JRTE | Vol. 43, No. 3, pp. 253–278 | ©2011 ISTE | iste.org

A Meta-Analysis Examining the Impact of Computer-Assisted Instruction on Postsecondary Statistics Education: 40 Years of Research

Karen Larwin
Youngstown State University

David Larwin
Kent State University at Salem

Abstract

The present meta-analysis is a comprehensive investigation of the effectiveness of computer-assisted instruction (CAI) on student achievement in postsecondary statistics education across a forty year period of time. The researchers calculated an overall effect size of 0.566 from 70 studies, for a total of 219 effect-size measures from a sample of n = 40,125 participants. These results suggest that the typical student moved from the 50th percentile to the 73rd percentile when technology was used as part of the curriculum. This study demonstrates that subcategories can further the understanding of how the use of CAI in statistics education might be maximized. The study discusses implications and limitations. (Keywords: statistics education, computer-assisted instruction, meta-analysis)

Discovering how students learn most effectively is one of the major goals of research in education. During the last 30 years, many researchers and educators have called for reform in the area of statistics education in an effort to more successfully reach the growing population of students, across an expansive variety of disciplines, who are required to complete coursework in statistics (e.g., Cobb, 1993, 2007; Garfield, 1993, 1995, 2002; Giraud, 1997; Hogg, 1991; Lindsay, Kettering, & Siegmund, 2004; Moore, 1997; Roiter, & Petocz, 1996; Snee, 1993;Yilmaz, 1996). These new populations of inexperienced statisticians are not mathematicians like the statistics apprentices from years past. Many of these students have very little interest in learning mathematics and even less interest in learning statistics. In light of this, reform efforts have proposed that statistics education should abandon the "…information transfer model in favor of a constructivist approach to learning…" (Moore, 1997, p. 124) in an effort to help students develop an understanding of statistical concepts beyond the use of mathematical formulas.

The use of technology in the statistics classroom has coincided with other reform efforts in statistics education. Students can now perform once laborious calculations instantaneously with statistical programs in and out of the classroom. More classroom time can be dedicated to exploring and understanding the underlying concepts behind statistics, as the mathematical mechanics of statistics have been reduced to inputting data so that a program can quickly perform the calculations. But has this progress in technology benefited student performance, or is it a double-edged sword?

Reform leaders suggest that traditional approaches to statistics education failed because students were not equipped to build conceptual understandings of the core concepts. Too much of the focus in these traditional approaches to statistics education is spent on grueling calculations and the difficult concepts of probability theory. At the same time, learning to successfully use statistical packages will not transform anyone into a statistician. Tools such as SPSS or Minitab 15.0 can help the student solve problems more quickly, therefore saving them from some of the discouragement of working with difficult statistical formulas, but these technological tools do not necessarily help the student to develop an understanding of statistical concepts. It is important to understand how the technology and statistical packages that reflect the nature of statistics in the workplace, research laboratory, and classroom can help students acquire the necessary conceptual understandings that make up the science of statistics.

Statement of the Problem

Over the last four decades, computer-assisted instruction (CAI) has become more prevalent in postsecondary educational systems. The use of technology has become commonplace in statistics education. Bratz and Sabikuj (2001) reported that when surveyed in 1982, 50% of universities responded that they were using CAI in introductory-level statistics courses in their respective psychology departments. More than two decades later, the number of universities reporting the use of CAI in teaching introductory statistics had grown to 80% (Bratz & Sabikuj 2001; Lindsay, et al., 2004). Today, CAI is used for tutorials, drill and practice, simulations, computation, and online or distance learning.

Research examining the effectiveness of CAI in each of these areas is limited. Many authors who have written on this subject have presented discussions on how to effectively implement CAI in the statistics classroom and have described methodologies to employ (Given-Larwin, 2004). Empirical studies that have investigated the impact of CAI on student achievement in statistics courses have reported mixed and conflicting findings. Considering that this research spans a 40-year period, during which the nature of the technology—along with its availability, capabilities, and student and instructor skill and comfort levels associated with its use—has changed dramatically, perhaps it isn't surprising that research results are often in conflict.

For example, some researchers have found that using CAI as a computational tool has a positive impact on student learning (Basturk, 2005;

McBride, 1996), but other research has suggested that, when used solely as a computation tool, CAI had no impact (Spinelli, 2001) or even a negative impact on student learning (Wang & Newlin, 2000). Wang and Newlin concluded that students who performed hand computations not only performed better on assessments, but also reported higher levels of self-confidence about their statistics abilities. Some researchers who incorporated CAI as a means to enhance lectures (e.g., animations, multimedia presentation, videos) found CAI had a positive impact on student performance on exams (Erwin & Rieppi, 1999; Wender & Muehloeck, 2003) and assisted students in comprehending abstract statistical concepts (Fusilier & Kelly, 1985). However, other research has suggested that CAI used for enhanced lectures resulted in a small negative impact on student performance on classroom assessments (Hilton & Christenson, 2002).

A number of researchers have found that CAI used as a tutorial has a positive impact on student learning (Aberson, 2003; Aberson, Berger, Healy, Kyle, & Romero, 2003; Bilwise, 2005; Marcoulides, 1990). However, Burruss and Farlow (2007) found no impact when the tutorial was used for reinforcing how to conduct a chi-square test. Similarly, other studies (Dimitrova, Percel, & Maisel, 1993; Gonzalez & Birch, 2000) found that although the CAI tutorial had a positive impact on student interest, it did not affect the learning or comprehension of students. Madigan (1991) found that the CAI tutorials were helpful to students understanding of probability, but this same type of tutorial had no impact on helping students to understand or conduct hypothesis testing. CAI in the form of simulations has been found to have a positive impact on students' learning of correlation (Morris, 2001) and abstract statistical concepts (Lane & Tang, 2000; Larwin & Larwin, 2011; Mills, 2004; Stockburger, 1982). At the same time, other researchers have found that CAI simulations have had a negative impact on student performance and learning (Lane & Aaleskic, 2002; Myers, 1990).

Results regarding CAI that use online delivery of instruction are equally conflicted. Much of the available research suggests that there is no difference in the learning and performance of students when instruction is delivered traditionally or using online mediums (Katz & Yablon, 2003; Palocsay & Stevens, 2008; Raymondo & Garret, 1998; Tsai & Pohl, 1980). However, Schutte (1996) found that the students in his virtually delivered class performed significantly better relative to the students in the same class with face-to-face delivery. Jones' (1999) attempt to replicate Schutte's (1996) study found the opposite pattern of results: The students in the face-to-face section performed significantly better. Potentially, these findings reflect the students' experiences and thus their attitudes, as suggested by Ware and Chastain (1989), and the students' freedom to make choices about their course delivery (Utts , Sommer, Acredolo, Maher, & Matthews, 2003).

