaction, attitude toward computers, attitude toward instruction, or attitude toward subject matter, as outcome measures.

In addition, studies were coded according to several additional study characteristics, including the year of the study, group means and standard deviations, number of participants in each group, use of a pretest, and *F* and/or *t* values if means and standard deviations are not available. A subset (20%) of all studies was coded by two raters, and the level of agreement among the raters was calculated using Cohen's kappa statistic (Rosenthal & Rosnow, 1991). If a disagreement occurred concerning the coding of variables for a particular study, the study was discussed among the raters so that a consensus can be reached.

Calculation of Effect Sizes

In order to compare the results across individual studies, standardized effect sizes were calculated (i.e., Cohen's *d*). The standardized effect size is calculated by subtracting the mean of the comparison or control group from the mean of the treatment group and dividing by the standard deviation of the control group, or the pooled standard deviation of the two groups (Wolf, 1986). The formula for the standardized mean difference between the experimental and control groups is:

$$d_j = (\overline{\mathrm{Y}}_{\mathrm{E}j} - \overline{\mathrm{Y}}_{\mathrm{C}j}) / S_j,$$

where

$\overline{\mathrm{Y}}_{\mathrm{E}j}$ is the mean of the outcome variable for the experimental group,

$\overline{\mathrm{Y}}_{\mathrm{C}j}$ is the mean of the outcome variable for the control group, and

S_j is the pooled, within-group standard deviation (Raudenbush & Bryk, 2002).

The formula for the pooled, within-group standard variance is:

$$S^2_j = [\Sigma_p (n_i^E - 1)(S_i^E)^2 + \Sigma_p (n_i^C - 1)(S_i^C)^2] / [\Sigma_p (n_i^E - 1) + \Sigma_p (n_i^C - 1)]$$

where

S_i^E are the experimental group standard deviations, and

S_i^C are the control group standard deviations (S. A. Kalaian, 2003).

Several authors provided effect size measures other than Cohen's *d*. When this occurred, the effect sizes were converted to Cohen's *d* using one of the formulas available in Wolf (1986) or Cooper and Hedges (1994). In addition, studies often included only *t* or *F* values or significance levels in their results. In such cases, the statistics were converted to standardized effect sizes using the formulas provided in Cooper and Hedges, Wolf, or Cortina and Nouri (2000). In addition, because studies often differ in the sample size used, the studies in the current meta-analysis were weighted by the inverse of the variance of the effect size. The variance of the estimated effect size is:

$$\sigma_d^2 = [(n_E + n_C) / (n_E n_C)] + \{d^2 / [2(n_E + n_C)]\}$$

(Cortina & Nouri, 2000). Using the inverse of this formula, effect sizes can be weighted so that studies with larger sample sizes are weighted higher than studies with lower sample sizes. Furthermore, the Cohen's *d* effect size has been shown to be a biased estimator, especially when computed using studies having small sample sizes, and was adjusted using the formula provided by Hedges (1981):

$$d^U = c(m)d$$

where,

d^U is the unbiased estimator of the effect size,

d is the biased estimator of the effect size, and

$c(m)$ is the correction factor:

$$c(m) \simeq 1 - 3/(4m - 1).$$

Analysis of Research Questions

The research questions were analyzed using the variance-known application of hierarchical linear modeling, as the sampling variances of the effect size estimates can be assumed to be known (Raudenbush & Bryk, 2002). The criterion variable for all analyses was the effect sizes obtained in the studies. The effect sizes associated with achievement and attitudinal outcome measures were analyzed separately. The predictor variables were dummy-coded so that the effect of each individual coding feature could be isolated. For example, studies that used tutorial software were coded as "1," whereas studies that used other types of software were coded as "0."

The hierarchical structure of meta-analytic data can be presented in two stages (S. A. Kalaian, 2003). The level-1, within-site (study) model specifies the observed effect size as a function of the true effect size and sampling error. The level-1, within-site model is:

$$d_j = \delta_j + e_j$$

for studies $j = 1,\ldots, J$, where e_j is the sampling error associated with d_j as an estimate of δ_j and for which we assume $e_j \sim N(0, V_j)$.

The level-2, between-sites (studies) model specifies the distribution of the true effect sizes as a function of study characteristics and random error (S. A. Kalaian, 2003). The level-2, between-sites model is:

$$\delta_j = \gamma_0 + \Sigma_s\, \gamma_s W_{sj} + u_j,$$

where

$W_{1j},\ldots, W_{sj}$ are study characteristics predicting effect sizes,

$\gamma_0,\ldots, \gamma_s$ are regression coefficients, and

u_j is a level-2 random error for which we assume $u_j \sim N(0, \tau)$ (Raudenbush & Bryk, 2002).

The combined model is:

$$d_j = \gamma_0 + \Sigma_s\, \gamma_s W_{sj} + u_j + e_j$$

The between-sites model can be expression in two forms, the unconditional and the conditional model. Analyzing the results of the unconditional model allows the researcher to examine the homogeneity of the study effect sizes. The conditional model assumes that the effect size depends on program or study characteristics and random error. In the unconditional model, no predictor variables are used, thus the combined model simplifies to:

$$d_j = \gamma_0 + u_j + e_j$$

This equation provides the grand-mean effect size (Raudenbush & Bryk, 2002). The unconditional model can then be tested to determine if significant variability exists across the true effect sizes. If so, the conditional model (i.e., the combined model) that includes predictor variables can be tested to determine if study or program characteristics account for significant reduction in variability among the true effect sizes. The software program, HLM 6.02, was used to conduct the statistical analyses.

CHAPTER IV

RESULTS

Chapter 4 provides the results of the analysis of the research questions. All analyses were conducted with the assistance of SPSS 12.0 (2005) or HLM 6.02 (Raudenbush et al., 2005). The chapter begins by providing an analysis of the inter-rater agreement on the coding of studies according to various characteristics. Second, the descriptive results for the meta-analysis of studies using measures of achievement as outcome variables are provided. Next, the hierarchical linear modeling results are provided, including both the results of the tests of the unconditional and conditional models. Finally, similar results are then provided for the meta-analysis of studies using attitudinal measures as outcome variables.

Inter-Rater Agreement

The author coded all primary studies. In addition, 20% (n = 11) of the studies were randomly selected to be coded by a second rater, who was trained in the coding process. When the results of the coding process were compared across raters, a level of agreement of 90.9% was obtained (Cohen's κ = 0.809). The raters then discussed the disagreements in order to obtain a consensus rating.

Achievement Results

A total of 117 effect sizes were obtained from 47 individual articles, thesis, or dissertations. The list of studies included in the meta-analysis of achievement outcomes is presented in Appendix B. Observed effect sizes varied from -0.59 to 2.93, with a mean of 0.443 and standard deviation of 0.626. Because Cohen's *d* effect size has been shown to be a biased estimator, especially when computed using studies having small sample sizes, effect sizes were adjusted using the formula provided by Hedges (1981). After the adjustment, the effect sizes varied from -0.58 to 2.92, with a mean of 0.436 and standard deviation of 0.615. The distribution of unbiased effect sizes is presented in Figure 1. The years of the studies ranged from 1974 to 2005. Total sample sizes of the individual studies ranged from 14 to 5599.

The frequencies of the study characteristics are presented in Table 1. As shown, the vast majority of the effect sizes were obtained using some form of statistics achievement test. However, a few effect sizes were calculated based on final course grades or total points scored in the course, or a retention test. A variety of types of technology were used throughout the studies obtained. However, few authors studied the effectiveness of computer laboratories, so this technology type was eliminated from subsequent analyses. Most forms of technology use were supplemental in nature, and most of the studies were conducted in the behavioral sciences fields, with psychology representing the majority. The education, mathematics, business, and nursing fields all contributed studies. However, only one effect size was found from the nursing discipline, so nursing was eliminated from subsequent analyses. The typical class used in the studies obtained was an undergraduate introductory statistics class. Most often, the studies made use of the same instructor to teach both the experimental and control conditions; however, several studies did not indicate whether the same or different instructors were

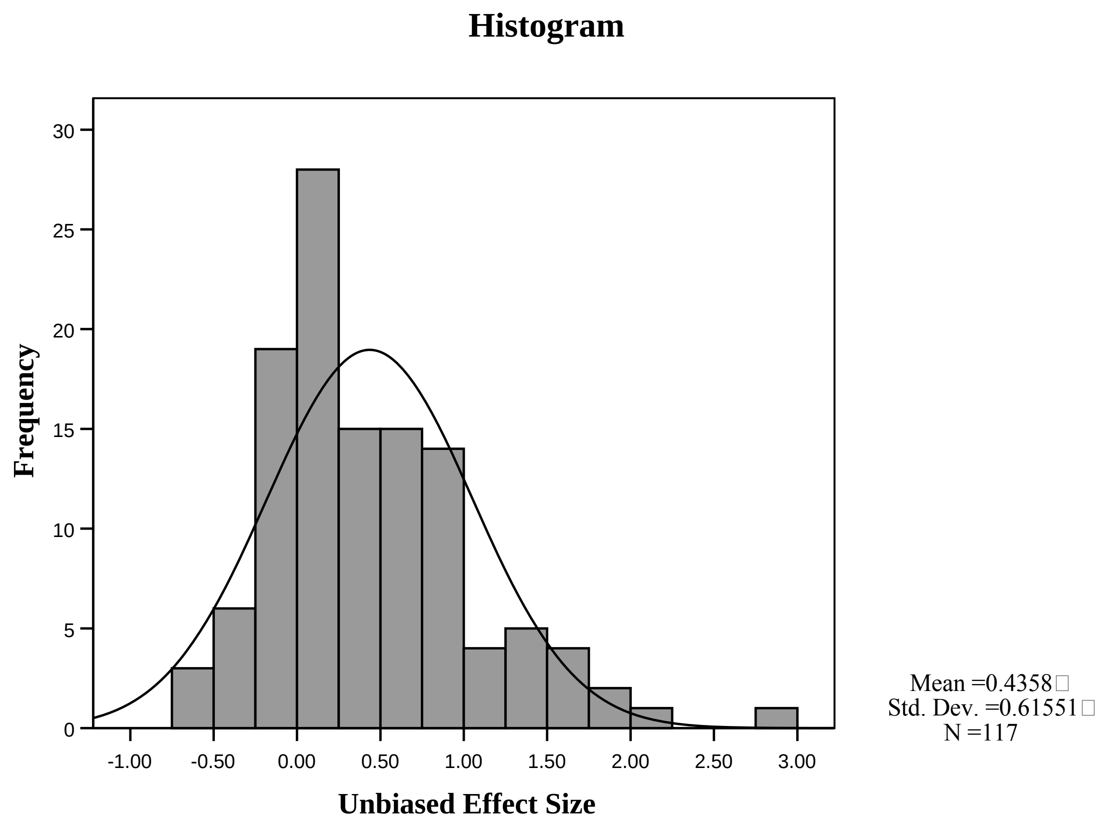

Figure 1. Distribution of unbiased effect sizes for studies measuring achievement.

used. Additionally, the studies included in the present meta-analysis were most often quasi-experimental in design, and included a control group that received some form of traditional instruction. However, a few studies included control groups that received no instruction at all. Furthermore, the majority of effect sizes came from studies that were published in professional journals. However, several effect sizes were obtained from dissertations, unpublished ERIC documents, and conference presentations.

Table 1

Frequencies of Primary Studies Measuring Achievement by Coding Characteristics

Variable	*n*	%
Outcome measure		
Total score in course or final grades	4	3.4
Achievement test	106	89.8
Retention test	8	6.8
Technology Type		
Drill and practice	11	9.4
Tutorial	21	17.9
Simulation	33	28.2
Statistical analysis software	26	22.2
Computer laboratory	2	1.7
Enhanced lecture	15	12.8
Web-based	21	17.9
Programmed instruction	12	10.3
Online learning	13	11.1
Function		
Supplemental	93	79.5
Substitutive	24	20.5
Duration		
Once	33	28.2
Several times to half semester/quarter	43	36.8
Full semester/quarter or longer	41	35.0

(*table continues*)

Table 1 (continued)

Frequencies of Primary Studies Measuring Achievement by Coding Characteristics

Variable	*n*	%
Academic discipline		
Social sciences (psychology, sociology, anthropology)	66	56.4
Business	13	11.1
Nursing	1	0.9
Mathematics	11	9.4
Education	26	22.2
Course type		
Statistics	112	95.7
Research methods	5	4.3
Course level		
Introductory	106	90.6
Intermediate or Advanced	11	9.4
Student academic standing		
Undergraduate	109	93.2
Graduate or both	8	6.8
Instructor bias		
Same instructor	98	83.8
Different instructor	19	16.2
Research design		
No random assignment	82	70.1
Random assignment	35	29.9

(*table continues*)

Table 1 (continued)

Frequencies of Primary Studies Measuring Achievement by Coding Characteristics

Variable	*n*	%
Instruction received by control group		
No instruction	19	16.2
Traditional instruction	98	83.8
Publication status		
Not published	20	17.1
Published	97	82.9

Hierarchical Linear Modeling Results: Unconditional Model

The level-1 analysis included 113 effect sizes from 46 studies, which were weighted by study and the inverse of their variances. Four effect sizes from one study were removed from the analysis because the sample size from this study was extremely large compared to all others, and thus was weighted much more heavily (i.e., Hilton & Christensen, 2002). Due to the potential impact of these effect sizes on the overall weighted effect size, the analysis was run without these data included. For those interested in viewing these results with all effect size estimates included, please see Appendix D.

The unbiased, unweighted mean effect size across the 113 effect sizes was found to be 0.45. The results of the unconditional HLM analysis are presented in Table 2.

Table 2

Hierarchical Linear Modeling Results for Primary Studies Measuring Achievement

Parameter	γ	$SE\ \gamma$	t	p
Unconditional Model				
Average achievement effect, γ_0	0.239	0.0690	3.463	0.001
Conditional Model				
Intercept, γ_0	8.268	16.777	0.493	0.626
Year, γ_1	-0.00395	0.00839	-0.471	0.641
Technology type				
Drill and practice, γ_2	-0.0641	0.129	-0.497	0.623
Tutorial, γ_3	0.0382	0.158	0.242	0.811
Simulation, γ_4	0.314	0.112	2.806	0.010
Statistical analysis software, γ_5	-0.314	0.162	-1.933	0.065
Enhanced lecture, γ_6	-0.264	0.168	-1.568	0.130
Web-based, γ_7	-0.208	0.147	-1.417	0.169
Programmed instruction, γ_8	-0.104	0.462	-0.225	0.824
Online learning, γ_9	-0.771	0.221	-3.485	0.002
Function, γ_{10}	-0.656	0.420	-1.562	0.131
Duration, γ_{11}	0.0237	0.101	0.234	0.817
Academic discipline				
Social sciences, γ_{12}	-0.00781	0.126	-0.062	0.951
Business, γ_{13}	-0.119	0.161	-0.735	0.469

(*table continues*)

Table 2 (continued)

Hierarchical Linear Modeling Results for Primary Studies Measuring Achievement

Parameter	γ	$SE\ \gamma$	t	p
Mathematics, γ_{14}	-0.147	0.129	-1.142	0.265
Instructor bias, γ_{15}	0.209	0.168	1.247	0.225
Course type, γ_{16}	0.158	0.191	0.831	0.414
Course level, γ_{17}	-0.915	0.238	-3.841	0.001
Student academic standing, γ_{18}	0.624	0.223	2.799	0.010
Research design, γ_{19}	0.000236	0.117	0.002	0.998
Type of control group used, γ_{20}	-0.411	0.199	-2.062	0.050
Publication status, γ_{21}	0.472	0.181	2.613	0.016

Note. Degrees of freedom for the unconditional model was 45. Degrees of freedom for the conditional model was 24.

