A Quantitative Synthesis of Recent Research on the Effects of Teaching and Learning With Technology on Student Outcomes

December 2002
A Quantitative Synthesis of Recent Research on the Effects of Teaching and Learning With Technology on Student Outcomes

December 2002

Hersh C. Waxman
Michael L. Connell
University of Houston

Jon Gray
Lamar University

Copyright © 2002 by the North Central Regional Educational Laboratory. All rights reserved.

This work was produced in whole or in part with funds from the Institute of Education Sciences (IES), U.S. Department of Education, under contract number ED-01-CO-0011. The content does not necessarily reflect the position or policy of IES or the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by the federal government.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>6</td>
</tr>
<tr>
<td>Method</td>
<td>7</td>
</tr>
<tr>
<td>Search and Selection Criteria</td>
<td>7</td>
</tr>
<tr>
<td>Procedure</td>
<td>8</td>
</tr>
<tr>
<td>Results</td>
<td>9</td>
</tr>
<tr>
<td>Description of Studies in Review</td>
<td>9</td>
</tr>
<tr>
<td>Overall Results</td>
<td>10</td>
</tr>
<tr>
<td>Discussion</td>
<td>12</td>
</tr>
<tr>
<td>Research Quality Issues</td>
<td>12</td>
</tr>
<tr>
<td>Limitations of the Present Study</td>
<td>12</td>
</tr>
<tr>
<td>Conclusions</td>
<td>14</td>
</tr>
<tr>
<td>References</td>
<td>15</td>
</tr>
<tr>
<td>Appendix: Information Coded for Each Study</td>
<td>21</td>
</tr>
</tbody>
</table>
Abstract

To estimate the effects of teaching and learning with technology on students’ cognitive, affective, and behavioral outcomes of learning, 138 effect sizes were calculated using statistical data from 20 studies that contained a combined sample of approximately 4,400 students. The mean of the study-weighted effect sizes averaging across all outcomes was .30 ($p < .05$), with a 95-percent confidence interval (CI) of .004 - .598. This result indicates that teaching and learning with technology has a small, positive, significant ($p < .05$) effect on student outcomes when compared to traditional instruction. The mean study-weighted effect size for the 13 comparisons containing cognitive outcomes was .39, and the mean study-weighted effect size for the 60 comparisons that focused on student affective outcomes was .208. On the other hand, the mean study-weighted effect size for the 30 comparisons that contained behavioral outcomes was -.154, indicating that technology had a small, negative effect on students’ behavioral outcomes. The overall study-weighted effects were constant across the categories of study characteristics, quality of study indicators, technology characteristics, and instructional/teaching characteristics.
Introduction

Education often has been characterized as the only field where personal experience and ideology are relied on to make policy choices because the research base is inadequate and rarely used (National Research Council, 1999). The federal No Child Left Behind Act of 2001, however, is placing a new emphasis on scientifically based research and is requiring states and school districts to choose “evidence-based” programs for their schools and classrooms. This change is providing support to the growing numbers of researchers (Glass, 2000) and organizations, such as the Campbell Collaboration (2002), which use the statistical technique of meta-analysis to synthesize findings from research. It is argued that these systematic reviews of the research will firm up the “soft science” of education and finally begin to provide empirical evidence that certain programs or approaches are effective in improving student outcomes (Viadero, 2002).

During the past three decades, a large number of meta-analyses have systematically examined the effects of technology on student outcomes. Several meta-analyses, for example, have investigated the impact of computer-assisted instruction on student outcomes (Lipsey & Wilson, 1993). Other meta-analyses have examined aspects such as the effects of microcomputer applications in elementary schools (Ryan, 1991) and the effects of computer programming on student outcomes (Liao & Bright, 1991). Niemiec and Walberg (1992) summarized the findings on 13 quantitative research syntheses that were conducted between 1975 and 1987 and found that the average effect size was .42, which indicated that the average student who received computer-based instruction scored at the 66th percentile of the control group distribution (i.e., the 50th percentile).

Overall, these meta-analyses—along with some recent, major studies and narrative reviews of the research—have documented the positive effects of educational technology on student achievement (Schacter, 2001; Sivin-Kachala, 1998; Wenglinsky, 1998). These studies, reviews, and meta-analyses, however, typically look at different aspects or types of technology. Furthermore, this knowledge base has not really provided information on how to appropriately integrate and use technology in schools and classrooms. In addition, recent improvements regarding the quality and quantity of technology in schools suggest that technology in schools today is dramatically different than the technology that was used in schools several years ago. This rapid growth and improvement in technology exceeds current knowledge of how to effectively use technology in schools (Allen, 2001) and suggests that the impact of technology is different today than it was in the past.