Although research has suggested that smaller class sizes can result in greater learning outcomes when using CAI (Given-Larwin, 2004), this is not a consistent finding across the available research. For example, studies in which class sections are below 30 students have shown CAI to have positive impacts (e.g., Athey, 1987; Dinkins, 1985; Frederikson & Clifford, 2005), whereas others have revealed negative impacts (e.g., Grandzol, 2004; McClaren, 2004). Similarly, studies in which class sections exceed 100 students have also revealed positive impacts associated with CAI (Lane & Aleksic, 2002; Stephenson, 2001) as well as negative impacts (Katz & Yablon, 2003; Utts et al., 2003).

In light of all of the conflicting research regarding CAI with these variables as well as others, meta-analysis is the tool that can provide a general measure of the impact of CAI on student achievement in statistics instruction that might otherwise be obscured by these conflicting results. A meta-analytic investigation is an appropriate and effective approach to synthesizing and integrating the conflicting results from this quantitative research (Cooper & Hedges, 2009; Johnson, 1989).

There are two types of meta-analysis, one that uses primary data or raw data and one that uses summary or secondary data. An example of the former is a meta-analytic summary, which compares primary data and synthesizes data across a number of contexts when raw data are available. With this type of meta-analysis, the researcher is "learning by comparing studies" through additional analysis in an effort to further explore the phenomena under study (Cooper & Hedges, 2009, p 18). The current investigation is an example of the second type of meta-analysis. It is a type of meta-analysis commonly referred to as a quantitative literature review, or research synthesis using secondary data (Cooper & Hedges, 2009). In this type of meta-analytic study, the researcher uses the existing available research on a specific topic area to establish the overall strength of an effect, according to research that has already been conducted (Glass, McGaw, & Smith, 1981). The current investigation used such summary data acquired from completed research studies. Meta-analysis was used because it not only allows the synthesis of data across all available existing research, but also provides a mechanism by which moderator variables can be examined.

The purpose of the present meta-analysis is to determine the overall effectiveness of CAI on student achievement in graduate- and undergraduate-level statistics courses. To fully assess the impact of CAI, it is necessary to investigate how the impact of CAI differs from traditional approaches to statistics education. To date, three prior meta-analyses have looked at the impact of CAI on student achievement. Christmann and Badgett (1999) and, later, Hsu (2003) incorporated studies to assess the impact of a variety of microcomputer-based software packages on student achievement in undergraduate statistics classes. Given-Larwin (2004) investigated CAI for undergraduate and graduate students, looking additionally at pedagogical approaches as a moderator variable. Given-Larwin's 2004 investigation revealed an overall mean effect for CAI of $d = 0.329$ from 39 studies; Hsu's study found an overall mean effect size of $d = 0.430$ from 25 studies, whereas Christmann and Badgett's overall mean effect size was calculated at $d = 0.256$ from nine studies. The authors of all these meta-analytic studies recommended that more research needs to be conducted. The present study expands on these prior investigations. This meta-analysis accomplishes this by quantitatively summarizing prior empirical research examining the impact of CAI in terms of its impact on student achievement in statistics education. Incorporating all available research on the use of CAI in postsecondary statistics instruction, while meeting the stated inclusion criteria, reveals that the available research spans a 40-year period (1969–2010). The present study includes many studies not incorporated in the three prior meta-analytic studies on this topic. Thus, this is the first study examining the influence of CAI for postsecondary statistics education across four decades, and this study is significantly more comprehensive than prior meta-analyses.

Method

Based on the recommendations of Glass, McGraw, and Smith (1981), the researchers employed the following steps for conducting this meta-analytic study: First, we gathered and examined research studies. Studies included in the meta-analysis must fit within the defined parameters selected for analysis while representing as much of the population of data available on the research area. Glass et al. maintain that a thorough search must be conducted of the subject area. The next step, according to Glass et al. (1981), is to describe, classify, and code all the research studies to be included in the meta-analysis. In this step, measurement consistency is imperative. Glass et al. suggest that studies should be coded independently at least twice to establish rater agreement. The moderator variables that have been included for consideration must be clearly defined so that raters are able to make clear distinctions between the various classifications. For the purposes of this meta-analysis, we coded a random sample of studies twice in order to establish the reliability of the coding procedures. We tested moderator variables for interrater reliability and found reliable classifications more than 97.2% of the time. The final step in performing the meta-analysis, according to Glass et al. (1981), is the analysis of the overall mean effect-size measures and the mean effect-size measures for each research characteristic being examined. Once we calculated the effect-size measures, we interpreted and reported the results.

Sample of Studies

We obtained the studies included in this meta-analysis through an intensive computer search. During a 6-month period, we searched various electronic databases, including Academic Premier, American Statistics Association (amstat.org), Digital Dissertations, Educational Resources Information Circuit (ERIC), EBSCO, Electronic Journal Center (EJC), Excite, Netscape, Journal of Statistics Education (ncsu.edu), JSTOR, and PsychInfo. The search examined research spanning the years 1960–2010. The descriptive search criteria we employed to identify related materials included such combinations as teaching statistics, statistics instruction, and statistics education, as well as each of these criteria with the addition of college, university, or college students. We inspected abstracts of articles and discarded those articles that did not appear to meet the initial inclusion criteria. The inclusion criteria were: (a) articles examining instructional methods used in statistics education, (b) articles examining the instruction of postsecondary students, (c) articles examining the use of CAI, (d) articles presenting quantitative data, and (e) articles based on experimental or quasi-experimental designs. Postsecondary students include students who attend colleges or universities, as well as those professionals participating in professional development training.

We printed the relevant literature that was electronically available and ordered other relevant sources through the university library system. Next, we searched the reference list of each relevant article in an effort to find any additional pertinent publications. We reviewed all obtained articles, dissertations, presentations, and project reports and included in this meta-analysis those primary-level studies with participant populations and treatment populations of interest as well as the necessary statistical information. In all, we initially identified and examined for possible inclusion in this meta-analysis more than 123 studies using these methods.

We obtained a few studies through this search process that initially looked like candidates for inclusion in this meta-analysis, but careful inspection revealed that they did not meet the inclusion criteria discussed above. For example, two studies originally identified as studies of CAI were eliminated after careful reading and consultation with another researcher. The focus of these studies was not the impact of CAI, specifically, though they used CAI in the process of examining other questions.

In addition, studies included in this meta-analysis reported data in such a way that we could recalculate an effect size. These studies provided sufficient descriptive and inferential data, such as means, standard deviations, variances, t-tests, F-tests, and chi-square information, to allow for the calculation of effect sizes. If the necessary descriptive and inferential data were not provided, we made an attempt to contact the author of the study to acquire such information. We eliminated three studies that failed to provide the necessary information for the meta-analysis.

Coding of Studies

We coded each study according to the following information: (a) year of study, (b) source of research study, (c) whether the study was published or not, (d) mode of CAI, (e) type of intervention, (f) locations of use, (g) course delivery, (h) duration of CAI use in class, (i) discipline area, (j) level of statistics class,

(k) academic level of students, (l) number of instructors, (m) study research design, (n) outcome measures used, and (o) sample size of the study. The variables of "source of research study" and "whether the study was published or not" are secondary-level variables. Specifically, these variables should not have a direct impact on the outcome of the individual studies, whereas they may have an impact on the overall effect-size measures.