The results indicated a mean study-weighted effect size of 0.239 across all studies. This value was statistically significant, t (46) = 3.463, p = 0.001, indicating that, overall, students who were instructed in statistics using some form of technology outperformed students in control or traditional instruction groups. Thus, the null hypothesis of a mean effect size of zero for achievement was rejected. However, the estimated variance of the effect parameter was also statistically significant, $\tau^2 = 0.161$, χ^2 (45) = 360.27, $p < 0.001$, indicating that considerable variability remained to be explained in the effect sizes. Thus, the second null hypothesis of the variance of the effect size parameters equaling zero was also rejected. As such, a conditional HLM analysis was conducted in order to examine the impact of several study characteristics on the variability of effect sizes.

In order to examine the “file drawer problem,” the fail-safe number (N_{fs}) was calculated to determine the additional number of studies that would need to be included in the current meta-analysis to overturn the results (Rosenthal, 1979, p. 638). Using an effect size value of 0.20 as a criterion, 8.97 additional primary studies each exhibiting a mean effect size of zero would be required to be added to the sample to reduce the current mean effect size to 0.20 or lower (N_{fs} = 46 * (0.239 - 0.20) / 0.20). Using an effect size value of 0.10 as a criterion, 63.94 primary studies each exhibiting a mean effect size of zero would have to be included to lower the current mean effect size to 0.10 (N_{fs} = 46 * (0.239 - 0.10) / 0.10).

Hierarchical Linear Modeling Results: Conditional Model

Because significant variability in achievement effect sizes was observed, a model was tested that included several study and methodological characteristics of the individual studies. These characteristics included the publication year, type of technology, function of technology, duration, academic discipline, course level, student academic standing, instructor bias, research design, and publication status. In addition, studies were coded based on the type of control group used in the study: traditional instruction versus no instruction. Because several studies included control groups that received no instruction, it seemed appropriate to include this variable in the analysis as one might expect larger effect sizes from studies where control group students received no instruction compared to studies where they received traditional instruction.

Variables were typically coded as dichotomous, with possession of a particular characteristic coded as “1” and lack of possession of that characteristic as “0.” For example, because a study could use more than one type of technology, technology type was not dummy-coded, but coded such that it would receive a code of “1” if a particular type was used, and a “0” if not. Additionally, studies were coded with a “1” if the technology used was supplemental and a “0” if it was substitutive, “1” for statistics courses and “0” for research methods, “1” for intermediate or advanced courses and “0” for introductory courses, “1” for graduate students and “0” for undergraduate students, “1” for different instructors and “0” for the same instructor, “1” for random assignment and “0” for non-random assignment, “1” if the control group received traditional instruction and “0” if it received no instruction, “1” if the study was published in a peer-reviewed journal and “0” if it was not, and “1” if the outcome was not a retention test and “0” if it was. Academic discipline was dummy-coded into three variables representing four fields: social sciences (i.e., psychology, sociology, anthropology), mathematics, business, nursing, and education, with education as the reference category. However, no studies in the nursing field were found that used achievement as an outcome variable, so nursing was eliminated from the analysis. Duration was not dichotomously coded or dummy-coded, but coded into three levels: one time (1), several times to half semester (2), and full semester or longer (3).

The results of the conditional analysis are presented in Table 2. Overall, the conditional model exhibited a statistically significant reduction in the variance of effect sizes over the unconditional model, χ^2 (21) = 203.65, $p < 0.001$. Thus, the third null hypothesis of all regression coefficients associated with study and methodological characteristics being equal to zero was rejected. As shown, several study and methodological characteristics were significantly related to effect size. Two uses of technology resulted in statistically significant differences in effect size: simulations and online learning. Simulations exhibited a mean effect size 0.314 standard deviations larger than studies examining other types of technology. Conversely, online learning exhibited a mean effect size 0.771 standard deviations smaller than other types of technology. Additionally, the course level variable had a statistically significant impact on the achievement effect sizes, with significantly larger effect sizes found in studies using introductory statistics classes rather than intermediate or advanced classes. Also, the student academic standing variable was statistically significant, with larger effect sizes found in studies using graduate students instead of undergraduate students. Furthermore, studies that included a control group that received traditional instruction had a significantly smaller mean effect size than studies where the control group received no instruction. Finally, studies published in peer-reviewed journals were found to exhibit significantly larger effect sizes than unpublished studies. Overall, the combination of predictor variables accounted for 41% of the variance in effect sizes. However, statistically significant variance in effect sizes remained after the conditional model was tested, $\tau^2 = 0.0952$, $\chi^2\ (25) = 156.61$, $p < 0.001$, suggesting that additional variables may exist that can further explain the differences between studies in observed effect sizes.

In order to determine the individual effect sizes, each variable was analyzed individually using separate hierarchical linear models. The results of the individual analyses are presented in Table 3, and the effect sizes based on those analyses are presented in Table 4. As shown, most of the types of technology used in statistics instruction exhibited positive effect sizes, with the exceptions of programmed instruction and online learning, which were close to zero and slightly negative, respectively.

Table 3

Individual Hierarchical Linear Modeling Results for Primary Studies Measuring Achievement

Parameter	γ	$SE\ \gamma$	*t*	*df*	*p*
Year					
Intercept	2.941	15.262	0.193	44	0.848
Coefficient	-0.00135	0.00766	-0.177	44	0.861
Drill and practice					
Intercept	0.209	0.0711	2.936	44	0.006
Coefficient	0.381	0.161	2.360	44	0.023
Tutorial					
Intercept	0.212	0.0727	2.913	44	0.006
Coefficient	0.217	0.205	1.056	44	0.297
Simulation					
Intercept	0.198	0.0752	2.631	44	0.012
Coefficient	0.184	0.186	0.992	45	0.327
Statistical analysis software					
Intercept	0.221	0.0717	3.086	44	0.004

(*table continues*)

Table 3 (continued)

Individual Hierarchical Linear Modeling Results for Primary Studies Measuring Achievement

Parameter	γ	$SE\ \gamma$	t	df	p
Coefficient	0.0669	0.179	0.374	44	0.710
Enhanced lecture					
Intercept	0.215	0.0784	2.742	44	0.009
Coefficient	0.142	0.144	0.986	44	0.330
Web-based					
Intercept	0.235	0.0792	2.973	44	0.005
Coefficient	0.0257	0.110	0.233	44	0.817
Programmed instruction					
Intercept	0.266	0.0756	3.521	44	0.001
Coefficient	-0.252	0.143	-1.769	44	0.083
Online learning					
Intercept	0.336	0.0729	4.606	44	0.000
Coefficient	-0.453	0.0958	-4.733	44	0.000
Function					
Intercept	0.265	0.0770	3.441	44	0.002
Coefficient	-0.180	0.157	-1.144	44	0.259
Duration					
Intercept	0.551	0.190	2.901	44	0.006
Coefficient	-0.133	0.0823	-1.613	44	0.113

(*table continues*)

Table 3 (continued)

Individual Hierarchical Linear Modeling Results for Primary Studies Measuring Achievement

Parameter	γ	*SE* γ	*t*	*df*	*p*
Academic discipline					
Intercept	0.485	0.173	2.798	42	0.008
Social sciences	-0.228	0.191	-1.194	42	0.240
Business	-0.380	0.257	-1.477	42	0.147
Mathematics	-0.497	0.179	-2.776	42	0.009
Instructor bias					
Intercept	0.278	0.0791	3.510	44	0.001
Coefficient	-0.201	0.145	-1.385	44	0.173
Course type					
Intercept	-0.0767	0.187	-0.410	44	0.683
Coefficient	0.327	0.200	1.638	44	0.108
Course level					
Intercept	0.244	0.0731	3.342	44	0.002
Coefficient	-0.0679	0.204	-0.333	44	0.740
Student academic standing					
Intercept	0.204	0.0622	3.281	44	0.002
Coefficient	0.320	0.322	0.994	44	0.326
Research design					
Intercept	0.220	0.0797	2.763	44	0.009

(*table continues*)

Table 3 (continued)

Individual Hierarchical Linear Modeling Results for Primary Studies Measuring Achievement

Parameter	γ	*SE* γ	*t*	*df*	*p*
Coefficient	0.0793	0.156	0.507	44	0.614
Type of control group used					
Intercept	0.468	0.0628	7.450	44	0.000
Coefficient	-0.245	0.0961	-2.552	44	0.015
Publication status					
Intercept	0.146	0.123	1.192	44	0.240
Coefficient	0.108	0.145	0.746	44	0.459

Table 4

Effect Sizes Based on Individual Hierarchical Linear Modeling Analyses by Variable for Primary Studies Measuring Achievement

Parameter	*d*
Year	
1975	0.275
1985	0.261
1995	0.248
2005	0.234

(*table continues*)

Table 4 (continued)

Effect Sizes Based on Individual Hierarchical Linear Modeling Analyses by Variable for Primary Studies Measuring Achievement

Parameter	*d*
Technology type	
Drill and practice	0.590
Tutorial	0.429
Simulation	0.382
Statistical analysis software	0.288
Enhanced lecture	0.357
Web-based	0.261
Programmed instruction	0.014
Online learning	-0.117
Function	
Supplemental (0)	0.265
Substitutive (1)	0.085
Duration	
One time (1)	0.418
Several times to half semester/quarter (2)	0.285
Full semester/quarter or longer (3)	0.152
Academic discipline	
Education (0, 0, 0)	0.485
Social sciences (1, 0, 0)	0.257

(*table continues*)

Table 4 (continued)

Effect Sizes Based on Individual Hierarchical Linear Modeling Analyses by Variable for Primary Studies Measuring Achievement

Parameter	*d*
Business (0, 1, 0)	0.105
Mathematics (0, 0, 1)	-0.012
Instructor bias	
Same instructor (0)	0.278
Different instructor (1)	0.077
Course type	
Research methodology (0)	-0.077
Statistics (1)	0.250
Course level	
Introductory (0)	0.244
Intermediate or advanced (1)	0.176
Student academic standing	
Undergraduate (0)	0.204
Graduate or both(1)	0.524
Research design	
No random assignment (0)	0.220
Random assignment (1)	0.299
Type of control group used	
No instruction (0)	0.468

(*table continues*)

Table 4 (continued)

Effect Sizes Based on Individual Hierarchical Linear Modeling Analyses by Variable for Primary Studies Measuring Achievement

Parameter	*d*
Traditional instruction (1)	0.223
Publication status	
Not published (0)	0.146
Published (1)	0.254

Attitude Results

A total of 91 effect sizes were obtained from 27 individual articles, thesis, or dissertations. The list of studies included in the meta-analysis of attitudinal outcomes is presented in Appendix C. Observed effect sizes varied from -1.59 to 2.08, with a mean of 0.158 and standard deviation of 0.681. Again, effect sizes were adjusted using the formula provided by Hedges (1981). After the adjustment, the effect sizes varied from -1.53 to 2.01, with a mean of 0.136 and standard deviation of 0.666. The distribution of adjusted effect sizes is presented in Figure 2. The years of the studies ranged from 1977 to 2005. Total sample sizes of the individual studies ranged from 9 to 291.

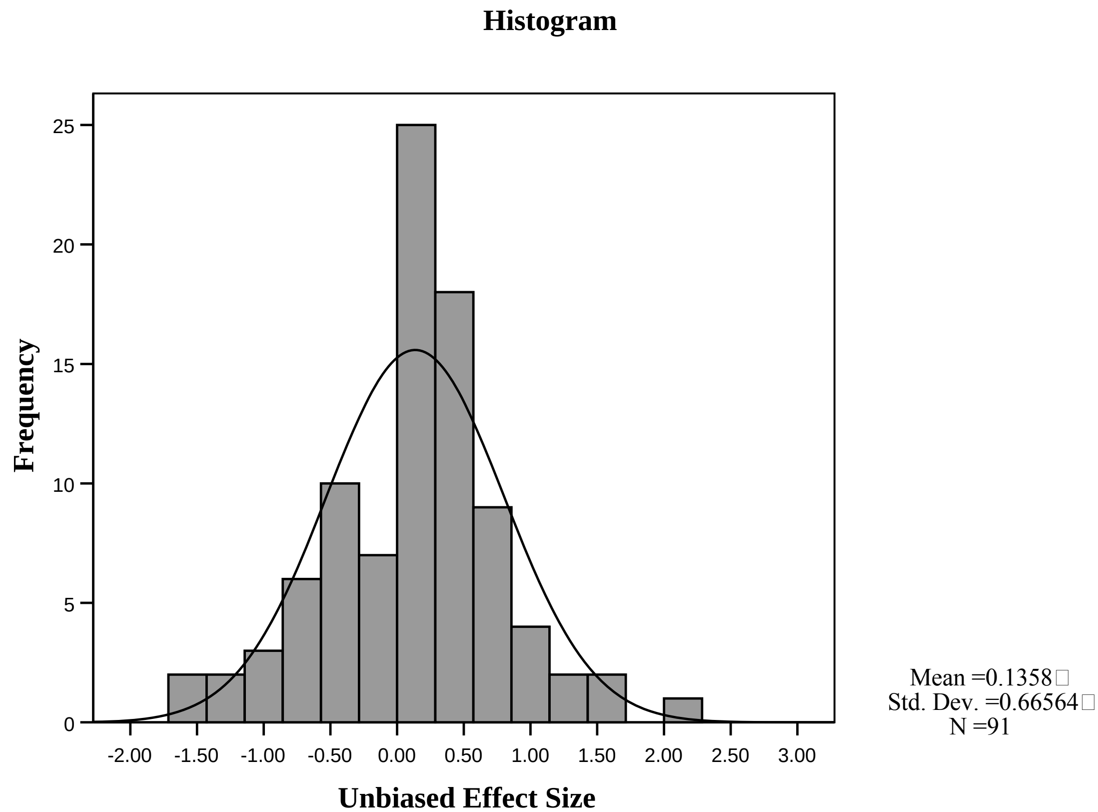

Figure 2. Distribution of unbiased effect sizes for studies measuring attitudes.