Although many of the meta-analyses examining the effects of technology on student outcomes were conducted more than a decade ago, several recent meta-analyses have focused on specific aspects of technology. Blok, Oostdam, Otter, and Overmaat (2002), for example, examined the effectiveness of computer-assisted instruction (CAI) programs in supporting beginning readers. Their review included 42 studies from 1990 onward, and they found the corrected overall effect size estimate was .19. Their findings were similar to earlier meta-analyses by Kulik and Kulik (1991) and Ouyang (1993), which also examined the effects of CAI and found it to have positive but small effects.
Lou, Abrami, and d’Apollonia (2001) examined the effects of students working in a small group versus working individually when students were using computer technology. They found that small-group learning had more positive effects than individual learning. Other recent meta-analyses in technology have examined topics such as the effectiveness of interactive distance education (Cavanaugh, 2001), computer-assisted instruction in science education (Bayraktar, 2001-2002), and computer-based instructional simulation (Lee, 1999). Furthermore, other recent meta-analyses have examined the effects of computer-assisted instruction on student achievement in differing science and demographic areas (Christmann & Badgett, 1999), microcomputer-based computer-assisted instruction within differing subject areas (Christmann, Badgett, & Lucking, 1997), gender differences in computer-related attitudes and behavior (Whitley, 1997), and the effectiveness of computer-assisted instruction on the academic achievement of secondary students (Christmann, Lucking, & Badgett, 1997). Some recent meta-analyses that have not yet been published have focused on the uses of educational technology in home and school (Penuel et al., 2002) and discrete educational software (Murphy, Penuel, Korbak, Whaley, & Allen, 2002).

Table 1 (on page 5) presents a summary of nine recent meta-analyses in the area of educational technology that have been published in peer-reviewed journals. The median effect size of the seven reported effect sizes is .21, which represents a small positive effect with the experimental group scoring at the 58th percentile of the control group distribution. These meta-analyses, however, also found that their particular treatments had several differential effects on their outcomes. Lee (1999), for example, found that although computer-based simulation had a modest, positive effect size of .41 on student achievement, it had a negative effect size of -.04 on student attitudes. The ability to examine differential effects of the treatment is one of the many advantages of meta-analysis as a meaningful method to aggregate and report educational findings.

One area in which there have not been many meta-analyses and systematic reviews of the research is in how teaching and learning with technology impacts student outcomes. This area is important because some studies have found that technology can change teachers’ pedagogic practices from a teacher-centered or teacher-directed model to a more student-centered classroom where students work cooperatively, have opportunities to make choices, and play an active role in their learning. Swan and Mitrani (1993), for example, compared the classroom interactions between high school students and teachers involved in (a) computer-based instruction and (b) traditional instruction. They found that student-teacher interactions were more student-centered and individualized during computer-based teaching and learning than in traditional teaching and learning. In another study that examined changes in classroom instruction as a result of technology, Sandholtz, Ringstaff, and Dwyer (1992) found that high access to computers enabled teachers to individualize instruction more. In a national study, Worthen, Van Dusen, and Sailor (1994) found that students using a computerized integrated learning system (ILS) in both laboratory and classroom settings were more actively engaged in learning tasks than students in the non-ILS classrooms.

Waxman and Huang (1996) similarly found that instruction in classroom settings where technology was not often used tended to be whole-class approaches, in which students generally listened or watched the teacher. Instruction in classroom settings where technology was
moderately used had much less whole-class instruction and much more independent work. Another important finding from the Waxman and Huang (1996) study is that students in classrooms where technology was moderately used (more than 20 percent of the time) were found to be on task significantly more of the time than students from the other two groups—in which technology was infrequently used (less than 10 percent of the time) or in which technology was slightly used (11 percent to 19 percent of the time). These findings are similar to prior studies that found that computer-based instruction increases students’ time-on-task (MacArthur, Haynes, & Malouf, 1986; Schofield & Verban, 1988; Worthen, Van Dusen, & Sailor, 1994). Although these individual studies have examined how technology impacts the teaching and learning process, little is known about how this intervention impacts student outcomes.