Publication bias is a concern when performing a meta-analysis, and a criticism of the meta-analytic approach (Wolf, 1986). Publication bias occurs when studies that find significant results for the effect being investigated are more likely to be published than studies that do not find significant findings. Publication bias has the potential of inflating the effect-size estimates (Hedges, 1986), and therefore it is important to include unpublished information when performing a meta-analysis. Thus, this meta-analysis addresses the issue of publication bias on two levels: First, this investigation includes 158 effect-size measures (72%) that are published and 61 effect-size measures (28%) that are not published. Second, many of the studies we included here were not investigating specifically the impact of CAI on student achievement as their primary focus. These studies focused on the impact of computer-assisted instruction on such issues as student anxiety or student attitudes toward statistics, with attention to the impact of CAI on achievement as a secondary concern. In these cases, a study might find significant results for the primary question about computer-assisted statistics instruction and attitudes and anxiety and nonsignificant results for the secondary questions about CAI and student performance. The study may have been published because it found significant results for the primary question or focus, yet because the nonsignificant results about the relationship between computer-assisted statistics instruction and performance were also reported, we classified these studies here as "published studies," resulting in a greater number of published studies with nonsignificant results than would otherwise be the case. For the present meta-analysis, 44 studies (20.1 %) reported that CAI did not have a positive impact on student achievement.

Calculation of Effect Sizes

Effect sizes are a statistical measure that attempts to represent the magnitude of the treatment effect in standard deviation units (Cohen, 1977). The effect size indicates the extent to which the treatment being tested is more effective than the condition of the control group—in other words, not simply indicating if differences exist, as represented by the p value, but indicating how big the differences are. Effect size can range from minus to plus infinity. For this meta-analytic study, we converted all statistics from each study to Hedges' d, a statistic defined as the difference between the means of the experimental and control groups divided by the intergroup standard deviation. Means and standard deviations were available to calculate 219 of the effect-size measures.

We calculated effect-size measures with means and standard deviations using the formula:

$$d = (ME\text{-}MC) / S_{pooled}$$

considering:

$$S_{pooled} = [(nE\text{-}1)\ (sE\ )^2 + (nC\text{-}1)\ (sC)\ ^2] / (nE\ \text{-}nC\ \text{-}\ 2)$$

(Cohen, 1977)

In this equation, ME is the mean for the experimental group, MC is the mean for the control group, nE is the number of participants in the experimental group, nC is the number of participants in the control group, sE is the standard deviation of the experimental group, and sC is the standard deviation of the control group. Additionally, we calculated within-group effect sizes using the Q statistic, a statistical test defined by Cochran (1954). The Q statistic is computed by summing the squared deviations of each study's effect estimate from the overall effect estimate, weighting the contribution of each study by its inverse variance using this formula:

$$Q = \sum wi\ (ESi - ES)2$$

In this equation, wi is the inverse variance weight for each effect size, ESi is the weighted mean effect size for each i, and ES is the weighted mean effect size across every effect size i (Cochran, 1954; Hedges & Olkin, 1985). Under the hypothesis of homogeneity among the effect sizes, the Q statistic follows a chi-square distribution with k – 1 degrees of freedom, where k is the number of effect sizes being assimilated (Given-Larwin, 2004). A significant test result indicates that heterogeneity exists with the group of effect sizes. Finally, 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:

$$\sigma d2 = [(nE + nC) / (nEnC)] + \{d2 / [2(nE + nC)]\}$$

(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. We computed a comprehensive effect-size measure and variable-based effect-size measures were using Comprehensive Meta-Analysis (2009), a software program designed specifically for meta-analytic research. Comprehensive Meta Analysis is a powerful computer program specifically designed for conducting meta-analytic investigations. With this program, researchers can choose the type of analyses

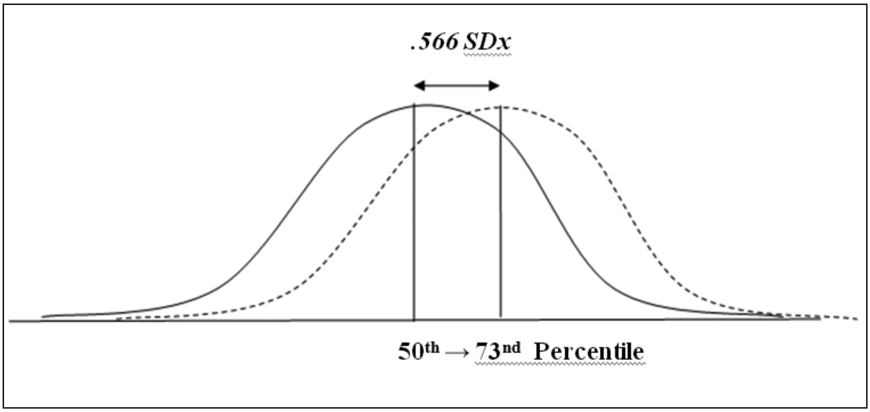

Figure 1. Graphical representation of standardized effect-size measures.

that best suit their study. CMA is currently the most flexible meta-analytic software available (W. Shadish, personal communication, November 2009).

Results

The primary purpose of this investigation was to determine the overall effectiveness of CAI in postsecondary statistics education. A comprehensive review of the literature yielded 71 studies meeting the inclusion criteria, including 56 peer-reviewed published journal articles, six conference papers, and nine dissertations. We removed one peer-reviewed study 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, N = 5,603). Due to the potential impact of these effect sizes on the overall weighted effect size, this study was deemed an influential outlier, and the analysis was run without these data included (Glass et al., 1981). Sample sizes prior to the elimination of Hilton and Christensen's (2002) study ranged from 16 to 5,603 students (M = 228.17, SD = 844.08). After the deletion, the sample sizes of the studies ranged from 16 to 480 students (M = 90.70, SD = 83.61).