The frequencies of the study characteristics are presented in Table 5. As shown, the majority of the effect sizes were obtained using either a single global satisfaction measure or a summated satisfaction measure. However, a few effect sizes were calculated based on more specific attitudinal measures. Although a variety of types of technology were used throughout the studies, none measured attitudes after using drill and practice software, and few used computer laboratories. As a result, these two types of technology were eliminated from subsequent analyses. Again, the form of the technology use was most often supplemental in nature, and most of the studies were conducted in the

Table 5

Frequencies of Primary Studies Measuring Attitude by Coding Characteristics

Variable	*n*	%
Outcome measure		
Overall/mean course satisfaction	31	34.1
Satisfaction with instructor	11	12.1
Attitude toward mathematics/statistics	11	12.1
Confidence in course/statistics	10	11.0
Difficulty/clarity of course	9	9.9
Perceived learning	6	6.6
Value/relevance of statistics	5	5.5
Attitude toward computers	5	5.5
Retention/Attendance	2	2.2
Statistics Anxiety	1	1.1
Technology Type		
Drill and practice	0	0.0
Tutorial	10	11.0
Simulation	7	7.7
Statistical analysis software	18	19.8
Computer laboratory	3	3.3
Enhanced lecture	6	6.6
Web-based	17	18.7
Programmed instruction	17	18.7

(*table continues*)

Table 5 (continued)

Frequencies of Primary Studies Measuring Attitude by Coding Characteristics

Variable	*n*	%
Online learning	33	36.3
Function		
Supplemental	75	82.4
Substitutive	16	17.6
Duration		
Once	23	25.3
Several times to half semester/quarter	7	7.7
Full semester/quarter or longer	61	67.0
Academic discipline		
Social sciences (psychology, sociology, anthropology)	42	46.2
Business	10	11.0
Nursing	19	20.9
Mathematics	6	6.6
Education	14	15.4
Course type		
Statistics	76	83.5
Research methods	15	16.5
Course level		
Introductory	87	95.6
Intermediate or Advanced	4	4.4

(*table continues*)

Table 5 (continued)

Frequencies of Primary Studies Measuring Attitude by Coding Characteristics

Variable	*n*	%
Student academic standing		
Undergraduate	78	85.7
Graduate or both	13	14.3
Instructor bias		
Same instructor	66	72.5
Different instructor	25	27.5
Research design		
No random assignment	64	70.3
Random assignment	27	29.7
Publication status		
Not published	17	18.7
Published	74	81.3

behavioral sciences fields, with psychology representing the majority. The education, mathematics, business, and nursing fields all contributed studies. The typical class used in the studies obtained was an undergraduate introductory statistics class. Most often, the studies made use of the same instructor to teach both the experimental and control conditions; however, several studies did not indicate whether the same or different instructors were used. Additionally, the studies included in the present meta-analysis were most often quasi-experimental in design, and included a control group that received some form of traditional instruction. However, a few studies included control groups that received no instruction at all. Furthermore, the majority of effect sizes came from studies that were published in professional journals. However, several effect sizes were obtained from dissertations, unpublished ERIC documents, and conference presentations.

Hierarchical Linear Modeling Results: Unconditional Model

The level-1 analysis included 91 effect sizes from 27 studies, which were weighted by study and the inverse of their variances. Because of the wide number of attitudinal measures used, all measures were analyzed as a group, rather than individually. The mean unbiased, unweighted effect size across the 91 effect sizes was found to be 0.14. The results of the unconditional HLM analysis are presented in Table 6. The results indicated a mean, study-weighted, effect size of 0.163 across all studies. This value was not statistically significant, t (26) = 1.116, p = 0.275, indicating that, overall, students who were instructed in statistics using some form of technology had similar attitudes to students in traditional instruction groups. Thus, the null hypothesis of a mean effect size equal to zero was not rejected. However, the estimated variance of the effect parameter was statistically significant, $\tau^2 = 0.263$, χ^2 (26) = 569.84, $p < 0.001$, indicating that considerable variability remained to be explained in the effect sizes. Thus, the second null hypothesis of the variance of effect size parameters being equal to zero was rejected. As such, a conditional HLM analysis was conducted in order to examine the impact of several study characteristics on the variability of effect sizes. The fail safe number was

Table 6

Hierarchical Linear Modeling Results for Primary Studies Measuring Attitudes

Parameter	γ	$SE\ \gamma$	t	p
Unconditional Model				
Average attitudinal effect, γ_0	0.162	0.147	1.102	0.281
Conditional Model				
Intercept, γ_0	-124.600	44.287	-2.813	0.031
Year, γ_1	0.0625	0.0225	2.785	0.032
Technology type				
Tutorial, γ_2	-0.745	0.627	-1.187	0.280
Simulation, γ_3	0.659	0.340	1.941	0.099
Statistical analysis software, γ_4	-1.542	0.405	-3.806	0.012
Enhanced lecture, γ_5	-2.034	0.578	-3.518	0.016
Web-based, γ_6	-1.019	0.231	-4.416	0.005
Programmed instruction, γ_7	0.578	0.400	1.444	0.199
Online learning, γ_8	-1.445	0.364	-3.967	0.010
Function, γ_9	-1.317	0.483	-2.729	0.035
Duration, γ_{10}	0.466	0.122	3.818	0.011
Academic discipline				
Social sciences, γ_{11}	0.359	0.185	1.941	0.099
Business, γ_{12}	-0.058	0.303	-0.193	0.854
Nursing, γ_{13}	-1.808	0.479	-3.774	0.012

(*table continues*)

Table 6 (continued)

Hierarchical Linear Modeling Results for Primary Studies Measuring Attitudes

Parameter	γ	$SE\ \gamma$	t	p
Mathematics, γ_{14}	1.055	0.246	4.282	0.006
Instructor bias, γ_{15}	0.347	0.187	1.852	0.113
Course type, γ_{16}	-1.223	0.480	-2.548	0.043
Course level, γ_{17}	-1.079	0.549	-1.963	0.096
Student academic standing, γ_{18}	-0.545	0.545	-1.000	0.356
Research design, γ_{19}	-0.980	0.443	-2.214	0.068
Publication status, γ_{20}	1.169	0.235	4.977	0.001

Note. Degrees of freedom for the unconditional model was 26. Degrees of freedom for the conditional model was 6.

not calculated for the results of the analysis of attitudes as the mean effect size was not statistically significant.

Hierarchical Linear Modeling Results: Conditional Model

Because significant variability in attitudinal effect sizes remained, a model was tested that included several study and methodological characteristics of the individual studies. These characteristics included the publication year, use of technology, function of technology, duration, academic discipline, course level, student academic standing, instructor bias, research design, and publication status. Unlike the analysis for the achievement outcome variable, the type of control group used was not included as a variable, as it was assumed that all the control groups would have had some form of instruction in order to compare attitudes. Additionally, the type of outcome measures was not included in the model as several outcome measures had very few studies associated with them. Again, the variables in the model were coded and dummy-coded in a similar fashion to the analysis of the achievement effect sizes. However, studies of attitudes were found in the nursing field, so the Academic Discipline variable was coded into four variables representing five fields, with Education again being the reference category.

The results of the conditional analysis are presented in Table 6. Overall, the conditional model exhibited a statistically significant reduction in the variance of effect sizes over the unconditional model, χ^2 (20) = 352.53, $p < 0.001$. Thus, the third null hypothesis of all regression coefficients associated with study and methodological characteristics being equal to zero was rejected. As shown, several study and methodological characteristics were significantly related to effect size. Four uses of technology resulted in significantly smaller than average effect sizes when all other variables are included in the model: statistical analysis software, enhanced lecture, Web-based, and fully online courses. For example, in this model, online learning had a mean effect size 1.445 standard deviations lower than the average of other uses of technology. Additionally, year was statistically significant, with newer studies exhibiting larger effect sizes. Furthermore, the function of technology was statistically significant, with students preferring supplemental over substitutive functions of technology. Also, students significantly preferred technology uses that lasted for longer durations over shorter ones. When comparing academic disciplines, the results indicated that studies conducted in mathematics provided significantly more positive student attitudes, while nursing provided significantly less positive student attitudes toward statistics classes using technology. Additionally, course type was statistically significant, with students in statistics courses having more positive attitudes than students in research methods courses. Finally, the source of the study was statistically significant, with published studies possessing much larger effect sizes than studies that had not been published. However, statistically significant variance in effect sizes remained after the conditional model was tested, $\tau^2 = 0.601$. χ^2 (6) = 217.31, $p < 0.001$, suggesting that additional variables may exist that can further explain the differences between studies in observed effect sizes.

In order to determine the individual effect sizes, each variable was entered into the hierarchical analysis individually. The results of the individual analyses are presented in Table 7, and the effect sizes based on those analyses are presented in Table 8. As shown, most of the types of technology used in statistics instruction exhibited positive effect sizes for attitudes with the exception of online learning. Additionally, students tended to prefer supplemental uses of technology, and disliked substitutive functions. Furthermore, students in mathematics, education, and the social sciences tended to have positive attitudes toward the use of technology in instruction, while students in the nursing and business fields did not.

Table 7

Individual Hierarchical Linear Modeling Results for Primary Studies Measuring Attitudes

Parameter	γ	*SE* γ	*t*	*df*	*p*
Year					
Intercept	31.151	31.619	0.985	25	0.334
Coefficient	-0.0155	0.0159	-0.976	25	0.339
Tutorial					
Intercept	0.166	0.177	0.942	25	0.356
Coefficient	-0.0251	0.189	-0.133	25	0.896
Simulation					
Intercept	0.163	0.153	1.065	25	0.298
Coefficient	-0.0257	0.320	-0.081	25	0.937
Statistical analysis software					
Intercept	0.106	0.184	0.575	25	0.570
Coefficient	0.228	0.218	1.049	25	0.305
Enhanced lecture					
Intercept	0.128	0.161	0.797	25	0.433
Coefficient	0.285	0.174	1.636	25	0.114
Web-based					
Intercept	0.158	0.173	0.916	25	0.369
Coefficient	0.0233	0.189	0.124	25	0.902

(table continues)

Table 7 (continued)

Individual Hierarchical Linear Modeling Results for Primary Studies Measuring Attitudes

Parameter	γ	$SE\ \gamma$	t	df	p
Programmed instruction					
Intercept	0.144	0.162	0.887	25	0.384
Coefficient	0.172	0.182	0.946	25	0.354
Online learning					
Intercept	0.291	0.0656	4.428	25	0.000
Coefficient	-0.400	0.383	-1.043	25	0.307
Function					
Intercept	0.272	0.0697	3.895	25	0.001
Coefficient	-0.386	0.431	-0.895	25	0.379
Duration					
Intercept	0.201	0.151	1.326	25	0.197
Coefficient	-0.0117	0.0823	-0.143	25	0.888
Academic discipline					
Intercept	0.361	0.150	2.411	22	0.025
Social sciences	-0.127	0.168	-0.753	22	0.459
Business	-0.429	0.472	-0.907	22	0.374
Nursing	-0.778	0.342	-2.274	22	0.033
Mathematics	-0.108	0.157	-0.689	22	0.498

(*table continues*)

Table 7 (continued)

Individual Hierarchical Linear Modeling Results for Primary Studies Measuring Attitudes

Parameter	γ	$SE\ \gamma$	t	df	p
Instructor bias					
Intercept	0.153	0.187	0.814	25	0.424
Coefficient	0.0407	0.206	0.198	25	0.845
Course type					
Intercept	0.147	0.0821	1.789	25	0.085
Coefficient	0.0172	0.186	0.093	25	0.927
Course level					
Intercept	0.173	0.108	1.602	25	0.121
Coefficient	-0.139	0.407	-0.341	25	0.736
Student academic standing					
Intercept	0.142	0.157	0.902	25	0.376
Coefficient	0.246	0.257	0.958	25	0.348
Research design					
Intercept	0.168	0.181	0.927	25	0.363
Coefficient	-0.0315	0.196	-0.160	25	0.874
Publication status					
Intercept	0.112	0.0907	1.235	25	0.229
Coefficient	0.0602	0.198	0.304	25	0.764

Table 8

Effect Sizes Based on Individual Hierarchical Linear Modeling Analyses by Variable for Primary Studies Measuring Attitudes

Parameter	*d*
Year	
1975	0.539
1985	0.384
1995	0.229
2005	0.074
Technology type	
Tutorial	0.140
Simulation	0.137
Statistical analysis software	0.334
Enhanced lecture	0.413
Web-based	0.181
Programmed instruction	0.316
Online learning	-0.109
Function	
Supplemental (0)	0.272
Substitutive (1)	-0.114
Duration	
One time (1)	0.189
Several times to half semester/quarter (2)	0.178

(*table continues*)

Table 8 (continued)

Effect Sizes Based on Individual Hierarchical Linear Modeling Analyses by Variable for Primary Studies Measuring Attitudes

Parameter	*d*
Full semester/quarter or longer (3)	0.166
Academic discipline	
Education (0, 0, 0, 0)	0.361
Social sciences (1, 0, 0, 0)	0.234
Business (0, 1, 0, 0)	-0.068
Nursing (0, 0, 1, 0)	-0.417
Mathematics (0, 0, 0, 1)	0.253
Instructor bias	
Same instructor (0)	0.153
Different instructor (1)	0.194
Course type	
Research methodology (0)	0.147
Statistics (1)	0.164
Course level	
Introductory (0)	0.173
Intermediate or advanced (1)	0.034
Student academic standing	
Undergraduate (0)	0.142
Graduate or both (1)	0.388

(*table continues*)

Table 8 (continued)

Effect Sizes Based on Individual Hierarchical Linear Modeling Analyses by Variable for Primary Studies Measuring Attitudes

Parameter	*d*
Research design	
No Random assignment (0)	0.168
Random assignment (1)	0.137
Publication status	
Not published (0)	0.112
Published (1)	0.172

CHAPTER V

DISCUSSION

The present meta-analysis endeavored to advance the research literature with an updated examination of the effectiveness of the use of technology for statistics instruction. The goal was to determine whether or not the use of technology was indeed effective, and which types were most effective in improving student achievement and attitudes. Additionally, other study and methodological characteristics were examined in order to determine several of the conditions under which technology might be most effective. The chapter begins with a summary of the major findings pertaining to the answering of the research questions. Next, a discussion of the results is presented in comparison to the results of previous meta-analyses covering the same or similar topics. Third, a discussion of the contributions of the study to the research literature is provided, including a discussion of the practical implications of the results. Finally, the chapter concludes with a listing of the limitations of the study, followed by recommendations for future research and concluding remarks.

Summary of Findings

Three research questions guided this study. The first research question involved determining the overall effect of using technology to enhance statistics instruction on student achievement and attitudes. The results indicated that technology did have an overall effect on achievement but not on attitudes. For achievement, a statistically significant mean effect size of 0.233 was found. However, the mean effect size for attitude was found to be 0.163, which was not statistically significant.

The second research question involved determining if significant between-study variation in effect sizes existed in order to warrant further analysis into the sources of this variation. Analyses revealed that significant variation did exist between studies in effect sizes for both achievement and attitude criterion variables. The variation across effect sizes was statistically significant, indicating the need for further exploration of study and methodological characteristics that could explain some of the differences between studies of technology use in statistics instruction.

Finally, the third research question involved determining if various study and treatment characteristics could help in explaining some of the variation in effect sizes between studies. The results indicated that several study and treatment characteristics could explain some of the variance in effect sizes between studies. For achievement outcome variables, several variables were found to explain significant variance in effect sizes, including technology type, course level, student academic standing, and publication status. For the attitudinal outcome variable, several variables were also related to effect size, including technology type, year, function of technology, duration of use, academic discipline, course type, and publication status. However, the third research question remained partially unsupported as statistically significant variance in effect sizes remained even after including all coded study and treatment characteristics, indicating that additional characteristics that were not accounted for may contribute to different effect sizes across studies.