**Table 1**

<table>
<thead>
<tr>
<th>Author(s) and Date</th>
<th>Focus</th>
<th>N of Studies</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayraktar (2001-2002)</td>
<td>CAI in secondary and college science</td>
<td>42</td>
<td>.273</td>
</tr>
<tr>
<td>Blok, Oostdam, Otter, and Overmaat (2002)</td>
<td>Computer-based instructional simulations</td>
<td>42</td>
<td>.190</td>
</tr>
<tr>
<td>Cavanaugh (2001)</td>
<td>Interactive distance education technologies</td>
<td>19</td>
<td>.147</td>
</tr>
<tr>
<td>Christmann and Badgett (1999)</td>
<td>CAI in science</td>
<td>11</td>
<td>.266</td>
</tr>
<tr>
<td>Christmann, Badgett, and Lucking (1997)</td>
<td>CAI in differing subject areas</td>
<td>27</td>
<td>.209</td>
</tr>
<tr>
<td>Christmann, Lucking, and Badgett (1997)</td>
<td>CAI in secondary schools</td>
<td>28</td>
<td>.172</td>
</tr>
<tr>
<td>Lou, Abrami, and d’Apollonia (2001)</td>
<td>Small group v. individualized learning with technology</td>
<td>122</td>
<td>.150</td>
</tr>
<tr>
<td>Whitley (1997)</td>
<td>Gender differences in computer-related attitudes and behavior</td>
<td>82</td>
<td>.209</td>
</tr>
</tbody>
</table>

Median = 28 .209
Purpose of the Study

Although there is an adequate knowledge base about the impact of technology on student outcomes, there are still several areas where the decision-making process is hampered due to the scant knowledge base in educational technology. One area in need of a synthesis of the research is examining the effects of teaching and learning with technology on student outcomes. The knowledge base is not consistent as to what type of classroom instruction and instructional setting is most beneficial for teaching and learning with technology in K–12 classrooms.

The purpose of the present study is to synthesize recent research on the effects of teaching and learning with technology on student outcomes. This quantitative synthesis investigates these results by addressing the following questions:

1. How extensive is the empirical evidence on the relationship between teaching and learning with technology and student outcomes?
2. What is the magnitude and direction of the relationship between teaching and learning with technology and student outcomes?
3. Are there certain social contexts or student characteristics that affect the relationship?
4. Are there particular methodological characteristics that affect the relationship?
5. Are there specific characteristics of the technology that affect its relationship with student outcomes?
6. Are there specific characteristics of instructional features that affect technology’s relationship with student outcomes?

To answer these questions, this study quantitatively synthesized experimental and quasi-experimental published research on the effects of teaching and learning with technology on student outcomes in naturalistic settings. The techniques of research synthesis that were applied derive from the work of Glass, McGaw, and Smith (1981) and Hunter, Schmidt, and Jackson (1982) on meta-analysis, as well as contributions from Arthur, Bennett, and Huffcutt (2001), Durlak (1995), and Lipsey and Wilson (2001).
Method

Search and Selection Criteria

A systematic search of research published from 1997 through 2002 investigating the effects of technology on student outcomes was conducted by accessing several sources. For this review, we used selection criteria and review methods that are similar to other recent major national reviews conducted in areas such as teacher preparation (Wilson, Floden, & Ferrini-Mundy, 2001) and reading (National Reading Panel, 2001).

Several criteria were established for inclusion in this synthesis. The synthesis included quantitative, experimental, and quasi-experimental research and evaluation studies that have been published in refereed journals during a five-year period (1997-2002). In order to be included, the study also needed to: (a) focus on teaching and learning with technology in K–12 classroom contexts where students and their teachers interact primarily face-to-face (> 50 percent of the time); (b) be conducted in the United States; (c) compare a technology group to a nontechnology comparison group, or compare the group at the beginning of the intervention (pretest) to a posttest measure; and (d) have reported statistical data (e.g., t tests or F tests) that allowed the calculation of effect sizes.