Many of the studies included multiple assessment results (e.g., quiz scores, exam scores, etc.), therefore resulting in multiple effect-size measures within many of the studies. As a result, the current investigation included a total of 70 studies with a total of 219 effect-size measures. The effect sizes of the studies included in this investigation range from -9.45 to 14.52, yielding a grand mean overall effect-size measure d = 0.566, p <.001, a moderate effect sized according to the rough standards established by Cohen (1977). Cohen recommends the use of these rough measures for estimating the effect-size index when there has not been a lot of research in the area of interest and a standard of consideration has not been established. A 95% confidence level ranges from 0.437 to 0.694. This confidence interval does not contain the value of zero, implying that the treatment of CAI had a significant impact (Johnson, 1989). This effect

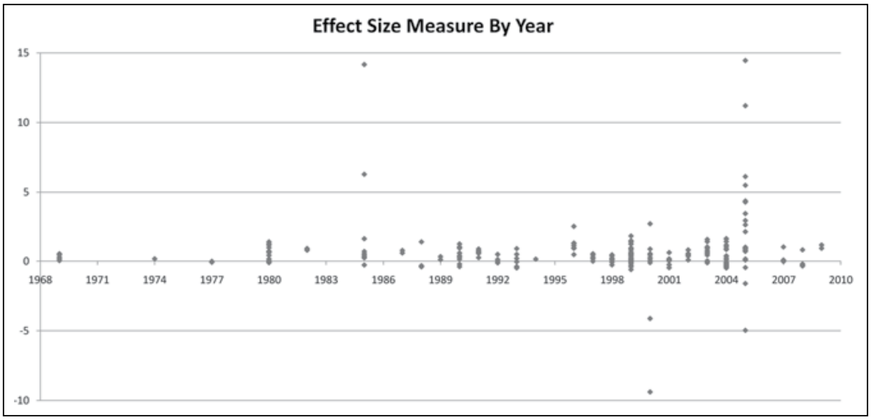

Figure 2. Time-series plot of the effect-size measures across time.

size suggests that the average student participating in a computer-assisted statistics instruction exceeds the academic achievement of approximately 73% of the students in traditional statistics instruction. Figure 1 (p. 261) presents a graphical representation of this difference.

Figure 2 presents a time-series plot of the effect-size measures across time. As Figure 2 reveals, the numbers of studies examining the impact of computer-assisted instruction on statistics education is increasing across time. With so few studies available across the first 20 years relative to the number of studies conducted during the second two decades, it can be valuable to examine these two periods of time more specifically. The increase in the number of studies during the second two-decade period increases the precision in estimating an effect size for that time span. An examination of the mean effect-size measure for the studies from 1969 through 1989 ($d = 0.447$) relative to the mean effect-size measure for studies from 1990 through 2010 ($d = 0.596$) indicates a significant change in effect-size measures ($Q\ [218] = 8.713, p < 0.001$).

One hundred and sixty-eight of the 219 effect sizes (76.7%) we included in these analyses were positive, indicating that computer-assisted statistics instruction had a positive impact on student learning. The remaining 51 studies (23.3%) had a negative effect, indicating that traditional approaches had a greater impact on student learning. These analyses also reveal that 98 (44.9%) of the 219 studies had an effect size of 0.5 or greater, indicating that the effect of CAI on student achievement was at least moderate. Table 1 (pp. 264–265) provides a breakdown of the studies meeting the inclusion criteria.

The grand mean analyses also revealed a $Q\ (218) = 3539.70, p < 0.001$ statistic, indicating significant heterogeneity across the 219 studies in this investigation. Therefore, further analyses are necessary to understand the variegation in effect sizes across the different studies. Table 2 (pp. 266–267) presents the summary of these analyses.

Specifically, these analyses explore the potential relationship between study characteristics and effect-size measures to determine which study characteristics influence the effect-size measures and which do not. As indicated above, we conducted analyses of within-group effects to assess if significant effect-size differences existed within each variable of study. Additionally, we also examined analyses of the standardized mean differences, revealing when effect-size measures were significantly different than what would be considered no significant effect. Significant mean effect-size measures, by category, are indicated with asterisk after the mean effect-size measure for each category.

Discussion, Limitations, and Conclusion

Discussion

The purpose of this meta-analysis was to investigate the impact of CAI on student achievement in postsecondary statistics education across four decades. The study also examined a number of variables that could potentially impact or mediate this relationship. This meta-analysis included a number of studies identified by a computerized literature search across many disciplines. With a total of 70 useable studies, we calculated 219 independent effect sizes. These studies comprise a total of 40,125 participants. The range of the effect sizes is 23.99, with a minimum effect-size measure of -9.45 and a maximum effect-size measure of +14.52. The overall mean effect measure for this group of effect sizes was d = 0.566. These findings indicate that the use of CAI can have a moderate impact on student achievement in postsecondary statistics education.

As these analyses revealed that significant heterogeneity existed across the 219 effect-size measures that were integrated into one overall mean effect-size estimate, it would not be adequate to attempt to describe this collection of studies with this single effect size. We conducted further analyses to explore the individual research characteristics and their potential influence on effect-size measures in an effort to explain this inconsistency across the individual effect-size measures.

An initial examination reveals significant variation in the secondary variables: source of research study and whether the study was published or not. Additionally, significant variation was revealed with a number of primary variables: year of study, mode of CAI, type of intervention, locations of use, course delivery, duration of CAI use in class, discipline area, level of statistics class, number of instructors, outcome measures used, and sample size of the study. We found no significant different effect-size measures across the categories for the variables of study research design and academic level of students.

Implications from 40 years

One outstanding feature of this meta-analysis is finding that the mean effect-size measures consistently increase across the four decades of research we examined in this study. The data suggest that CAI did not reveal

Table 1. The Primary Studies Included in the Meta-Analysis with Effect Sizes

Study	n of ES	ES range
Aberson et al. (2003)	1	0.249
Aberson et al. (2000)	2	0.821 to 1.457
Aberson et al. (2002)	1	0.061
Athey (1987)	2	0.621 to 0.803
Basturk (2005)	8	0.709 to 2.93
Benedict & Anderson (2004)	2	0.374 to 0.391
Bliwise (2005)	10	-4.98 to 14.528
Burruss & Farlow (2007)	5	-0.021 to 1.035
Christmann & Badgett(1997)	1	0.187
Collis, et al. (1988)	3	-0.327 to -0.403
Cybinski & Selvanathan (2005)	2	0.082
Debord, et al. (2004)	4	-0.139 to 0.228
Dinkins (1985)	1	1.695
Dimitrova, et al. (1993)	2	-0.398 to -0.400
Dixon & Judd (1977)	2	-0.074 to -0.005
Dorn (1993)	3	0.203 to 0.923
Erwin & Rieppi (1999)	3	0.190 to 1.428
Frederickson & Clifford (2005)	1	0.169
Fusilier & Kelly (1985)	2	6.33 to 14.168
Gilligan (1990)	2	0.290 to 0.399
Gonzalez & Birch (2000)	2	0.231 to 0.555
Grandzol (2004)	2	-0.505 to 1.099
Gratz, et al. (1993)	2	-0.487 to -0.036
Hall, et al. (1999)	12	0.409 to 1.88
Harrington (1999)	1	0.48
High (1998)	1	-0.269
Hollowell & Duch (1991)	2	-0.9
Hurlburt (2001)	3	0.061 to 0.118
Jones (1999)	3	-0.020 to 0.136
Katz & Yablon (2003)	2	-0.106 to -0.111
Koch & Gobell (1999)	1	0.951
Lane & Aleksic (2002)	3	0.369 to 0.497
Lane & Tang (2000)	2	0.493 to 0.879
Larwin & Larwin (2011)	2	0.943 to 1.183