Discussion

In the present study, a statistically significant mean effect size was found for achievement outcomes, but not attitude. The modest, but significant, mean achievement effect size of 0.239 found in this study was comparable to effect sizes found in previous studies. For example, Christmann and Badgett (1999) conducted a meta-analysis of the effectiveness of computer-based software packages for teaching statistics, and found an almost identical mean effect size of 0.256 across 14 effect sizes from 9 studies. However, the present study included much broader uses of technology beyond simple software programs. Hsu (2003) also included a wider array of technology types, and found a larger mean effect size of 0.43 from 31 effect sizes found in 25 studies. Furthermore, the mean effect size found here was similar to those found in meta-analyses that included studies across a variety of academic topics, such as C. C. Kulik et al. (1980) who found a mean effect size of 0.28; Kulik and Kulik (1986) who found a mean effect size of 0.26; Christmann et al. (1997b) who found a mean effect size of 0.19; and Timmerman and Kruepke (2006) who found a mean effect size of 0.24.

Also, like several previous meta-analyses, the mean effect size for attitude outcome measures for this study did not significantly differ from zero. For example, C. C. Kulik et al. (1980) obtained an effect size of 0.10 for students' ratings of course quality. Similarly, Lee (1999) obtained a mean effect size of -0.04 for attitudes across 5 effect sizes, and Allen et al. (2002) found that students exhibited little difference in satisfaction for online learning over traditional instruction. Additionally, Kulik and Kulik (1986) found a mean effect size of -0.03 for attitude toward subject matter, but did find higher effect sizes for attitude toward computers (0.37) and attitude toward instruction (0.31), indicating that different types of attitudes may be differentially affected by technology use. Additional studies have found larger effect sizes as well, including Waxman, Lin, and Michko (2003), who obtained a study-weighted mean effect size for affective outcomes of 0.464, although this meta-analysis included only 10 studies and the confidence interval around this effect size included zero. Also, Kulik and Kulik (1987) found that computer-based instruction resulted in a mean effect size of 0.28 for attitude-toward-instruction and 0.33 for attitudes toward computers when compared to traditional instruction. However, they also found a mean effect size of 0.05 for attitude toward subject, suggesting that the difference between the results found here and those found by Kulik and Kulik might be the result of having combined all affective measure into one index in the present study instead of separating them into separate affective measures. Also, these studies examined technology use across a variety of disciplinary fields, whereas the present study focused only on statistics instruction.

Several explanations are plausible as to why technology was found to have a significant effect on achievement, but not attitudes. First, a student’s achievement might be affected by fewer factors than his or her attitude. Overall, the student’s own motivation and the quality of instruction may lead to greater achievement, both of which could be influenced by the use of technology to enhance instruction. However, attitudes may be affected by, or include, far more factors, such as students’ enjoyment of the material, liking of the instructor, liking of other classmates, perceived difficulty of the course, the number of assignments, perceived fairness of the grading, or even the time of day when the class meets. Technology is unlikely to have much of an effect on several of these factors. Additionally, students are generally apprehensive toward statistics courses, and the addition of technology simply might not be enough to help them overcome their anxiety. Furthermore, attitude variables may be less stable than achievement variables leading to greater error in measurement. Although authors, such as Schimmack, Diener, and Oishi (2002) have shown that global attitudes (e.g., life satisfaction) tend to be relatively stable, more specific attitudes are less so (e.g., current mood). Applied to the current context, students' attitudes toward school may be fairly stable, whereas their attitudes toward a particular instructor or class may vary much more from day to day. Finally, this result could have occurred due to differences in statistical power. The number of studies included and the resulting mean effect size were both lower for the analysis of attitude effect sizes than for the analysis of achievement effect sizes.

Overall, a great deal of variance was found in the observed achievement and attitude effect sizes, indicating that several study and methodological characteristics could be related to the magnitude of the effect sizes. For example, the present study found smaller achievement effect sizes for studies that examined online learning when compared to other types of technology use. Olson and Wisher (2002) found similar results when comparing web-based instruction and fully online courses. Web-based instruction was found to have a moderate effect size of 0.48, whereas fully online courses had a mean effect size near zero, at 0.08. However, researchers commonly have found near zero mean effect sizes for the effectiveness of distance learning courses for improving students' achievement throughout the literature, indicating that online and distance courses are often no worse, but no better than traditional instruction (Allen et al., 2004; Bernard et al., 2004). Perhaps statistics courses are particularly ill-suited for online learning environments, as such courses are generally qualitatively different from other subject areas. For example, several topic areas are much more content-based and require more rote memorization of material, whereas statistics requires a greater conceptual understanding of ideas and theories. Instructional methods such as programmed instruction and online learning may be more suitable for classes that are based on content knowledge and/or discussion of course material, where the instructor's primary role would be to guide students to sources of information. Also, students often struggle, initially, with learning statistical concepts, and not having an instructor in close proximity and readily available may be discouraging for many students. Students in statistics courses often benefit from having individual concepts explained in a multitude of ways and using a variety of examples; tactics that instructors may not be able to fully implement online. With face-to-face contact, students can more easily ask questions and obtain immediate feedback. For example, verbally conveying a concept to a student may have an entirely different effect than conveying the same topic through only emails, text messages, or discussion boards. Furthermore, online courses require students to have a certain level of self-motivation and self-discipline, such that they may not be entirely suitable for all students.

Additionally, studies using simulation software programs were found to possess significantly larger effect sizes than other technology types. Authors such as Bayraktar (2002) also found larger effect sizes for simulation software programs than other technology types, although he did find similar effect sizes for tutorials and simulations. Likewise, Lee (1999) found an overall effect size of 0.41 for simulations. Simulation software programs may be more effective than other types simply because they often combine the elements of other types such as drill and practice, tutorials, or statistical analysis software, and add a more interactive component which allows students to manipulate data and explore relationships (J. A. Kulik, C. C. Kulik, & Cohen, 1980). For example, although a statistical analysis software program might only provide a table with summary statistics and perhaps a graphing option, a simulation might include an interactive graph that allows students to examine how the graph changes when various characteristics of the data are changed, such as the addition or subtraction of outliers, the increase or decrease of sample size, and so forth. Additionally, a simulation program could be used to visually demonstrate concepts such as sampling or the Central Limit Theorem in an interactive manner (e.g., Aberson et al., 2000), whereas drill and practice and tutorial software programs may only provide descriptions or definitions, perhaps augmented with graphs or charts, more in the fashion of a computerized textbook.

Furthermore, the current study echoed the results of previous studies, such as Kulik and Kulik (1987, 1989), which found larger effect sizes for published rather than unpublished studies. This could be the result of the “file drawer” problem whereby studies are not as likely to be published if statistically significant results are not obtained. An alternative explanation is that many of the unpublished effect sizes may have come from thesis and dissertations, which might not be as methodologically sound as published studies.

However, the present study obtained several results that differed from those of previous meta-analyses. For example, several studies, including C. C. Kulik et al. (1980), found larger effect sizes (0.55) for programmed or personalized systems of instruction than for other types of technology. In the present study, however, the mean effect size of programmed or personalized systems of instruction was not found to differ significantly from that of other types of technology, and was found to be near zero (0.014). Additionally, Kulik and Kulik (1989) found that effect sizes were considerably larger when different instructors (0.39) taught the experimental and control groups rather than the same instructor (0.25) teaching both. Whereas this difference was found here, it was relatively small (0.194 vs. 0.153) and not statistically significant. Kulik and Kulik also found larger effect sizes for shorter duration interventions (0.42) over longer ones (0.26). Again, this result was not found to be statistically significant here, but the trend was in the same direction, with one time uses of technology exhibiting a mean effect size of 0.418 compared to a mean effect size of 0.152 for semester long uses of technology. Additionally, Christmann et al. (1997b) and Timmerman and Kruepke (2006) found smaller effect sizes for newer studies over older ones. Christmann et al. obtained a correlation of -0.762 between year and effect size, whereas Timmerman and Kruepke found a statistically significant difference between year groups, with an effect size of 0.22 for more recent studies and 0.28 for older studies. Again, the trend in the present study was in the same direction, despite not being statistically significant. For example, a study published in 1975 would have been expected to obtain an effect size of 0.275, whereas a study published in 2005 would have been expected to obtain an effect size of

0.234. Furthermore, Kulik et al. (1986) found larger effect sizes for studies of shorter (0.45), rather than longer (0.37), durations, results that were not found to be statistically significant here. Again, however, the trend was in the same direction, with effect sizes ranging from 0.418 for one-time uses of technology to 0.152 for semester-long uses of technology. Finally, Bayraktar (2002) found a mean effect size for drill and practice software programs (-0.107), which differed from simulations (0.391) and tutorials (0.369). However, in the present study, all three exhibited positive effect sizes: 0.590, 0.382, and 0.429, respectively. However, Bayraktar specifically studied uses of technology in science, whereas this study was limited to statistics instruction. Also, Bayraktar found different mean effect sizes for supplemental (0.288) versus substitutive (0.178) uses of technology. Again, although not statistically significant in the present study, the trend was similar, with a larger mean effect size for supplement (0.265) over substitutive (0.085) forms of technology.

As indicated, several previous meta-analyses found larger effect sizes for older, rather than newer, studies (Bayraktar, 2002; Christmann et al., 1997b; Kulik, 1983; Timmerman & Kruepke, 2006). Although the regression coefficient was in the same direction, it was not statistically significant in the current study. This was a somewhat surprising finding since it appears fairly consistent across previous meta-analyses, suggestive of a novelty effect (Kulik & Kulik, 1989). However, statistics was among the first subjects to regularly incorporate technology into instruction (R. L. Rogers, 1987), so the novelty effect may have worn off much sooner, or the types of technology used in statistics are so different from those used in other disciplines that they have remained novel to the present day. For example, students make regular use of the Internet and word processing and presentation software in a variety of classes, but data analysis software programs may be entirely unique to statistics courses, and therefore completely new to beginning statistics students. Also, continual improvement of technology throughout the years could offset some of the diminishing novelty effect.

Additionally, although various authors have posited larger effect sizes in later years due to continual improvement of technology, this result was not obtained in the present study, and has infrequently been demonstrated in the research literature (Bayraktar, 2002). Several explanations may account for this counterintuitive finding. The first, and most obvious, explanation is that this is again the result of a novelty effect, whereby technology was new, interesting, and exciting to students in early studies, whereas less so in more recent studies. Second, although software programs are expected to become better and more sophisticated over the years, this is often not the case, especially for the software that was often used in the primary studies included in the present meta-analysis. Software programs such as SPSS, and Microsoft® Word, Excel, and PowerPoint® have not fundamentally changed in at least a decade. They may have gained additional features, but the basic, most frequently used features remain intact. Third, year may be much less important than other variables, such as technology type or various methodological variables. Furthermore, year might be correlated with technology type, such that certain types of technology were more common in early studies, while different types were more common in later studies. For example, studies of online learning tend to be more recent than studies of other types of technology, and online learning was found to be less effective than other types.

Also, previous studies have found that treatments that involved technology for shorter durations were more effective than those lasting longer durations (Bayraktar, 2002; Kulik, 1983; Kulik & Kulik, 1987). Again, this is often attributed to a novelty affect (Kulik & Kulik, 1989). However, this result was not obtained here, although the statistically non-significant regression coefficient trended in that direction. This is another counterintuitive result throughout the research literature, as one might expect longer studies, where technology was used more extensively, to exhibit larger effect sizes. Indeed, some authors have found this (Khalili & Shashaani, 1994). However, the meta-analytic study by Khalili and Shashaani examined the use of software programs, specifically, rather than a variety of different technologies. Overall, however, meta-analyses have generally shown shorter duration interventions to be associated with larger effect sizes than longer ones. Several plausible reasons for this exist. First, shorter studies tend to be methodologically stronger, and less affected by outside variables or sources of error. Longer studies may suffer from greater diffusion of treatment, whereby students in the experimental group provide information about the intervention to members of the control group. For example, members of the experimental and control groups may form study groups, where information is passed back and forth between them. Additionally, shorter duration studies may use technology less extensively, but more intensively. A longer duration study might use the technology on an "as-needed" basis throughout a semester, whereas a shorter study might use the technology in one session for the sole purpose of studying its effects. Finally, like year, the duration variable may be correlated with other variables, such as technology type. Studies of online learning and statistical analysis software, for example, tend to be longer in duration (i.e., entire semesters), whereas studies of tutorials, simulations, and drill and practice software programs tend to be shorter (i.e., one time).

Furthermore, this study found no effect for instructor bias. Studies with the same instructor teaching both the treatment and control groups did not significantly differ in mean effect size from studies with different instructors teaching each treatment condition. Several previous studies have found instructor bias, but this result has been rather inconsistent. For example, J. A. Kulik, C. C. Kulik, and Cohen (1980) found larger effect sizes when the study used different instructors than when the same instructor was used. Bayraktar (2002) and Khalili and Shashaani (1994) found similar results. Liao (1998), on the other hand, found that studies where the same instructor taught both treatment conditions exhibited larger effect sizes than those with different instructors. Overall, little consistency exists across the results of previous meta-analyses concerning instructor bias, so the result here is not as discrepant as it might immediately appear.

The present study also examined the same set of study and methodological characteristics in relation to attitudinal outcomes. Unfortunately, few previous studies thoroughly examined these characteristics in relation to attitudinal outcomes; rather, most simply reported overall mean effect sizes. In the present study, online learning, enhanced lectures, web-based instruction, and statistical analysis software were found to result in significantly smaller effect sizes in relation to attitudinal outcomes than other technology types. This result could be due to the fact that technology types such as enhanced lectures, web-based instruction, and statistical analysis software may only result in relatively minor changes in the instructional methods or classroom environment, and may, in some cases result in additional problems or difficulties. For example, some students may find statistical analysis software programs difficult to use, and therefore more of a hassle than they are worth. Web-based instruction could be particularly difficult for students who do not readily have access to the Internet; for example, the student who has to drive to campus to use the computer laboratory in the library to gain access.

Additionally, significantly smaller, and possibly negative, effect sizes for attitudes were found for online learning. Several possible explanations exist for why students do not prefer online learning environments over other types of technology and possibly even traditional instruction. For example, some students might feel that they do not have adequate access to or contact with the instruction, and might not feel that responses to their questions or feedback are provided in a timely manner. Additionally, students may have misconceptions about online courses, believing that they may be easier or require less work than other types of courses. Others may, after beginning an online course, find the distractions of working from home difficult to bear. Also, students in online courses may experience technical difficulties with some of the course software, or find that their home computers are not compatible. Finally, as mentioned above, students in online courses must be more self-motivated, making such courses less suitable for some students.

Several additional results were found concerning the relationship between various study and methodological characteristics and attitudinal effect sizes. For example, effect sizes were found to be significantly larger for supplemental forms of technology rather than substitutive forms. Students may feel more isolated when using substitutive forms of technology, and therefore prefer supplemental forms where they have increased access to the instructor and other students. However, technology was found to have a stronger effect on attitudes the longer it was used. Perhaps this is due to a learning curve, where students may spend the first few instances of using a software program simply learning how to operate it, which can be frustrating in some cases. However, once that period is over, students might find the software much more enjoyable and useful. Additionally, research methods courses exhibited significantly larger effect sizes than statistics courses. This result is likely caused by using different types of technology for each type of course. For example, in the current sample, research methods courses most often used tutorials, whereas statistics courses used statistical analysis software programs, web-based instruction, and online learning. All three of these types were found to exhibit significantly smaller effect sizes than other types of technology. Furthermore, the effectiveness of using technology to improve attitudes varied by academic discipline. When compared to education, technology was significantly more effective in mathematics courses, and significantly less effective in nursing courses. Perhaps this is due to the nature of the students who enroll in different academic disciplines, or nature of the instructors who teach in them. Perhaps those involved in the mathematics discipline are much more open to the use of technology than those in other disciplines, especially nursing. Finally, and unsurprisingly, effect sizes were found to be larger in published, rather than unpublished, studies, likely a result of publication bias.