We identified studies by examining data-base searches, using relevant keywords, and searching the Education Resources Information Center (ERIC). We located additional studies by examining the reference lists of relevant literature reviews and reports. In addition, we specifically examined several major journals in the field of educational technology, such as: Computers & Education, Journal of Educational Technology Research and Development, Computers in the Schools, Journal of Technology Education, Journal of Computers in Math and Science Teaching, International Journal of Instructional Media, Journal of Educational Media, Journal of Education Technology Systems, Journal of Technology Studies, Educational Technology, Learning and Leading With Technology, and Journal of Educational Computing Research. In addition, other education journals such as American Educational Research Journal, Journal of Educational Research, Journal of Educational Psychology, and Elementary School Journal were examined. In addition, several Web sites provided comprehensive lists of technology-based journals with links to journal Web sites. Some were links to specific journals, and others that were only print-based provided a fairly comprehensive index to their journals. Entering the keywords educational technology, evaluation, and instruction and research into a search engine (e.g., Metacrawler and Google) provided a number of other sites (e.g., dmoz.org) that were searched.

Certain types of studies and reports were excluded from the synthesis. Many studies were eliminated because they did not report the appropriate statistics necessary to calculate effect sizes. Some of these studies, for example, provided raw scores for a few “select” participants in the treatment group, but they did not report aggregate scores for both groups (i.e., experimental and control groups). Other studies were eliminated because students in the control group either had access to or used computers. There were many studies, for example, that used research designs where technology was held as a constant and comparisons were made between factors such as differential feedback or instructional approaches. Most of these studies, however, were eliminated because all the students in the control groups had access to and used technology.
The search and selection procedures resulted in a collection of 20 studies. Of these, 14 are published articles from technology journals, three are published articles from education journals, and three are published in refereed conference proceedings.

Procedure

To calibrate the studies’ results, or place them on a common scale, effect sizes were calculated. These effect sizes consist of the treatment group mean minus the control mean divided by the control standard deviation. Effect sizes can be considered a standardized estimate of where the treatment group stands in comparison with the control group distribution. In the case of articles examined for this study, a positive effect size indicated that the instructional technology group received higher (i.e., more desirable) scores than the control group. The formulas of Glass, McGaw, and Smith (1981) were employed for studies that did not report group means or standard deviations but contained $F$ or $t$ values, correlations, or other statistics from which effect sizes could be calculated.

For this synthesis, three investigators recorded 68 codable characteristics and other data for each of the 138 effect sizes from the 20 studies. The 68 categorical variables were employed as factors in an analysis of variance (ANOVA). Each investigator independently coded three studies from each of two investigators. The intercoder agreement for each study reviewed exceeded the 85-percent criterion.

The coding categories are listed in the Appendix. The methodological threats to validly were adapted from Cook and Campbell (1979). Most of the technology characteristics were adapted from other meta-analyses in the area. The teaching variables were adapted from the Five Standards for Effective Pedagogy developed by the Center for Research on Education, Diversity, and Excellence (Dalton, 1998; Tharp, 1997). The five standards are: (1) Teachers and Students Producing Together (Joint Productive Activity), (2) Developing Language and Literacy Across the Curriculum (Language Development), (3) Making Meaning: Connecting School to Students’ Lives (Contextualization), (4) Teaching Complex Thinking (Challenging Activities), and (5) Teaching Through Conversation (Instructional Conversation). These standards are based on the best theoretical and empirical knowledge in the field, and there is ample evidence that their use in classrooms may lead to dramatic improvements for the education of all students (Tharp, Estrada, Dalton, & Yamauchi, 2000).

The studies varied by the number of comparisons they reported. Therefore, those studies with a greater number of comparisons (e.g., those that reported separate results by ability level, sex, or race) would have weighted more heavily than others if each comparison had been given equal weight. To give all studies the same unit weight in the analysis, each comparison was weighted in inverse proportion to the number of comparisons in the study from which it was taken (i.e., $1/n$ where $n =$ number of comparisons in the study). Each of the three comparisons of Michael (2001), for example, received a weight of .333. For studies in which multiple comparisons were made by the percentages of computer use or number of computers, the comparison between the high and low categories were used to calculate the effect size. Most of the studies had multiple outcomes, but the only comparisons included were those that had the appropriate statistics to calculate effect sizes.
Results

The results comprise two sections. The first section summarizes the quantity, type, and quality of studies included in the review. The second section summarizes the overall findings from the studies.