Lesser (1998)	2	0.447 to 0.463
Madigan (1991)	6	-0.610 to 0.590
Marcoulides (1990)	3	0.594 to 1.252
McBride (1996)	2	1.177 to 2.612
McClaren (2004)	5	-0.408 to -0.081
Mills (2004)	6	0.865 to 1.676
Morris (2001)	2	-0.478 to 0.638
Morris et al. (2002)	4	0.094 to 0.831
Myers (1989)	6	-0.395 to 0.952
Olsen & Bozeman(1988)	1	1.432
Oswald (1996)	1	0.481
Petta (1999)	1	0.098
Palocsay & Stevens (2008)	3	-0.348 to -0.199
Porter et al. (2003)	3	0.942 to 1.333
Ragasa (2008)	1	0.827
Raymondo & Garrett (1998)	1	-0.072
Rosen et al. (1994)	1	0.151
Schutte (1996)	2	0.986 to 1.305
Skavaril (1974)	1	0.173
Smith (2003)	4	0.975 to 1.593
Song (1992)	4	-0.120 to 0.500
Spinelli (2001)	4	-0.254 to 0.022
Stephenson (2001)	1	0.173
Sterling & Gray (1991)	1	0.888
Stockburger (1982)	3	0.803 to 0.928
Summers et al.(2005)	1	-0.463
Tsai & Pohl (1980)	18	-0.101 to 1.463
Tubb (1977)	4	0.002 to 0.555
Utts et al. (2003)	2	-0.085 to 0.028
Wang & Newlin (2000)	4	-9.449 to 2.706
Ware & Chastain (1989)	2	0.114 to 0.331
Wassertheil (1969)	6	0.045 to 0.555
Wender & Muehlboeck (2003)	4	0.442 to 0.662
Weir, et al. (1991)	5	0.263 to 0.768
White (1986)	6	-0.283 to 0.732
Wilmouth & Wybraniec (1998)	4	0.100 to 0.390

Table 2. Summary of Analysis Results across Study Characteristics

Variables and Categories	Number of Effect Sizes (N)	Within-Group Effects	Mean Effect Size (d+)
Year of Study		21.47*	
1960–1969	6		0.342
1970–1979	7		0.085
1980–1989	36		0.386*
1990–1999	75		0.420*
2000–2010	95		0.761*
Study Source		10.64*	
Journal Publication	158		0.638*
Conference Paper	31		0.438*
Dissertations/Theses	30		0.248*
Publication Status		7.562*	
Published	158		0.683*
Not Published	61		0.348*
Focus of CAI Use		28.56*	
Tutorial	32		1.670*
Drill and Practice	39		0.478*
Online Delivery	37		-0.156
Computation	37		0.514*
Simulation	48		0.579*
Enhanced Lecture	26		0.441*
Supplement or Replacement		36.04*	
Supplemental	154		0.795*
Replacement	65		0.060
Location of CAI Use		30.84*	
In Class	165		0.738*
Online Delivery	44		0.033
Homework	5		0.011
In Class and Homework	5		0.370
Course Delivery		48.11*	
Face-to-Face	183		0.706*
Online Delivery	34		-0.149
Hybrid	2		-0.035
Duration of CAI Use		17.25*	
Entire Semester	138		0.360*
Several Classes	63		1.033*
One Lesson	18		0.700*

Discipline of Students		20.53*	
Biology	8		0.182
Business	38		0.259*
Catch-All Class	55		0.543*
Education	21		0.531*
Mathematics	8		0.204*
Psychology	76		0.851*
Sociology/Social Work	8		0.502*
Criminal Justice	5		0.222
Level of Statistics Class		9.38*	
Introductory	203		0.596*
Intermediate/Advanced	16		0.257*
Academic Level of Students		1.63	
Undergraduate	175		0.529*
Graduate	26		0.754*
Mixed Class	18		0.604*
Number of Instructors (Bias)		12.14*	
One Instructor	167		0.688*
Multiple Instructors	52		0.207
Study Research Design		0.360	
Experimental	50		0.506*
Quasi-Experimental	169		0.572*
Outcome Measures Used		17.98*	
Homework Grades	6		1.026*
Quiz Grades	91		0.852*
Exam Grades	104		0.430*
Class GPA	18		-0.054
Sample Size of Study		19.00*	
up to 25	41		0.669*
26 to 50	51		0.514*
51 to 100	57		0.273*
multiple classes ($\geq$100)	70		0.786*

Note: * $p < 0.05$

any significant impact on student achievement until the 1980s (*d* = 0.386). Since the 1980s, the level of impact has consistently increased in the research across the 1990s (*d* = 0.420) and 2000s (*d* = 0.761), with the greatest gains in impact found between 1990 and 2000. This is as would be expected. Students and instructors are more technologically competent and comfortable with technology, computers and software have continued to evolve, and the variety of available software for computation, simulation, and tutorial applets continues to expand. Additionally, reform efforts in statistics education have encouraged and enabled statistics instructors to incorporate technology into their pedagogy (Cobb, 2007).

Implications of Delivery

Another notable feature of this study is the results regarding the mode of CAI. Specifically, the data revealed that the use of CAI is most beneficial when CAI was used as a tutorial (*d* = 0.849), for computation purposes (*d* = 0.525), and for simulations (*d* = 0.461). Using CAI for drill-and-practice or to enhance lectures produced smaller yet significant effect-size measures (*d* = 0.361 and 0.372 respectively). The use of CAI, strictly in an online format, actually produced a negative effect size (*d* = -0.035). These results again reflect the growing number of resources that are available for CAI in the classroom and on the Internet.

Consistent with these results, where the CAI was used also impacted the magnitude of effect size on student achievement. Location of CAI revealed that CAI was not effective if its use was located strictly online (*d* = 0.052) or independently as homework (*d* = 0.066). Using CAI during face-to-face class meetings (*d* = 0.509) or in both face-to-face class meetings (including lab time) and as homework (*d* = 0.379) produced the largest effects. Finally, the course delivery variable revealed that face-to-face courses produced the greatest effect on student achievement when using CAI (*d* = 0.706), whereas the use of CAI with courses categorized as "online delivery" produced a negative effect on student achievement (*d* = -0.149).

Additionally, how much the CAI was implemented impacted the level of effect on student achievement. Specifically, we found that using CAI to supplement instruction had a good impact on student achievement (*d* = 0.539), whereas CAI as the only means of instruction provided no impact on student achievement (*d* = 0.06). One explanation for these results might be that many of the studies included in the category of "complete replacement" were studies of online delivery courses, which also did not reveal a positive impact on student achievement.

The poor results associated with online instruction in the current investigation are likely a reflection of the early stages of online instruction. Exclusive online delivery of courses is a relatively new phenomenon that has become widely accepted only within the last decade. Much like CAI, the effect-size measures associated with online instruction will likely improve over the next decade as research expands on best practices with this medium of delivery and as this knowledge finds its way into the virtual classroom. It is also possible that, due to its abstract nature, the effectiveness of using online instruction in statistics education might lag behind other subject areas. One other possible consideration, which can only be decided with time, is whether online instruction is suitable for the delivery of statistics education. This is a question that might better be visited after another decade of online instruction.