Contributions and Implications

The present meta-analysis provided an update to previous meta-analyses that have been conducted on the use of technology to enhance statistics instruction, such as Christmann and Badgett (1999), and Hsu (2003). Additionally, this meta-analysis included studies of statistics instruction taught fully online, which have not been included in previous meta-analyses. Also, this study included a far greater number of studies than those included in both Christmann and Badgett and Hsu, and also included an examination of attitudinal variables. Furthermore, this study took advantage of a mixed effects methodological approach, which treats the individual studies as a random sample from a population of similar studies, while the others implemented fixed effects approaches, which assume that all variation in effect sizes is explained by known study characteristics (Kalaian et al., 1999).

The present study indicated that the use of technology in statistics courses does have a modest, positive effect on student performance. Specifically, the use of technology was more effective in introductory courses, courses including graduate students, and courses that use technologies such as simulations. A possible reason for the greater effectiveness of the use of technology in introductory courses might be that technology helps illustrate the basic concepts and theories (e.g., probability, central limit theorem, etc.) that underlie statistics courses in a visual medium. However, once these concepts are understood by the learner, the use of technology might provide little additional benefit.

Alternatively, all students in certain academic fields are required to enroll in an introductory statistics course, whereas few are required to enroll in intermediate or advanced courses. As a result, those who already have an aptitude in statistics may enroll in more advanced courses, limiting the size of the effect that is attainable with the addition of technology. In other words, students proficient in statistics will continue to be even in more advanced courses, regardless of whether or not such courses are enhanced with technology. In addition, the use of technology to enhance statistics instruction might be more effective in courses with graduate students because those students might be more accepting of, and familiar with, the use of technology in their courses. Having taken more courses over their academic careers, graduate students are simply more likely to have encountered various uses of technology in their courses than undergraduate students.

However, the use of technology to enhance statistics instruction had little effect on attitudes overall, but did positively affect attitudes under specific circumstances. For example, students preferred uses of technology that were longer in duration, as well as those that were included in research methods rather than statistics courses. Additionally, when compared to students in the education field, students in mathematics had more positive attitudes toward technology, whereas students in nursing had more negative attitudes toward technology use. This could reflect a relationship between gender and attitude toward technology, as students in mathematics tend to be overwhelming male, whereas students in nursing tend to be overwhelmingly female. However, this explanation is not conclusive, as students in education are also overwhelmingly female, and yet tend to view technology use in instruction fairly positively. Also, the effect size was larger for newer studies than for older studies, perhaps indicating a greater acceptance of using technology to enhance instruction over time. Modern students are simply more accustomed to using technology throughout various aspects of their lives, and newer technology is likely to be increasingly user-friendly. However, early uses of technology might have created a great deal of frustration among students.

Limitations

Although the present study made several valuable contributions, several limitations of the study were apparent. First, the coding of the studies was imperfect. Although a high level of inter-rater agreement was obtained, the raters did not obtain perfect agreement on how all study characteristics should be coded. Also, instructional technologies might differ considerably even within the categories in which they were coded here. For example, online learning can vary from simply posting lecture notes and quizzes online asynchronously, to much more interactive synchronous and asynchronous activities, such as online discussions, links to small tutorial programs, games, or interactive presentations, links to additional resources, and so forth. In fact, some have criticized the value of using meta-analysis in technology research on the grounds that individual uses of technology for instruction are so idiosyncratic that they cannot, or should not, be classified into the broad categories necessary for conducting meta-analyses. As stated by Bangert-Drowns (1997), “Some critics argue that meta-analysis, in its effort to be comprehensive, necessarily mixes elements that are too dissimilar to warrant integration” (p. 235). However, although broad categories are often considered for practical purposes, such as sample size, meta-analysts can create categories as broad or narrow as they deem appropriate. Additionally, this is more generally a criticism of the generalizability of quantitative research studies than meta-analysis studies in particular. In other words, if each study were so idiosyncratic that it could not be combined with or compared to other studies covering similar topics in a meta-analysis, then its generalizability would be severely limited for other purposes as well, such as informing practice. Furthermore, other researchers have argued that pooling data from across a large number of quite dissimilar studies is similar to pooling data from across a large number of dissimilar participants, a common and accepted practice in primary quantitative studies (Slavin, 1984).

Second, insufficient information was often provided in the original manuscripts to make adequate, confident, or complete coding decisions. For example, some studies may have referred to the program being studied as a simulation, without providing an adequate description of the program to determine whether or not the program was indeed a simulation according to the definition used in the present study. Other studies simply did not provide some of the relevant information at all. For example, some authors may have provided only group means and not standard deviations, making it impossible to calculate effect sizes. Finally, the coding decisions often had to be made for practical considerations, such as sample size, rather than for logical or theoretical reasons. For instance, several academic disciplines, such as psychology, sociology, and anthropology, were combined into a more global “social science” category simply because few studies emanated from sociology departments, and even fewer from anthropology departments.

A third limitation was that authors used a variety of outcome measures, and one could question whether combining these different measures across studies was appropriate. For example, when authors studied student achievement, the measures varied from individual tests and quizzes, to combinations of tests and quizzes or final course grades. Individual tests and quizzes may target the specific concepts taught in the experiment, whereas combinations of measures and final grades may target a great deal of content, or other variables, outside of that taught specifically in the experiment. This limitation was even more apparent in studies where attitudes were used, as authors included attitudinal measures that varied from overall course satisfaction, satisfaction with instructor, attitude toward technology, attitude toward statistics, attitude toward mathematics, difficulty of course, clarity of presentation of course material, and so forth. Needless to say, one's attitude toward a particular academic subject might vary considerably from that person's opinion of the instructor of that subject.

Fourth, a high likelihood exists that not all relevant studies were found and included in the analysis. Obviously, unpublished studies can be difficult to obtain. In addition, even published studies can be somewhat difficult to find if they are older or published in more obscure journals. However, the use of a mixed effects analysis approach accounts for this, as such an approach assumes that the studies included in the analyses represent a sample from a universe of similar studies that have either been conducted or could potentially be conducted (Raudenbush, 1994).

Finally, even after including several study and methodological variables into the analysis, statistically significant variation in effect sizes remained for both the achievement and the attitude variables. This suggests that not all appropriate variables were identified and taken into account. Additionally, a portion of the remaining variability might be attributed to interaction effects between some of these variables which were not included in the models for practical reasons, including a lack of sufficient sample size and the large number of variables already included in the models. Also, several of the variables that were included in the analysis failed to account for statistically significant variation in effect sizes. This limitation could be related to the imperfection of the coding process, as a great deal of variation within each level of the variables may have remained. For example, a great deal of variance may have remained within each category of technology type related to the manner in which the technology was used or the intervention was implemented. However, the possibility exists that there are subtle differences in the implementation of each intervention that contribute to variation in effect sizes, but defy accountability.

Recommendations and Conclusions

This study provided useful information for statistics instructors who are interested in incorporating greater use of technology into their teaching. However, several questions remain unanswered. First, although differences were found between some technology types in achievement and attitude effect sizes in the present study, this study did not explore the specific characteristics or attributes that could contribute to different levels of effectiveness. Future studies could examine different implementations of technology based on specific characteristics or attributes of that technology rather than more broad types of technology. For example, studies could examine one specific type, such as tutorials, and then examine differences between various characteristics of different tutorial software programs. Additionally, although the present study indicated that online learning was no better than traditional instruction when used for teaching statistics, future studies could examine variations characteristics, attributes, or activities that could be included as part of an online instruction statistics course that could enhance its effectiveness. For example, researchers could compare two different implementations of online learning for statistics instruction, one with weekly synchronous class discussions and one without. Also, future studies could further demarcate various outcome measures. Although all achievement measures were combined in the present study, future studies could examine the outcome measures more specifically, such as short quizzes, mid-term and final examinations, assignments, course points, course grades, and so forth. This would be particularly appropriate for attitudinal measures, as attitudes can vary quite considerably across different targets, such as the course itself, the instructor, technology, the subject matter, and so forth. However, as a result of practical concerns, attitudinal outcome measures were not distinguished here. Furthermore, as mentioned previously, this study did not include analyses of the interaction effects of study characteristics in the model because of the number of variables included relative to the sample size. Future studies could examine the effects of some possible interactions more closely. In addition, multivariate statistical procedures could be used to analyze the data in future studies. For example, multiple dependent variables could be included in the same analysis, such as including both attitudes and achievement variables in one analysis, or different attitude measures could be kept as separate variables but included in the same analysis. This was not conducted in the present meta-analysis because many primary studies did not report both attitude and achievement results, so such an analysis would have greatly reduced the sample size. Also, a multivariate analysis might not be necessary, as studies have generally found small relationships between attitudes and achievement, indicating that the two variables may be reasonably independent (Aleamoni, 1999). Furthermore, this study obtained relatively small mean effect sizes which may not be practically significant. For example, Clark (1983), citing Kulik, Kulik, and Cohen (1979), commented that an effect size of 0.20 amounted to a 1.6 point increase on a 100 point test, which he did not view as practically significant. However, future studies could address how large an effect size is necessary to constitute a practically significant increase in achievement, perhaps by conducting a cost-benefit analysis in conjunction with a meta-analysis. Finally, this study did not specifically examine the relationship between class size and effect size. Perhaps technology is more effective in smaller classes, where students have greater access to technology and the instructor can provide more individualized instruction if students are struggling.

Furthermore, additional variables could be examined that were not included in the present study. For example, several variables related to the type of technology used might account for differences in effectiveness, such as total actual minutes the technology was used rather than duration, the difficulty or learning curve of the technology for instructors to implement and/or students to use, and level of interactivity of the technology. Different technologies allow or require students to interact or submerge themselves in the technology to different degrees. Second, several researcher and methodological variables that may lead to differences in effectiveness include the researcher's skill and experience in implementing experimental interventions and conducting experimental studies more generally, the researcher's knowledge and familiarity with the technology under study, and the overall quality of the study. In the present study, publication status was included as a proxy variable for study quality. However, not all studies published in peer reviewed journals are of equal quality. Perhaps a more systematic rating of study quality would be more appropriate. Third, some instructor variables that might be related to effectiveness include the instructor's knowledge or familiarity with the technology, the amount of experience or training in using and implementing the technology for instructional purposes, and instructor's level of enthusiasm, open-mindedness, or interest in using the technology. Finally, student variables that might relate to effectiveness include the average socioeconomic status within each class or treatment condition, students' average level of experience with using technology, and students' average level of anxiety toward technology use.

Overall, this study assisted in compiling evidence that enhancing statistics instruction with technology can be, under certain conditions, a worthwhile endeavor. Certain uses of technology can be modestly more effective than a lack of technology, and students do enjoy using some types of technology under certain conditions. Additionally, this study indicated that even uses of technology that were not particularly effective were often no worse than not using technology. However, instructors must take great care in introducing technology in a manner that will benefit students most. Students appear to benefit most in statistics courses that are enhanced with technology that allows them to interact with and manipulate data, and does not hinder the instructor's availability and presence to answer questions, clarify concepts, and otherwise provide further assistance when necessary.

APPENDICES

APPENDIX A

META-ANALYSIS DATA CODING SHEET

Authors: ____

Year: ____________

Name of Study:
____

1. Use of technology: ____________________________

2. Supplemental or Substitutive

3. Duration: ____________________________

4. Academic Discipline: ____________________________

5. Course Level: Introductory Intermediate/Advanced

6. Student Academic Standing: Undergraduate Graduate

7. Instructor Bias: Same Instructor Different Instructors

8. Research Design: Experimental Quasi-Experimental

9. Source of Study: ____________________________

10. Outcome measure(s): ____________________________

Data:

	Control	Experimental
Mean		
Standard Deviation		
Sample Size		

t-value:	
F-value:	
Effect size:	

Additional comments: ____
____
____

APPENDIX B

STUDIES INCLUDED IN META-ANALYSIS OF ACHIEVEMENT EFFECT SIZES

Studies Included in the Meta-Analysis of Achievement Effect Sizes

First Author	Year	Technology type	N	d^a	σ_d^2
Aberson	2000	Tutorial	109	-0.247	0.036
Aberson	2002	Tutorial	23	1.409	0.241
Aberson	2003	Tutorial	23	0.786	0.180
Basturk	2005	Statistical analysis software	203	2.922	0.044
		Statistical analysis software	203	1.007	0.025
Benedict	2004	Drill and practice, Web-based	121	0.389	0.033
Bliwise	2005	Tutorial, Web-based	150	0.156	0.027
		Tutorial, Web-based	143	0.027	0.028
		Tutorial, Web-based	147	0.710	0.029
		Tutorial, Web-based	140	0.509	0.029
		Tutorial, Web-based	146	0.508	0.028
		Tutorial, Web-based	139	0.106	0.028
		Tutorial, Web-based	149	-0.184	0.027
		Tutorial, Web-based	142	-0.585	0.029
		Tutorial, Web-based	149	0.631	0.028
		Tutorial, Web-based	143	0.402	0.028
		Tutorial, Web-based	149	1.660	0.036
		Tutorial, Web-based	140	1.330	0.034
Christmann	1997	Statistical analysis software	48	0.184	0.100
Dimitrova	1993	Simulation	66	-0.384	0.061
Dixon	1977	Programmed instruction	109	-0.005	0.036

First Author	Year	Technology type	N	d^a	σ_d^2
		Programmed instruction	109	-0.073	0.036
Dorn	1993	Tutorial	72	0.726	0.058
Erwin	1999	Enhanced lecture	73	1.414	0.067
Frederickson	2005	Online learning	14	0.123	0.251
Fusilier	1985	Simulation, Statistical analysis software	103	0.363	0.039
		Simulation, Statistical analysis software	103	1.385	0.048
Gilligan	1990	Statistical analysis software	76	0.283	0.052
Gonzalez	2000	Programmed instruction	27	1.740	0.193
		Programmed instruction	27	0.229	0.139
		Programmed instruction	26	2.008	0.219
		Programmed instruction	26	0.539	0.148
Gratz	1993	Statistical analysis software	53	-0.036	0.073
		Statistical analysis software	53	-0.480	0.075
		Statistical analysis software	53	0.551	0.076
Harrington	1999	Programmed instruction, Online learning	92	-0.508	0.048
High	1998	Statistical analysis software	85	-0.267	0.046
Hilton	2002	Enhanced lecture	5,597	-0.024	0.001
		Enhanced lecture	5,597	0.002	0.001
		Enhanced lecture	5,597	0.000	0.001
		Enhanced lecture	5,597	-0.020	0.001
Hollowell	1991	Statistical analysis software	131	-0.088	0.032
Hurlburt	2001	Web-based, Online learning	162	0.120	0.038