Results suggest that the greatest impact was found when CAI was incorporated into several class meetings (d = 1.033), followed by a single class (d = 0.700) and over the entire semester (d = 0.360). One explanation for the high level of impact for a single class is that many of these studies were investigations in which a quiz was used as the means to assess student understanding. The studies included in this investigation demonstrated greater effect-size measures when quizzes were used as the assessment of learning.

Most studies used quizzes (n = 91) and exams (n = 104) to assess student achievement; quizzes were associated with a large effect-size measure (d = 0.852), and exams were associated with a moderate effect-size measure (d = 0.430). Although only six studies (2%) used homework assignments, this category was associated with highest effect-size measures (d = 1.026). Unfortunately, homework assignments are problematic in testing individual achievement, as there is nothing to stop students from working cooperatively. Class GPA (n = 18) was associated with no significant effect-size measures (d = -0.054).

One potential explanation for the substantially higher effect-size measure associated with quizzes is suggested by the idea of dynamic assessment (Sternberg, 2007). According to Sternberg, assessment processes that take place more immediately after instruction are likely to result in higher effects relative to assessment that is more stagnate in nature, such as exams. For the current investigation, quizzes produced an effect-size measure that was approximately twice as large as it was for exams throughout the studies included in this investigation. Additionally, the information that is covered on a quiz-type assessment is generally more focused and potentially produces less anxiety for the student relative to what might be experienced on an exam.

As indicated above, the academic level of the student (and therefore the course) did not reveal significant differences across the different categories. However, the level of the statistics class (whether covering introductory or advanced statistical topics) did reveal statistically different effect-size measures. Based on the studies included in the meta-analytic investigation, students in the introductory courses benefited more from CAI (d = 0.596) relative to students in the advanced sections (d = 0.257). Unfortunately, only 16 studies (7%) included data from intermediate/advanced statistics sections, which makes drawing strong conclusions from these differences difficult. This result may simply reflect the fact that a prevailing amount of the tutorials, simulations, etc., are focused around introductory topics and concepts.

The variable examining the potential of instructor bias revealed a significantly different magnitude of effect when one instructor teaches both control and experimental conditions ($d = 0.688$) relative to multiple instructors teaching separate control and experimental conditions ($d = 0.207$). Instructor bias is a concern if one instructor teaching both conditions teaches in a manner in which he or she consciously or unconsciously provides an advantage or reports better results for the experimental group. However, several meta-analytic studies examining technology use in instruction have actually found higher effect sizes when different instructors teach each condition (Cohen, 1980; Kulik & Kulik, 1986). For the current investigation, only 52 studies (23%) included multiple instructors, but many of these studies included larger course sections ($n = 41$, 72%), potentially explaining the weaker effects for multiple instructors. As indicated by the examination of the study size variable, smaller studies (and classes) resulted in the greatest impact when using CAI ($n = 41$, $d = 0.669$). Studies examining multiple sections of the same class did not give detailed information for the class sizes specifically; however, these also resulted in a large impact when CAI was employed ($n = 70$, $d = 0.786$). A breakdown of this data revealed that when information was available, smaller sections produced the greatest impact, as would be expected. Consistent with these results, Given-Larwin (2004) found a $d = 0.685$ effect-size measure for classes of 25 students or less.

Finally, the home discipline of students taking the class created variation in the effect-size measures revealed. Data for biology ($n = 8$) and criminal justice (n = 5) majors revealed no significant effect-size measures; however, this is likely due to the few studies represented for these disciplines. The greatest impact was observed for psychology ($n = 76$, $d = 0.851$) , followed by catch-all sections ($n = 55$, $d = 0.543$), education sections ($n = 21$, $d = 0.531$), sociology sections ($n = 8$, $d = 0.502$), business sections ($n = 38$, $d = 0.259$), and mathematics sections ($n = 8$, $d = 0.204$). Catch-all sections included sections that statistics departments generally offered for the greater population of undergraduates at the respective university. These general statistics sections are often allowed to be substituted for a general mathematics requirement, such as college algebra. These types of sections attract students from all disciplines and are often the only class in statistics (or mathematics) that these students will be asked to complete. Because these types of sections tend to draw students who are trying to avoid the otherwise required college algebra class, these students often have a dislike for mathematics, which is potentially one reason why CAI was so effective with this group of students.

Limitations

There were several unavoidable limitations associated with this research study. A number of the individual constructs were represented in a small sample of studies. Although this might have occurred as the result of an insufficient computer literature search strategy, that was not the case for the current investigation. The literature search process was thorough and exhaustive and turned up a number of additional empirical research studies not included in the past meta-analytic reviews of this subject area.

For example, this initial limitation applies to the variable students' education level. Although many graduate programs require that their graduate students take at least one statistics course during their course of study, a large percentage of the available research on CAI investigates undergraduate statistics education and predominantly students enrolled in introductory-level statistics courses. The result is that there are relatively few studies that look at CAI in graduate statistics courses in particular. As indicated above, this might simply be the result of little CAI that is geared to graduate students, other than the use of statistical software packages, such as SAS and SPSS.

Another limitation of the present study is associated with meta-analytic studies in general. It can be the case with the meta-analytic approach that it is difficult to break categories down enough to examine as much information as possible without creating too much overlap in the results. Although these overlaps in categories can be used as a form of "triangulation" and a reliability check, they can also cause redundancy and useless repetition. Also with the meta-analytic approach, the meta-analytic researcher is at the mercy of the authors who have conducted research in the area. The researcher has to rely on the authors or individual researchers to report results accurately, describe the studies well, report statistics appropriately, and respond to inquiries about their research if there are any questions or discrepancies.

This results in another layer of concern that has to do with the selection of variables for possible examination in a meta-analysis. The variables the meta-analytic researcher wishes to examine may not be variables on which the authors of the individual studies typically report usable data. For example, it might be desirable to examine variables such as gender; student characteristics, such as traditional versus nontraditional status; or school variables, such as public versus private institutions of learning. The examination of these kinds of variables may have revealed additional insight, complexities, and moderator effects about the effectiveness of CAI on student performance. However, because an insufficient number of studies conducted included the necessary information, data for variables such as gender cannot be specifically examined.

Finally, interpreting the results of meta-analytic studies such as this one can be challenging when the content area is itself undergoing dynamic transition over the period the study spans. For example, what constituted CAI in statistics 30 or 40 years ago is different than what it involves today. This is in part due to the fact that technology itself is changing. Computer-assisted instruction making use of simulations in the 1980s (e.g., Stockburger, 1982) looks different and works differently than simulations used in the mid to late

2000s (e.g., Lane & Tang, 2000), even if the underlying idea the simulation was intended to convey remains the same. Thus, comparing studies using CAI across a long span of years may not always involve comparisons of the same thing.

However, this consideration may also be at the heart of the gains in effect size observed in this meta-analysis across the 40-year span it covers. If the nature of the underlying technology and the CAI that students receive is changing over the years, and if instructors' comfort level with, understanding of, and willingness to incorporate CAI is also improving, that might help explain these changes. If this results in statistics instructors better able to link pedagogical content with the use of CAI, such gains in effect size as those observed in the present study would be expected. And changes in the students' experience may be a contributing factor as well. As students have better access to technology and their comfort level with it increases, we would expect CAI capable of having a greater effect. Thus, these dynamic changes in CAI content, delivery, and experience over the past several decades may be responsible for the observed gains in effect size, while also making detailed comparisons across lengthy time spans somewhat challenging.

Conclusions

The implications of the present research are extensive. The primary question this meta-analysis sought to examine is whether or not students enrolled in postsecondary statistics courses benefit from CAI, as evidenced by their achievement scores. The overall mean effect-size measure of $d = 0.566$ obtained in this meta-analysis indicates that CAI has a moderate impact on student achievement in postsecondary statistics education, according to Cohen's (1977) classification of effect-size measures. This effect-size measure indicates that CAI does have an impact on student performance, and the data across the four decades included in this meta-analytic study suggest that the impact of CAI is growing. However, this moderate impact also suggests that CAI is not a panacea for all the ills that might plague statistics education. There may be ways to make the use of CAI in statistics education more effective, so that by the next decade, the benefits of CAI can be maximized. The findings of this study suggest a number of ways that this might happen.

For example, the investigation of the particular research characteristics included in this study revealed several characteristics that had a greater impact on student performance than others. Incorporating these variables, along with CAI, into a broader pedagogical philosophy with respect to statistics education may boost student performance and understanding beyond what could be achieved with CAI alone. With this in mind, the findings of this meta-analysis suggest several components that could be incorporated into a formula or model for the optimal delivery of statistics instruction that would maximize student achievement. Such a formula or model would include small class sections that use CAI to enhance or supplement activities, and using the computer for the purposes of tutorial, simulation, computation and drill and practice.

As discussed earlier, some researchers (e.g., Song, 1992; Wang, 1999) have suggested that students can easily become dependent on CAI, and at the same time fail to fully grasp the statistical concepts they are supposed to be learning. Others researchers (Harrington, 1998; High, 1999; Sterling & Gray, 1991; Utts et al., 2003) have found that students do not necessarily evaluate CAI as a positive addition to their statistics education, and if given the choice, these students would opt for a course without the computer intervention. With the formula or model described above, students have the opportunity and challenge to develop their conceptual understanding of statistics through activities and practical application that are provided via CAI. The component of smaller classes with cooperative learning groups can provide the bridge for the computer novice to find success in a computer-based learning environment (Given-Larwin, 2004).

CAI in statistics education can be a double-edged sword: It may better prepare students to use computers and statistics in an increasingly technology-based society, and it may create more time and space in the statistics course for instructors to focus on something other than endless chalkboard calculations. Unfortunately, it doesn't necessarily ensure that students will learn the fundamental statistical concepts, and in some cases it may compete or interfere with such concept mastery. Haphazardly applied, CAI in statistics education could create a dangerous lot of statistical technicians who will use statistical packages they do not fully understand to pump out results that they cannot reliably or accurately interpret.

Based on the results of this meta-analysis, CAI, properly applied as an enhancement or supplement to small class sections, can be very beneficial to student achievement. Coupling introductory courses with activity-based learning in which CAI is used for drill and practice, as a form of tutorial, and to provide students with computational experiences and simulations of concepts, can have a positive impact on student performance. CAI has the potential to equip students with the necessary tools to effectively use the knowledge they have acquired to become more critical consumers of information and users of statistical concepts and applications.

Author Notes

Karen Larwin received her PhD in evaluation, measurement, and statistics in 2007. She has been teaching statistics and quantitative methods to undergraduate and graduate students since 2006. She is currently the lead program chair of the American Evaluation Association Quant TIG as well as an active member of the United States Consortium for the Advancement of Undergraduate Statistics Education. She works at Youngstown State University as a member of the doctoral faculty for the department of Foundations, Research, Technology, and Leadership. Please send correspondence to Karen Larwin, Youngstown State University, 14601 Seacrist Road, Salem, Ohio 44460. E-mail: drklarwin@yahoo.com

David Larwin, MA, is a psychology professor for Kent State University at Salem, where he teaches personality psychology, social psychology, general psychology, and quantitative methods in psychology. He is an active member of the American Psychological Association as well as a contributing member of the American Evaluation Association. Please send correspondence to David Larwin, Kent State University at Salem, 14601 Seacrist Road, Salem, Ohio 44460. E-mail: dlarwin@kent.edu

References

References marked with an asterisk (*) indicate studies included in the meta-analysis.

*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*, 75–79.

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

*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), 1–7. Retrieved February 28, 2006, from http://www.amstat.org/publications/jse/v10n3/aberson.html

*Athey, S. (1987). A mentor system incorporating expertise to guide and teach statistical decision making. (Doctoral dissertation. The University of Arizona, 1987). *Dissertation Abstracts, 48*, 0238.

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

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

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

Bratz, A., & Sabolik, M. A. (2001). Computer and software use in teaching the beginning statistics courses. *Teaching of Psychology, 28*, 147–151.

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.

*Burruss, G. W., & Furlow, M. H. (2007). Teaching statistics visually: A quasi-experimental evaluation of teaching chi-square through computer tutorials. *Journal of Criminal Justice Education, 18*(2), 209–230.

*Christmann, E. P., & Badgett, J. L. (1997). The effects of computers and calculators on male and female statistics achievement. *Journal of Computing in Higher Education, 9*, 49–58.

Cobb, G. W. (1993). Reconsidering statistics education: A National Science Foundation Conference. *Journal of Statistics Education, 1*, 1–33.

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

Cochran, W. G. (1954). Some methods for strengthening the common $\chi 2$ tests. *Biometrics, 10*, 417–451.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). NY: Academic Press.

*Collis, B., O'Berg, A., & Shera, W. (1988). An evaluation of computer-based instruction in statistical techniques for education and social work students. *Journal of Educational Technology Systems, 17*, 59–71.

Comprehensive Meta-Analysis. (2009). Retreived October 20, 2009, from http://www.meta-analysis.com

Cooper, H. M., & Hedges, L. V. (2009). Research synthesis as a scientific process. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 1–18). New York: Russell Sage Publications.

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.

*Cybinski, P., & Selvanathan, S. (2005). Learning experience and learning effectiveness in undergraduate statistics: Modeling performance in traditional and flexible learning environments. *Decision Sciences Journal of Innovative Education, 3*(2), 251–271.

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

*Dinkins, P. (1985). *Development of a computer-assisted instruction courseware package in statistics and a comparative analysis of three management strategies.* (Doctoral dissertation, Louisiana State University, 1985). Dissertation Abstracts International, 1-47.

*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 of Carbondale, 1993). *Dissertation Abstracts International, 55*, 1–277.

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

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

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

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education, 1*, 1–11.

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*, 25–34.

Garfield, J. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education, 10*, 1–19.