First Author	Year	Technology type	N	d^a	σ_d^2
Koch	1999	Tutorial, Web-based	24	1.537	0.203
		Tutorial, Web-based	24	0.277	0.156
		Tutorial, Web-based	39	0.918	0.110
Lane	2000	Enhanced lecture	108	0.449	0.037
Lesser	1998	Enhanced lecture, Web-based	106	0.447	0.049
		Enhanced lecture, Web-based	106	-0.206	0.048
		Enhanced lecture, Web-based	106	-0.357	0.049
Marcoulides	1990	Drill and practice, Tutorial, Simulation	82	0.803	0.052
		Tutorial, Simulation, Statistical analysis software	82	1.248	0.057
McLaren	2004	Online learning	205	-0.227	0.020
Mills	2004	Simulation	28	0.921	0.151
		Simulation	26	0.379	0.152
Morris	2001	Simulation, Statistical analysis software	32	0.623	0.124
		Drill and practice	31	-0.466	0.125
Morris	2002	Simulation, Statistical analysis software	32	0.811	0.128
		Drill and practice	32	0.206	0.119
		Simulation	32	0.550	0.123
		Simulation	32	0.459	0.121
Myers	1990	Simulation	50	0.148	0.078
		Simulation	50	0.938	0.087
		Simulation	50	0.551	0.081
		Simulation	50	1.306	0.095

First Author	Year	Technology type	N	d^{a}	σ_d^2
Petta	2000	Web-based	33	0.093	0.133
Porter	1996	Drill and practice	37	0.788	0.111
		Drill and practice	37	1.171	0.121
		Drill and practice	37	0.151	0.103
Raymondo	1998	Statistical analysis software	85	0.143	0.048
		Statistical analysis software	132	-0.141	0.030
Skavaril	1974	Programmed instruction	118	0.169	0.034
Smith	2003	Simulation	41	0.925	0.103
		Simulation	41	1.855	0.135
		Simulation	41	0.862	0.102
		Simulation	41	0.736	0.100
		Simulation	41	1.983	0.141
Song	1992	Simulation, Statistical analysis software	55	-0.119	0.070
		Simulation, Statistical analysis software	55	0.123	0.070
		Simulation, Statistical analysis software	55	-0.010	0.070
		Simulation, Statistical analysis software	55	0.493	0.072
Spinelli	2001	Statistical analysis software	108	0.022	0.039
		Statistical analysis software	108	-0.253	0.039
Sterling	1991	Simulation	64	0.880	0.068
Stockburger	1982	Drill and practice	52	0.000	0.074
		Drill and practice	52	0.917	0.082
		Drill and practice	52	0.832	0.081
		Drill and practice	52	0.000	0.074

First Author	Year	Technology type	N	d^a	σ_d^2
Summers	2005	Online learning	36	0.381	0.108
Tsai	1980	Programmed instruction	94	-0.052	0.061
		Programmed instruction	114	0.610	0.067
		Programmed instruction	16	-0.046	0.222
		Programmed instruction	16	1.573	0.298
		Online learning	88	0.020	0.062
		Online learning	89	0.159	0.069
Utts	2003	Simulation, Online learning	273	-0.090	0.018
		Simulation, Online learning	236	0.027	0.023
Wang	2000	Online learning	113	-0.211	0.036
		Online learning	113	-0.506	0.037
		Online learning	113	0.087	0.036
		Online learning	113	-0.006	0.036
Ware	1989	Statistical analysis software	94	0.113	0.043
		Statistical analysis software	77	0.972	0.066
		Statistical analysis software	94	0.328	0.043
		Statistical analysis software	77	1.208	0.069
Weir	1991	Simulation	37	0.530	0.106
		Simulation	42	0.104	0.091
		Simulation	40	0.702	0.101
		Simulation	37	-0.064	0.103
		Simulation	42	0.601	0.095
		Simulation	42	0.754	0.098

First Author	Year	Technology type	N	d[a]	σ_d^2
Wender	2003	Enhanced lecture	110	0.437	0.037
Wilmouth	1998	Enhanced lecture	231	0.251	0.017
		Enhanced lecture	250	0.367	0.016
		Enhanced lecture	231	0.160	0.017
		Enhanced lecture	243	0.100	0.016
		Enhanced lecture	237	0.229	0.017

[a]The effect size presented in the table is Cohen's d, adjusted for bias.

APPENDIX C

STUDIES INCLUDED IN META-ANALYSIS OF ATTITUDE EFFECT SIZES

Studies Included in Meta-analysis of Attitude Effect Sizes

First author	Year	Technology type	N	d[a]	σ_d^2
Aberson	2000	Tutorial	152	-0.060	0.026
		Tutorial	152	0.261	0.026
		Tutorial	152	-0.080	0.026
		Tutorial	152	0.020	0.026
Aberson	2003	Tutorial	23	2.007	0.253
Bliwise	2005	Tutorial, Web-based	224	0.100	0.021
Brown	2004	Online learning	73	0.916	0.059
Dixon	1977	Programmed instruction	109	0.163	0.036
		Programmed instruction	109	0.775	0.039
Frederickson	2005	Online learning	14	-0.421	0.256
		Online learning	14	0.302	0.253
		Online learning	14	-0.481	0.258
Gilligan	1990	Statistical Analysis Software	76	0.204	0.052
		Statistical Analysis Software	76	0.071	0.052
Gonzalez	2000	Programmed instruction	27	0.135	0.138
		Programmed instruction	26	0.246	0.144
		Programmed instruction	27	-0.227	0.139
		Programmed instruction	26	-0.610	0.150
		Programmed instruction	27	0.765	0.149
		Programmed instruction	26	0.292	0.144
Gratz	1993	Statistical Analysis Software	53	0.266	0.073
		Statistical Analysis Software	53	-0.193	0.073

First author	Year	Technology type	N	d[a]	σ_d^2
		Statistical Analysis Software	53	0.032	0.073
		Statistical Analysis Software	53	0.168	0.073
		Statistical Analysis Software	53	0.184	0.073
Hollowell	1991	Statistical Analysis Software	132	0.466	0.032
		Statistical Analysis Software	132	0.265	0.031
		Statistical Analysis Software	132	0.248	0.031
Hurlburt	2001	Web-based	81	0.409	0.139
		Web-based	81	0.596	0.141
		Web-based	81	1.392	0.150
		Web-based	81	0.268	0.139
		Web-based	81	1.081	0.145
		Web-based	81	0.348	0.139
Koch	1999	Tutorial, Web-based	24	1.185	0.184
		Tutorial, Web-based	24	0.467	0.159
		Tutorial, Web-based	24	0.540	0.161
		Tutorial, Web-based	24	0.139	0.155
Lesser	1998	Enhanced lecture	106	0.248	0.049
McLaren	2004	Online learning	289	-0.947	0.015
Messecar	2003	Programmed instruction	7	0.413	0.462
		Enhanced lecture	9	0.767	0.399
		Online learning	7	1.593	0.628
Mills	2004	Simulation	29	1.619	0.175
Myers	1990	Simulation	50	0.091	0.078

First author	Year	Technology type	N	d[a]	σ_d^2
Petta	2000	Web-based	33	-0.396	0.135
		Web-based	33	-0.382	0.135
		Web-based	33	0.142	0.133
		Web-based	33	0.174	0.133
		Web-based	33	-0.549	0.137
Rendulic	2000	Statistical Analysis Software	85	0.268	0.046
Song	1992	Simulation, Statistical Analysis Software	55	-0.276	0.071
Spinelli	2001	Statistical Analysis Software	108	0.795	0.042
		Statistical Analysis Software	108	1.043	0.044
		Statistical Analysis Software	108	0.956	0.043
		Statistical Analysis Software	108	0.651	0.041
Suanpang	2004	Online learning	233	0.762	0.018
		Online learning	233	0.466	0.018
		Online learning	233	0.599	0.018
		Online learning	233	0.296	0.017
Summers	2005	Online learning	29	-1.084	0.156
		Online learning	31	-0.449	0.127
		Online learning	31	-0.319	0.126
		Online learning	30	-0.670	0.137
		Online learning	30	-0.737	0.138
		Online learning	30	-0.656	0.137
		Online learning	29	-0.473	0.140

First author	Year	Technology type	N	d[a]	σ_d^2
		Online learning	30	-0.111	0.130
		Online learning	28	-1.266	0.172
		Online learning	30	-1.471	0.163
		Online learning	29	-0.977	0.152
		Online learning	26	-1.427	0.194
		Online learning	31	-0.637	0.131
		Online learning	31	-0.342	0.126
		Online learning	29	-0.581	0.142
		Online learning	23	-1.534	0.270
Varnhagen	1990	Programmed instruction	88	-0.464	0.046
		Programmed instruction	88	-0.199	0.045
		Programmed instruction	88	0.257	0.045
		Programmed instruction	88	0.269	0.045
		Programmed instruction	91	0.822	0.047
		Simulation, Programmed instruction	91	0.371	0.044
		Simulation, Programmed instruction	91	0.061	0.043
		Simulation, Programmed instruction	91	0.444	0.044
Wang	2000	Online learning	113	0.000	0.036
Ware	1989	Statistical Analysis Software	94	0.532	0.044
		Statistical Analysis Software	77	0.516	0.062

First author	Year	Technology type	N	d[a]	σ_d^2
Wilmouth	1998	Enhanced lecture	183	0.497	0.022
		Enhanced lecture	182	0.387	0.022
		Enhanced lecture	182	0.477	0.022
		Enhanced lecture	183	0.552	0.023

[a]The effect size presented in the table is Cohen's *d*, adjusted for bias.

APPENDIX D

HLM 6.02 OUTPUT INCLUDING OUTLIER STUDY

Program: HLM 6 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2000
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (6.02.25138.2)
Date: 14 July 2007, Saturday
Time: 11:46:14

SPECIFICATIONS FOR THIS HLM2 RUN

Problem Title: no title

The data source for this run = achieve
The command file for this run = whlmtemp.hlm
Output file name = D:\dissertation files\hlm2.txt
The maximum number of level-1 units = 118
The maximum number of level-2 units = 48
The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood

Weighting Specification

	Weighting?	Weight Variable Name	Normalized?
Level 1	yes	WEIGHT	yes
Level 2	yes	WEIGHT	yes
Precision	yes	VARIANCE	no

The outcome variable is AEFFECT

The model specified for the fixed effects was:

Level-1 Coefficients	Level-2 Predictors
INTRCPT1, B0	INTRCPT2, G00

The model specified for the covariance components was:

Sigma squared (constant across level-2 units)

Tau dimensions

INTRCPT1

Summary of the model specified (in equation format)

Level-1 Model

Y = B0 + R

Level-2 Model
B0 = G00 + U0

Iterations stopped due to small change in likelihood function

******* ITERATION 16 *******

Tau
INTRCPT1,B0 0.04176

Tau (as correlations)
INTRCPT1,B0 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.848

The value of the likelihood function at iteration 16 = -3.201973E+002
The outcome variable is AEFFECT

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	0.059754	0.053670	1.113	47	0.272

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	0.20436	0.04176	47	514.22530	0.000

Statistics for current covariance components model

Deviance = 640.394677
Number of estimated parameters = 1

Program: HLM 6 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2000
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (6.02.25138.2)
Date: 14 July 2007, Saturday
Time: 11:47:50

SPECIFICATIONS FOR THIS HLM2 RUN

Problem Title: no title

The data source for this run = achieve
The command file for this run = whlmtemp.hlm
Output file name = D:\dissertation files\hlm2.txt
The maximum number of level-1 units = 118
The maximum number of level-2 units = 48
The maximum number of iterations = 100
Method of estimation: restricted maximum likelihood

Weighting Specification

	Weighting?	Weight Variable Name	Normalized?
Level 1	yes	WEIGHT	yes
Level 2	yes	WEIGHT	yes
Precision	yes	VARIANCE	no

The outcome variable is AEFFECT

The model specified for the fixed effects was:

```
 Level-1              Level-2
 Coefficients          Predictors
 ----------------------   ---------------
      INTRCPT1, B0     INTRCPT2, G00
                        YEAR, G01
                       DRILL, G02
                       TUTOR, G03
                       SIMUL, G04
                       STAT, G05
                       LECTURE, G06
                       WEB, G07
                       PROGRAM, G08
                       ONLINE, G09
                       FORM, G010
                       DURATION, G011
                       DEPT1, G012
                       DEPT2, G013
                       DEPT4, G014
                       CTYPE1, G015
                       CLEVEL, G016
                       SLEVEL, G017
                       BIAS2, G018
                       DESIGN, G019
                       CONTROL, G020
```

The model specified for the covariance components was:

Sigma squared (constant across level-2 units)

Tau dimensions
INTRCPT1

Summary of the model specified (in equation format)

Level-1 Model

Y = B0 + R

Level-2 Model
B0 = G00 + G01*(YEAR) + G02*(DRILL) + G03*(TUTOR) + G04*(SIMUL)
+ G05*(STAT) + G06*(LECTURE) + G07*(WEB) + G08*(PROGRAM)
+ G09*(ONLINE) + G010*(FORM) + G011*(DURATION) + G012*(DEPT1)
+ G013*(DEPT2) + G014*(DEPT4) + G015*(CTYPE1) + G016*(CLEVEL) + G017*(SLEVEL) + G018*(BIAS2) + G019*(DESIGN) + G020*(CONTROL) + U0

Iterations stopped due to small change in likelihood function

******* ITERATION 1963 *******

Tau
INTRCPT1,B0 0.00000

Tau (as correlations)
INTRCPT1,B0 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.001

The value of the likelihood function at iteration 1963 = -2.560716E+002
The outcome variable is AEFFECT

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	1.220665	16.352760	0.075	27	0.941
YEAR, G01	-0.000128	0.008194	-0.016	27	0.988
DRILL, G02	0.009873	0.143275	0.069	27	0.946
TUTOR, G03	-0.134750	0.183432	-0.735	27	0.469
SIMUL, G04	0.389413	0.103547	3.761	27	0.001
STAT, G05	-0.257949	0.196204	-1.315	27	0.200
LECTURE, G06	-0.233697	0.183244	-1.275	27	0.213
WEB, G07	0.024245	0.119929	0.202	27	0.842
PROGRAM, G08	-0.138140	0.490558	-0.282	27	0.780
ONLINE, G09	-0.678742	0.190133	-3.570	27	0.002
FUNCTION, G010	-0.359269	0.401343	-0.895	27	0.379
DURATION, G011	0.023652	0.108476	0.218	27	0.829
DEPT1, G012	-0.030165	0.154008	-0.196	27	0.846
DEPT2, G013	-0.131068	0.181628	-0.722	27	0.477
DEPT4, G014	-0.277483	0.164935	-1.682	27	0.104
CTYPE1, G015	-0.195217	0.178706	-1.092	27	0.285
CLEVEL, G016	-0.836147	0.216831	-3.856	27	0.001
SLEVEL, G017	0.789236	0.272155	2.900	27	0.008
BIAS2, G018	-0.006331	0.134400	-0.047	27	0.963
DESIGN, G019	0.062530	0.083151	0.752	27	0.459
CONTROL, G020	-0.315913	0.211287	-1.495	27	0.146

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	0.00040	0.00000	27	178.12548	0.000

Statistics for current covariance components model

Deviance = 512.143160
Number of estimated parameters = 1

REFERENCES

REFERENCES

References marked with an asterisk indicate those included in the meta-analyses.

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

Aberson, C. L., Berger, D., Healy, M. R., & Romero, V. L. (2001). Teaching statistics with web technology: The WISE project. *Syllabus, 14*(1), 43-45.