*Gilligan, W. P. (1990). The use of computer statistical package in teaching a unit of descriptive statistics. (Doctoral dissertation, Boston University, 1990). *Dissertation Abstracts International, 51*, 1–230.

Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education, 5*, 1–14.

Given-Larwin, K. H. (2004). *The impact of computer-assisted instruction on student achievement in postsecondary statistics education: A meta-analysis.* (Unpublished Master's Thesis). Kent State University, Kent, Ohio.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* London: Sage Publications.

*Gonzalez, G. M., & Birch, M. A. (2000). Evaluating the instructional efficacy of computer-mediated interactive multimedia: Comparing three elementary statistics tutorial modules. *Journal of Education Computing Research, 22*, 411–430.

*Grandzol, J.R. (2004). Teaching MBA statistics online: A pedagogically sound process approach. *Journal of Education for Business, 3*, 237–244.

*Gratz, Z. S., Volpe, G. D., & Kind, B. M. (1993). *Attitudes and achievement in introductory psychological statistics classes: Traditional versus computer supported instruction.* Proceeding of the Annual Conference on Undergraduate Teaching of Psychology, Ellenville, New York. (ERIC Document Reproduction Service No. ED365405).

*Hall, R., Pilant, M, & Strader, R. (1999). *The impact of Web-based instruction on performance in an applied statistics course*. In D. A. Thomas (Ed.), Proceedings of International Conference on Mathematics/Science Education and Technology, (pp. 261–266). Chesapeake, VA: AACE Publications.

*Harrington, D. (1999). Teaching statistics: A comparison of traditional classroom and programmed instruction/distance learning approaches. *Journal of Social Work Education, 9*, 1–11.

Hedges, L. V. (1986). Issues in meta-analysis. In E. Z. Zothkoph (Ed.), *Review of Research in Education, 13*, 353–398. Washington, DC: American Education Research Association.

Hedges, L. V., & Olkin, D. (1985). *Statistical methods for meta-analyses.* San Diego, CA: Academic.

*High, R. V. (1998). *Some variables in relation to students' choice of statistics classes: Traditional versus computer-supported instruction.* (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] (pp. 1-6). Cape Town: International Statistical Institute.

Hogg, R. V. (1991). Statistics education: Improvements badly needed. *The American Statistician, 45*, 342–343.

*Hollowell, K. A., & Duch, B. J. (1991). *Functions and statistics with computers at college level.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No.336090).

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

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

Johnson, B. T. (1989). *DSTAT: Software for the meta-analytic review of research literatures.* NJ: Lawrence Erlbaum Associates.

*Jones, E. R. (1999). *A comparison of an all Web-based class to a traditional class.* In J. Price et al. (Eds.), Proceedings of Society for Information Technology & Teacher Education International Conference 1999 (pp. 1372–1377). Chesapeake, VA: AACE.

*Katz, Y. F., & Yablon, Y. B. (2003). Online university learning: cognitive and affective perspectives. *Campus Wide Information Systems, 20*, 48–54.

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

*Lane, J. L., & Aleksic, M. (2002). *Transforming elementary statistics to enhance student learning.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

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

*Larwin, K., & Larwin, D. (2011). Evaluating the use of random distribution theory to introduce statistical inference concepts to business students. *Journal of Education for Business, 86*(1), 1–9.

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

Lindsay, B., Kettenring, J., & Siegmund, D. (Ed.) (2004). *Statistics: Challenges and opportunities for the 21st century.* NSF Report. Retrieved August 27, 2004, from http://www.amstat.org/news/nsf4Aug04.pdf

*Madigan, E. (1991). *An expert system as a study guide in an elementary statistics class and its effect on learning.* (Doctoral Dissertations, Florida Institute of Technology, UMI No. 9137077).

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

*McBride, A. B. (1996). Creating a critical thinking learning environment: Teaching statistics to social science undergraduates. *PS: Political Science & Politics, 29*, 517–521.

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

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

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. (1989). *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,1989). Dissertation Abstracts International, 51, 1–441.

*Olsen, C. R. & Bozeman, W. C. (1988). Decision support systems: Applications in statistics and hypothesis testing. *Journal of Research on Computing in Education, 20*, 206–212.

*Oswald, P.A. (1996). Classroom use of personal computers to teach statistics. *Teaching of Psychology, 23*, 124–126.

*Palocsay, S.W., & Stevens, S.P. (2008). A study of the effectiveness of Web-based homework in teaching undergraduate business statistics. *Decision Sciences Journal of Innovative Education, 6*(2), 213–232.

*Petta, N. A. (1999). *The impact of Web-based class management system on student performance and attitudes in a quantitative statistics class.* (Doctoral Dissertations, University of Nevada, UMI order no. ADG75-10784).

*Porter, A., Griffiths, D., & Hedberg, J. (2003). From classroom to online teaching: Experiences in improving statistics education. *Journal of Applied Mathematics and Decision Sciences, 7*, 35–73.

*Ragasa, C.Y. (2008). A comparison of computer-assisted instruction and traditional methods of teaching basic statistics. *Journal of Statistics Education, 16*(1), 1–20.

*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.

Roiter, K., & Petocz, P. (1996). Introductory statistics courses: A new way of thinking. *Journal of Statistics Education, 4*, 1–16.

*Rosen, E. F., Feeney, B., & Petty, L. C. (1994). An introductory statistics class and examination using SPSS/PC. *Behavior Research Methods, Instruments, and Computers, 26*, 242–244.

*Schutte, J.G. (1996). *Virtual teaching in higher education: The new intellectual superhighway or just another traffic jam?* Retrieved on September 5, 2004, from http://www.csun.edu/sociology/virexp.htm

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

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

Snee, R. D. (1993). What's missing in statistical education? *The American Statistician, 47,* 149–154.

*Song, P. W. (1992). *The effect of using computer software and computer simulations to teach statistics and probability in a Korean college.* University of Iowa, Doctoral Dissertation. Available from ProQuest Dissertations and Theses Database. (UMI No. 9237014)

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

*Stephenson, W. R. (2001). Statistics at a distance. *Journal of Statistics Education, 9,* 1–12.

*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.

Sternberg, R. J. (2007). Culture, instruction, and assessment. *Comparative Education, 43*(1), 5–22.

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

*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.

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

*Tubb, G. W. (1977). *Current use of computers in the teaching of statistics.* Paper presented at the meeting of the Computer Science and Statistics annual symposium, Gaithersbury, Maryland. (ERIC Document Reproduction Service No. ED432286).

*Utts, J., Sommer, B., Acredolo, 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*, 1–18.

*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.

*Wassertheil, S. (1969). Computer assistance in statistics. *Improving College and University Teaching, 17*(4), 264–266.

*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.

*White, S. L. (1986). *Teaching introductory statistics: Hand calculations versus computer data analysis.* California State University Doctoral Dissertation. Available from ProQuest Dissertations and Theses Database. (UMI No. 1328044).

*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.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* London: Sage Publications.

Yilmaz, M. R. (1996). The challenges of teaching statistics to non-specialists. *Journal of Statistics Education, 4*, 1–12.