*Aberson, C. L., Berger, D. E., Healy, M. R., & Romero, V. L. (2002). An interactive tutorial for teaching statistical power. *Journal of Statistics Education, 10*(3). Retrieved February 28, 2006, from http://www.amstat.org/publications/jse/v10n3/aberson.html

*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.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 12*(2), 153-166.

Allen, M., Bourhis, J., Burrell, N., & Mabry, E. (2002). Comparing student satisfaction with distance education to traditional classrooms in higher education: A meta-analysis. *The American Journal of Distance Education, 16*(2), 83-97.

Allen, M., Mabry, E., Mattrey, M., Bourhis, J., Titsworth, S., & Burrell, N. (2004). Evaluating the effectiveness of distance learning: A comparison using meta-analysis. *Journal of Communication, 54*(3), 402-420.

Balkovich, E., Lerman, S., & Parmelee, P. (1985). Computing in higher education: The Athena experience. *Communications of the ACM, 28*(11), 1214-1224.

Bangert-Drowns, R. L. (1997). Some limiting factors in meta-analysis. In W. J. Bukoski (Ed.), *Meta-analysis of drug abuse prevention programs* (Monograph 170, pp. 234-252). Washington, DC: National Institute of Drug Abuse Research.

*Basturk, R. (2005). The effectiveness of computer-assisted instruction in teaching introductory statistics. *Educational Technology & Society, 8*(2), 170-178.

Bayraktar, S. (2002). A meta-analysis of the effectiveness of computer-assisted instruction in science education. *Journal of Research on Technology in Education, 34*(2), 173-188.

Beard, L. A., Harper, C., & Riley, G. (2004). Online versus on-campus instruction: Student attitudes & perceptions. *TechTrends, 48*(6), 29-31.

Becker, B. J. (1996). A look at the literature (and other resources) on teaching statistics. *Journal of Educational and Behavioral Statistics, 21*(1), 71-90.

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1&2), 127-155.

*Benedict, J. O., & Anderton, J. B. (2004). Applying the just-in-time teaching approach to teaching statistics. *Teaching of Psychology, 31*(3), 197-199.

Benjamin, L. T., Jr. (1988). A history of teaching machines. *American Psychologist, 43*(9), 703-712.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., et al. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379-439.

Bertie, A., & Farrington, P. (2003). Teaching confidence intervals with Java applets. *Teaching Statistics, 25*(3), 70-74.

*Bliwise, N. G. (2005). Web-based tutorials for teaching introductory statistics. *Journal of Educational Computing Research, 33*(3), 309-325.

Botstein, L. (2001). High-tech's "Brave New World" braggadocio. *The Education Digest, 66*(9), 11-16.

Brett, P., & Nagra, J. (2005). An investigation into students' use of a computer-based social learning space: Lessons for facilitating collaborative learning. *British Journal of Educational Technology, 36*(2), 281-292.

Brooks, G. P. (2003). Using Monte Carlo methods to teach statistics: The MC2G computer program. *Understanding Statistics, 2*(2), 137-150.

*Brown, S. W., & Kulikowich, J. M. (2004). Teaching statistics from a distance: What have we learned? *International Journal of Instructional Media, 31*(1), 19-35.

Buzhardt, J., & Semb, G. B. (2002). Item-by-item versus end-of-test feedback in a computer-based PSI course. *Journal of Behavioral Education, 11*(2), 89-104.

Chambers, J. A., & Sprecher, J. W. (1980). Computer assisted instruction: Current trends and critical issues. *Communications of the ACM, 23*(6), 332-342.

Christmann, E., & Badgett, J. (1999). The comparative effectiveness of various microcomputer-based software packages on statistical achievement. *Computers in Schools, 16*(1), 209-220.

*Christmann, E., Badgett, J., & Lucking, R. (1997a). Microcomputer-based computer-assisted instruction within differing subject areas: A statistical deduction. *Journal of Educational Computing Research, 16*(3), 281-296.

Christmann, E., Badgett, J., & Lucking, R. (1997b). Progressive comparison of the effects of computer-assisted instruction on the academic achievement of secondary students. *Journal of Research on Computing in Education, 29,* 325-337.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-459.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis.* New York: Sage.

Cortina, J. M., & Nouri, H. (2000). *Effect size for ANOVA designs.* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-129). Thousand Oaks, CA: Sage.

Dallal, G. E. (1990). Statistical computer packages: Dare we abandon their teaching to others? *The American Statistician, 44*(4), 265-266.

DeBord, K. A., Aruguete, M. S., & Muhlig, J. (2004). Are computer-assisted teaching methods effective? *Teaching of Psychology, 31*(1), 65-68.

delMas, R. C., Garfield, J., & Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education, 7*(3). Retrieved March 2, 2006, from http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm

*Dimitrova, G., Persell, C. H., & Maisel, R. (1993). Using and evaluating ISEE, a new computer program for teaching sampling and statistical inference. *Teaching Sociology, 21,* 341-351.

*Dixon, P. N., & Judd, W. A. (1977). A comparison of computer-managed instruction and lecture mode for teaching basic statistics. *Journal of Computer-Based Instruction, 4*(1), 22-25.

*Dorn, M. J. (1993). The effect of an interactive, problem-based hypercard modular instruction on statistical reasoning (Doctoral dissertation, Southern Illinois University at Carbondale, 1994). *Dissertation Abstracts International, 55,* 09-A.

Duhaney, D. C. (2005). Technology and higher education: Challenges in the halls of academe. *International Journal of Instructional Media, 32*(1), 7-15.

Eisenberg, Y. (1986). The effects of computer-based instruction on college student's interest and achievement. *Educational Technology, 26*(4), 40-43.

*Erwin, T. D., & Rieppi, R. (1999). Comparing multimedia and traditional approaches in undergraduate psychology classes. *Teaching of Psychology, 26*(1), 58-61.

Feenberg, A. (2001). Whither educational technology? *International Journal of Technology and Design Education, 11*, 83-91.

Fernandez, G. C. J., & Liu, L. (1999). A technology-based teaching model that stimulates statistics learning. *Computers in the Schools, 16*(1), 173-191.

Fitch, J. L. (2004). Student feedback in the college classroom: A technology solution. *Educational Technology Research & Development, 52*(1), 71-81.

Forsyth, D. R., & Archer, C. R. (1997). Technologically assisted instruction and student mastery, motivation, and matriculation. *Teaching of Psychology, 24*(3), 207-212.

Forte, J. (1995). Teaching statistics without sadistics. *Journal of Social Work Education, 31*(2), 204-218.

*Frederickson, N., Reed, P., & Clifford, V. (2005). Evaluating web-supported learning versus lecture-based teaching: Quantitative and qualitative perspectives. *Higher Education, 50,* 645-664.

Frith, V., Jaftha, J., & Prince, R. (2004). Evaluating the effectiveness of interactive computer tutorials for an undergraduate mathematical literacy course. *British Journal of Educational Technology, 35*(2), 159-171.

*Fusilier, M. R., & Kelly, E. P. (1985) An evaluation of computer-assisted instruction for learning business statistics. *AEDS Journal, 18*, 237-242.

Garland, K., & Noyes, J. (2004). The effects of mandatory and optional use on students' ratings of a computer-based learning package. *British Journal of Educational Technology, 35*(3), 263-273.

Garson, G. D. (1999). The role of technology in quality education. Retrieved July 13, 2006, from http://hcl.chass.ncsu.edu/sscore/garson2.htm

Geoghegan, W. (1994) *Whatever happened to instructional technology?* Paper presented at the 22nd Annual Conference of the International Business Schools Computing Association, Baltimore MD, July 17-20.

*Gilligan, W. P. (1990). The use of a computer statistical package in teaching a unit of descriptive statistics. (Doctoral dissertation, Boston University, 1990). *Dissertation Abstracts International, 51,* 07-A.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5,* 351-379.

*Gonzalez, G. M., & Birch, M. A. (2000). Evaluating the instructional efficacy of computer-mediated interactive multimedia: Comparing three elementary statistics tutorial modules. *Journal of Educational Computing Research, 22*(4), 411-436.

Goolkasian, P. (1985). A microcomputer-based lab for psychology instruction. *Teaching of Psychology, 12*(4), 223-225.

Goolkasian, P., & Lee, J. A. (1988). A computerized laboratory for general psychology. *Teaching of Psychology, 15*(2), 98-100.

Grant, L. K., & Spencer, R. E. (2003). The personalized system of instruction: Review and applications to distance education. *International Review of Research in Open and Distance Learning, 4*(2). Retrieved September 11, 2006, from http://www.irrodl.org/index.php/irrodl/article/view/152/233

*Gratz, Z. S., Volpe, G. D., & Kind, B. M. (1993, March). *Attitudes and achievement in introductory psychological statistics classes versus computer supported instruction.* Paper presented at the annual conference of the Teachers of Psychology, Ellenville, NY.

Gretes, J. A., & Green, M. (2000). Improving undergraduate learning with computer-assisted assessment. *Journal of Research on Computing in Education, 33*(1), 46-54.

Grineski, S. (1999). Questioning the role of technology in higher education: Why is this the road less traveled? *The Internet and Higher Education, 2*(1), 45-54.

Groves, M. M., & Zemel, P. C. (2000). Instructional technology adoption in higher education: An action research case study. *International Journal of Instructional Media, 27*(1), 57-65.

Hacker, R., & Sova, B. (1998). Initial teacher education: A study of the efficacy of computer mediated courseware delivery in a partnership context. *British Journal of Educational Technology, 29*(4), 333-341.

*Harrington, D. (1999). Teaching statistics: A comparison of traditional classroom and programmed instruction/distance learning approaches. *Journal of Social Work Education, 35*(3), 343-352.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.

*High, R. V. (1998). *Some variables in relation to students' choice of statistics classes: Traditional versus computer-supported instruction.* Garden City, NY: Adelphi University. (ERIC Document Reproduction Service No. ED427762)

*Hilton, S. C., & Christensen, H. B. (2002). Evaluating the impact of multimedia lectures on student learning and attitudes. In B. Phillips (Ed.), *Developing a statistically literate society. Proceedings of the Sixth International Conference on Teaching of Statistics* [CD-ROM]. Cape Town: International Statistical Institute.

*Hollowell, K. A., & Duch, B. J. (1991). *Functions and statistics with computers at the college level*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

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

Hunt, N., & Tyrrell, S. (2000). Learning statistics on the web-DISCUSS. *Teaching Statistics, 22*(3), 85-90.

*Hurlburt, R. T. (2001). "Lectlets" deliver content at a distance: Introductory statistics as a case study. *Teaching of Psychology, 28*(1), 15-20.

Hyden, P. (2005). Teaching statistics by taking advantage of the laptop's ubiquity. *New Directions for Teaching and Learning, 101*, 37-42.

Jacobsen, D. M. (1998, June). *Adoption patterns of faculty who integrate computer technology for teaching and learning in higher education.* Proceedings of the ED-MEDIA AND ED-TELECOM 98: World Conference on Educational Multimedia and Hypermedia & World Conference on Educational Telecommunications, Freiburg, Germany.

Johnson, H. D., & Dasgupta, N. (2005). Traditional versus non-traditional teaching: Perspectives of students in introductory statistics classes. *Journal of Statistics Education, 13*(2). Retrieved June 16, 2006, from http://www.amstat.org/publications/jse/v13n2/johnson.html

Kalaian, H. A., Mullan, P. B., & Kasim, R. M. (1999). What can studies of problem-based learning tell us? Synthesizing and modeling PBL effects on National Board of Medical Examination performance: Hierarchical linear modeling meta-analytic approach. *Advances in Health Sciences Education, 4,* 209-221.

Kalaian, S. A. (2003). Meta-analysis methods for synthesizing treatment effects in multisite studies: Hierarchical linear modeling (HLM) perspective. *Practical Assessment, Research & Evaluation, 8*(15). Retrieved November 22, 2005, from http://PAREonline.net/getvn.asp?v=8&n=15

Keller, F. S. (1968). "Goodbye, Teacher...". *Journal of Applied Behavior Analysis, 1*(1), 79-89.

Khalili, A., & Shashaani, L. (1994). The effectiveness of computer applications: A meta-analysis. *Journal of Research on Computing in Education, 27*(1), 48-62.

Khamis, H. J. (1991). Manual computations: A tool for reinforcing concepts and techniques. *The American Statistician, 45*(4), 294-299.

*Koch, C., & Gobell, J. (1999). A hypertext-based tutorial with links to the Web for teaching statistics and research methods. *Behavior Research Methods, Instruments, & Computers, 31*(1), 7-13.

Kozma, R. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42*(2), 7-19.

Kulik, C. C., & Kulik, J. A. (1986). Effectiveness of computer-based education in colleges. *AEDS Journal, 19*(1-2), 81-108.

Kulik, C. C., Kulik, J. A., & Cohen, P. A. (1980). Instructional technology and college teaching. *Teaching of Psychology, 7*(4), 199-205.

Kulik, C. C., Kulik, J. A., & Shwalb, B. J. (1986). The effectiveness of computer-based adult education: A meta-analysis. *Journal of Educational Computing Research, 2*(2), 235-252.

Kulik, J. A. (1983). Synthesis of research on computer-based instruction. *Educational Leadership, 41*(1), 19-21.

Kulik, J. A., Cohen, P. A., & Ebeling, B. J. (1980). Effectiveness of programmed instruction in higher education: A meta-analysis of findings. *Educational Evaluation and Policy Analysis, 2*(6), 51-64.

Kulik, J. A., & Kulik, C. C. (1987). Review of recent research literature on computer-based instruction. *Contemporary Educational Psychology, 12*, 222-230.

Kulik, J. A., & Kulik, C. C. (1989). Effectiveness of computer-based instruction. *School Library Media Quarterly, 17*(3), 156-159.

Kulik, J. A., Kulik, C. C., & Bangert-Drowns, R. (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior, 1*, 59-74.

Kulik, J. A., Kulik, C. C., & Carmichael, K. (1974). The Keller plan in science teaching. *Science, 183*, 379-383.

Kulik, J. A., Kulik, C. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist, 34*(4), 307-318.

Kulik, J. A., Kulik, C. C., & Cohen, P. A. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. *Review of Educational Research, 50*(4), 525-544.

*Lane, D. M., & Tang, Z. (2000). Effectiveness of simulation training on transfer of statistical concepts. *Journal of Educational Computing Research, 22*(4), 383-396.

Lau, L. K. (2003). Institutional factors affecting student retention. *Education, 124*(1), 126-136.

Lawson, T. J. (2000). Teaching a social psychology course on the web. *Teaching of Psychology, 27*(4), 285-289.

Lee, J. (1999). Effectiveness of computer-based instructional simulation: A meta analysis. *International Journal of Instructional Media, 26*(1), 71-85.

*Lesser, L. M. (1998). Technology-rich standards-based statistics: Improving introductory statistics at the college level. *Technological Horizons in Education Journal, 25*(7), 54-57.

Levine, A. (2001). The remaking of the American university. *Innovative Higher Education, 25*(4), 253-267.

Lewis, D. R., Dalgaard, B. R., & Boyer, C. M. (1985). Cost effectiveness of computer-assisted economics instruction. *The American Economic Review, 75*(2), 91-96.

Liao, Y. C. (1998). Effects of hypermedia versus traditional instruction on students' achievement: A meta-analysis. *Journal of Research on Computing in Education, 30*(4), 341-359.

Liao, Y. C. (1999). Effects of hypermedia on students' achievement: A meta-analysis. *Journal of Educational Multimedia and Hypermedia, 8*(3), 255-277.

Lloyd, D., & Martin, J. G. (1996). The introduction of computer-based testing on an engineering technology course. *Assessment & Evaluation in Higher Education, 21*(1), 83-90.

Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician, 54*(3), 1-11.

Luyben, P. D., Hipworth, K., & Pappas, T. (2003). Effects of CAI on the academic performance and attitudes of college students. *Teaching of Psychology, 30*(2), 154-158.

Lynch, T. G., Steele, D. J., Palensky, J. E. J., Lacy, N. L., & Duffy, S. W. (2001). Learning preferences, computer attitudes, and test performance with computer-aided instruction. *The American Journal of Surgery, 181*, 368-371.

Malec, M. A. (1982). A PSI statistics course. *Teaching Sociology, 10*(1), 84-87.

Maltby, J. (2001). Learning statistics by computer software is cheating. *Journal of Computer Assisted Learning, 17*, 329-330.

Marasinghe, M. G., Meeker, W. Q., Cook, D., & Shin, T. (1996). Using graphics and simulation to teach statistical concepts. *The American Statistician, 50*(4), 342-351.

*Marcoulides, G. A. (1990). Improving learner performance with computer based programs. *Journal of Educational Computing Research, 6*(2), 147-155.

Massy, W. F., & Wilger, A. K. (1998). Technology's contribution to higher education productivity. *New Directions for Higher Education, 103*, 49-59.

Mausner, B., Wolff, E. F., Evans, R. W., DeBoer, M. M., Gulkus, S. P., D'Amore, A., et al. (1983). A program of computer assisted instruction for a personalized instructional course in statistics. *Teaching of Psychology, 10*(4), 195-200.

*McLaren, C. H. (2004). A comparison of student persistence and performance in online and classroom business statistics experiences. *Decision Sciences Journal of Innovative Education, 2*(1), 1-10.

McOmber, J. B. (1999). Technological autonomy and three definitions of technology. *Journal of Communication, 49*(3), 137-153.

*Messecar, D. C. (2003). Reading statistics in nursing research: A self-study CD-ROM module. *Journal of Nursing Education, 42*(5), 220-226.

Michael, K. Y. (2001). The effect of a computer simulation activity versus a hands-on activity on product creativity in technology education. *Journal of Technology Education, 13*(1), 31-43.

Miller, J. W., Martineau, L. P., & Clark, R. C. (2000). Technology infusion and higher education: Changing teaching and learning. *Innovative Higher Education, 24*(3), 227-241.

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1). Retrieved March 1, 2006, from http://www.amstat.org/publications/jse/v10n1/mills.html

Mills, J. D. (2003). A theoretical framework for teaching statistics. *Teaching Statistics, 25*(2), 56-58.

*Mills, J. D. (2004). Learning abstract statistics concepts using simulation. *Educational Research Quarterly, 28*(4), 18-33.

Mitra, A., & Steffensmeier, T. (2000). Changes in student attitudes and student computer use in a computer-enriched environment. *Journal of Research on Computing in Education, 32*(3), 417-433.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*, 123-165.

*Morris, E. (2001). The design and evaluation of Link: A computer-based learning system for correlation. *British Journal of Educational Technology, 32*(1), 39-52.

*Morris, E. J., Joiner, R., & Scanlon, E. (2002). The contribution of computer-based activities to understanding statistics. *Journal of Computer Assisted Learning, 18*, 114-124.

*Myers, K. N. (1990). An exploratory study of the effectiveness of computer graphics and simulations in a computer-student interactive environment in illustrating random sampling and the central limit theorem (Doctoral dissertation, Florida State University, 1990). *Dissertation Abstracts International, 51,* 441.

Niemiec, R. P., & Walberg, H. J. (1985). Computers and achievement in the elementary schools. *Journal of Educational Computing Research, 1*(4), 435-440.

Olson, T. M., & Wisher, R. A. (2002). The effectiveness of web-based instruction: An initial inquiry. *The International Review of Research in Open and Distance Learning, 3*(2). Retrieved September 11, 2005, from http://www.irrodl.org/index.php/irrodl/article/viewArticle/103/182

Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research & Evaluation*, 7(1). Retrieved November 22, 2005, from http://PAREonline.net/getvn.asp?v=7&n=1

Paden, D. W., Dalgaard, B. R., & Barr, M. D. (1977). A decade of computer-assisted instruction. *The Journal of Economic Education, 9*(1), 14-20.

Pear, J. J., & Crone-Todd, D. E. (1999). Personalized system of instruction in cyberspace. *Journal of Applied Behavior Analysis, 32*(2), 205-209.

Pear, J. J., & Novak, M. (1996). Computer-aided personalized system of instruction: A program evaluation. *Teaching of Psychology, 23*(2), 119-123.

Pearson, F. S., & Lipton, D. S. (1999). A meta-analytic review of the effectiveness of the corrections-based treatments for drug abuse. *The Prison Journal, 79*(4), 384-410.

Pearson, G., & Young, A. T. (Eds.). (2002). *Technically speaking: Why all Americans need to know more about technology.* Washington, DC: National Academy Press.

Peluchette, J. V., & Rust, K. A. (2005). Technology use in the classroom: Preferences of management faculty members. *Journal of Education for Business, 80*(4), 200-205.

*Petta, N. A. (2000). The impact of a web-based class management system on student performance and attitudes in a quantitative statistics class. (Doctoral dissertation, University of Nebraska, 1999). *Dissertation Abstracts International, 60,* 3889.

*Porter, T. S., & Riley, T. M. (1996). The effectiveness of computer exercises in introductory statistics. *Journal of Economic Education, 27*(4), 291-299.

Rainbow, S. W., & Sadler-Smith, E. (2003). Attitudes to computer-assisted learning amongst business and management students. *British Journal of Educational Technology, 34*(5), 615-624.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321). New York: Russell Sage Foundation.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2005). *HLM 6: Hierarchical linear and nonlinear modeling.* Chicago: Scientific Software International.

Raymondo, J. C. (1996). Developing a computer laboratory for undergraduate sociology courses. *Teaching Sociology, 24*, 305-309.

*Raymondo, J. C., & Garrett, J. R. (1998). Assessing the introduction of a computer laboratory experience into a behavioral science statistics course. *Teaching Sociology, 26*, 29-37.

Reiser, R. A. (2001). A history of instructional technology: Part II: A history of instructional design. *Educational Technology Research & Development, 49*(2), 57-67.

*Rendulic, P., & Terrell, S. (2000). Anxiety toward statistics and the use of computers: A study of graduate level education majors. *Journal of Computing in Higher Education, 11*(2), 104-120.

Rickman, J., & Grudzinski, M. (2000). Student expectations of information technology use in the classroom. *Educause Quarterly, 1*, 24-30.

Rogers, D. L. (2000). A paradigm shift: Technology integration for higher education in the new millennium. *Educational Technology Review, 13*, 19-27, 33.

Rogers, R. L. (1987). A microcomputer-based statistics course with individualized assignments. *Teaching of Psychology, 14*(2), 109-111.

Romero, V. L., Berger, D. E., Healy, M. R., & Aberson, C. L. (2000). Using cognitive learning theory to design effective on-line statistics tutorials. *Behavior Research Methods, Instruments, & Computers, 32*(2), 246-249.

Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Rowell, G. H. (2004, August). *Assessment of using technology for teaching statistics*. Paper presented at the ARTIST Roundtable Conference on Assessment in Statistics, Appleton, WI.

Sanders, M. (1996). Teaching statistics with computer networks. *Mathematics Teacher, 89*, 70-72.

Schimmack, U., Diener, E., & Oishi, S. (2002). Life-satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources. *Journal of Personality, 70*(3), 345-384.

Searle, S. R. (1989). Statistical computing packages: Some words of caution. *The American Statistician, 43*(4), 189-190.

Shuell, T. J., & Farber, S. L. (2001). Students' perceptions of technology use in college courses. *Journal of Educational Computing Research, 24*(2), 119-131.

*Skavaril. R. V. (1974). Computer-based instruction of introductory statistics. *Journal of Computer-Based Instruction*, *1*(1), 32-40.

Slavin, R. E. (1984). Meta-analysis in education: How has it been used? *Educational Researcher, 13*(8), 6-15.

Sliwinski, M. J., & Hall, C. B. (1998). Constraints on general slowing: A meta-analysis using hierarchical linear models with random coefficients. *Psychology and Aging, 13*(1), 164-175.

*Smith, B. (2003). Using and evaluating resampling simulations in SPSS and Excel. *Teaching Sociology, 31,* 276-287.

Smith, S. G., & Sherwood, B. A. (1976). Educational uses of the PLATO computer system. *Science, 192*, 344-352.

*Song, P. W. (1992). *The effect of using computer software and computer simulations to teach statistics and probability in a Korean college.* (Doctoral dissertation, The University of Iowa, 1992). Dissertation Abstracts International, 53, 07-A.

*Spinelli, M. A. (2001). The use of technology in teaching business statistics. *Journal of Education for Business, 77,* 41-44.

SPSS, Inc. (2005). *SPSS 14.0 for Windows.* Chicago: SPSS, Inc.

*Sterling, J., & Gray, M. W. (1991). The effect of simulation software on students' attitudes and understanding in introductory statistics. *Journal of Computers in Mathematics and Science Teaching, 10*(4), 51-56.

*Stockburger, D. W. (1982). Evaluation of three simulation exercises in an introductory statistics course. *Contemporary Educational Psychology, 7*, 365-370.

Stork, D. (2003). Teaching statistics with student survey data: A pedagogical innovation in support of student learning. *Journal of Education for Business, 78*(6), 335-339.

Strasser, S. E., & Ozgur, C. (1995). Undergraduate business statistics: A survey of topics and teaching methods. *Interfaces, 25*(3), 95-103.

*Suanpang, P., Petocz, P., & Kalceff, W. (2004). Student attitudes to learning business statistics: Comparison of online and traditional methods. *Educational Technology & Society, 7*(3), 9-20.

*Summers, J. J., Waigandt, A., & Whittaker, T. A. (2005). A comparison of student achievement and satisfaction in an online versus a traditional face-to-face statistics class. *Innovative Higher Education, 29*(3), 233-250.

Suppes, P., & Morningstar, M. (1969). Computer-assisted instruction: Two computer-assisted instruction programs are evaluated. *Science, 166*, 343-350.

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

Tengs, T. O., & Lin, T. H. (2003). A meta-analysis of quality-of-life estimates for stroke. *Pharmacoeconomics, 21*(3), 191-200.

Thelwall, M. (2001). Understanding and assessment methodology in an introductory statistics course. *Journal of Computers in Mathematics and Science Teaching, 20*(3), 251-263.

Thoennessen, M., & Harrison, M. J. (1996). Computer-assisted assignments in a large physics class. *Computers & Education, 27*(2), 141-147.

Timmerman, C. E., & Kruepke, K. A. (2006). Computer-assisted instruction, media richness, and college student performance. *Communication Education, 55*(1), 73-104.

Tromater, L. J. (1985). Teaching a course in computer-assisted statistical analysis. *Teaching of Psychology, 12*(4), 225-226.

*Tsai, S. W., & Pohl, N. F. (1980). Computer-assisted instruction augmented with planned teacher/student contacts. *Journal of Experimental Education, 49*, 120-126.

*Utts, J., Sommer, B., Acredelo, C., Maher, M. W., & Matthews, H. R. (2003). A study comparing traditional and hybrid internet-based instruction in introductory statistics classes. *Journal of Statistics Education, 11*(3). Retrieved September 9, 2006 from http://www.amstat.org/publications/jse/v11n3/utts.html

Van Dusen, G. C. (1998). Technology: Higher education's magic bullet. *Thought & Action, 14*(1), 59-67.

Varnhagen, C. K., Drake, S. M., & Finley, G. (1997). Teaching statistics with the Internet. *Teaching of Psychology 24*(4), 275-278.

*Varnhagen, C. K., & Zumbo, B. D. (1990). CAI as an adjunct to teaching introductory statistics: Affect mediates learning. *Journal of Educational Computing Research, 6*(1), 29-40.

Velleman, P. F., & Moore, D. S. (1996). Multimedia for teaching statistics: Promises and pitfalls. *The American Statistician, 50*(3), 217-225.

Vodanovich, S. J., & Piotrowski, C. (2001). Internet-based instruction: A national survey of psychology faculty. *Journal of Instructional Psychology, 28*(4), 253-255.

Wagner, G. R., & Motazed, B. (1972). The proctorial system of instruction combined with computer pedagogy for teaching applied statistics. *The American Statistician, 26*(3), 36-39.

*Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology web-based classes. *Journal of Educational Psychology, 92*(1), 137-143.

*Ware, M. E., & Chastain, J. D. (1989). Computer-assisted statistical analysis: A teaching innovation? *Teaching of Psychology, 16*(4), 222-227.

Warner, C. B., & Meehan, A. M. (2001). Microsoft Excel™ as a tool for teaching basic statistics. *Teaching of Psychology, 28*(4), 295-298.

Warren, L. L., & Holloman, Jr., H. L. (2005). On-line instruction: Are the outcomes the same? *Journal of Instructional Technology, 32*(2), 148-151.

Waxman, H. C., Lin, M., & Michko (2003). *A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes.* Naperville, IL: Learning Point Associates. Retrieved January 15, 2005, from http://www.ncrel.org/tech/effects2/

*Weir, C. G., McManus, I. C., & Kiely, B. (1991). Evaluation of the teaching of statistical concepts by interactive experience with Monte Carlo simulations. *British Journal of Educational Psychology, 61*, 240-247.

*Wender, K., & Muehlboeck, J. (2003). Animated diagrams in teaching statistics. *Behavior Research Methods, Instruments, & Computers, 35*(2), 255-258.

Welsh, J. A., & Null, C. H. (1991). The effects of computer-based instruction on college students' comprehension of classic research. *Behavior Research Methods, Instruments, & Computers, 23*(2), 301-305.

Weston, T. J. (2005). Why faculty did-and did not-integrate instructional technology software in their undergraduate classrooms. *Innovative Higher Education, 30*(2), 99-115.

*Wilmouth, J., & Wybraniec, J. (1998). Profits and pitfalls: Thoughts on using a laptop computer and presentation software to teach introductory social statistics. *Teaching Sociology, 26*(3), 166-178.

Wilson, S. P., & Harris, A. (2002). Evaluation of The Psychology Place: A web-based instructional tool for psychology courses. *Teaching of Psychology, 29*(2), 165-168.

Wit, E. (2003). Who wants to be... The use of a personal response system in statistics teaching. *MSOR Connections, 3*(2), 14-20.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-04). Newbury Park, CA: Sage.

Worthington, E. L., Jr., Welsh, J. A., Archer, C. R., Mindes, E. J., & Forsyth, D. R. (1996). Computer-assisted instruction as a supplement to lectures in an introductory psychology class. *Teaching of Psychology, 23*(3), 175-181.

Yilmaz, M. R. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistics Education, 4*(1) Retrieved March 1, 2006, from http://www.amstat.org/publications/jse/v4n1/yilmaz.html