Description of Studies in Review

Initially, a total of nearly 100 potentially applicable articles were retrieved. Upon further application of the criteria for the synthesis, however, only 20 articles were included in the final synthesis. Many of the articles were eliminated because they did not provide the relevant statistics for calculating effect sizes. Other studies were eliminated because students in the control groups had access to or used technology. The final sample of studies included 20 journal articles. A total of 138 effect sizes were calculated from the 20 studies. The studies contained a combined sample of 4,314 students and 6,944 schools. The total number of schools is larger than the total number of students because three studies used the school as the unit of analysis. Most of the studies in the synthesis, however, were smaller-scale studies focusing on a limited number of schools, typically between four and six. The median number of schools in those studies was 76 and the range was from 40 to 6,800. The median number of students in the sample of studies was 65 and the range was from 32 to 2,802. About half of the studies had sample sizes of less than 50, and only 25 percent of the studies had sample sizes greater than 100.

Thirty percent of the studies were published in 1999, 25 percent in 1997, and 20 percent in 2001. The average number of comparisons in each study was approximately 6, but the range was from 1 to 27. About 40 percent of the studies focused on elementary school (Grades K–5), 40 percent on middle-level school (Grades 6–8), and 20 percent on secondary school (Grades 9–12). In terms of research design, about 70 percent of the studies included in the synthesis were quasi-experimental, using either a nonrandomized static-group posttest comparison design (20 percent), a nonrandomized one-group pretest-posttest design (25 percent), or a nonrandomized pretest-posttest control group design (25 percent). Only 20 percent of the studies used an experimental (randomized) pretest-posttest control group design or a randomized posttest-only control group design.

In terms of type of technology, 45 percent of the studies used personal computers, 20 percent used networked laboratories, 10 percent used multimedia, and the other 25 percent used a variety of other technology resources. In terms of instructional software, 35 percent of the studies used an exploratory environment such as simulations, hypermedia, and hypertext. About 20 percent used drill-and-practice software, 10 percent used tools for other tasks such as word processing or e-mail, and 10 percent used mixed forms of technology. About 15 percent of the studies did not specify the software they used.

Evidence for the use of Five Standards for Effective Pedagogy was not very prevalent in the studies reviewed. In 85 percent of the studies, for example, there was no evidence that instructional conversations (extended dialogue between teachers and students) occurred in the classroom. In more than half of the studies, the use of joint productive activities, language and literacy activities, contextualization/making meaning, and challenging activities was not
described. The cognitive outcomes used in the 20 studies varied widely. The most common cognitive outcomes were a researcher-based test (30 percent), followed by standardized tests (20 percent), and a creativity test (20 percent). About 10 percent of the studies used teacher-made tests, and 10 percent used school-level tests. About 85 percent of the affective outcomes were student attitudes towards computers, and 15 percent were students’ motivation. All of the behavioral outcomes examined in the studies in this synthesis focused on the outcome of student attendance.

Overall Results

Table 2a and Table 2b list the mean study-weighted means and the unweighted means for each of the three outcomes and the overall mean. The standard deviations, confidence intervals, and number of comparisons also are included in Table 2a and Table 2b. The mean of the study-weighted effect sizes averaging across all outcomes was .30 \((p < .05)\), with a 95-percent confidence interval (CI) of .004 - .598. This result indicates that teaching and learning with technology has a small, positive, significant \((p < .05)\) effect on student outcomes when compared to traditional instruction.

### Table 2a
**Summary of Mean Study-Weighted Effect Sizes for Student Outcomes**

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Number of Weighted Comparisons</th>
<th>Study-Weighted Effect Sizes</th>
<th>95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>13</td>
<td>.390</td>
<td>.711</td>
</tr>
<tr>
<td>Affective</td>
<td>5</td>
<td>.279</td>
<td>.493</td>
</tr>
<tr>
<td>Behavioral</td>
<td>2</td>
<td>-.154</td>
<td>.402</td>
</tr>
<tr>
<td>Overall</td>
<td>20</td>
<td>.301</td>
<td>.633</td>
</tr>
</tbody>
</table>

### Table 2b
**Summary of Unweighted Effect Sizes for Student Outcomes**

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Number of Unweighted Comparisons</th>
<th>Unweighted Effect Sizes</th>
<th>95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>48</td>
<td>.327</td>
<td>.589</td>
</tr>
<tr>
<td>Affective</td>
<td>60</td>
<td>.208</td>
<td>.397</td>
</tr>
<tr>
<td>Behavioral</td>
<td>30</td>
<td>.096</td>
<td>.235</td>
</tr>
<tr>
<td>Overall</td>
<td>138</td>
<td>.223</td>
<td>.455</td>
</tr>
</tbody>
</table>
In addition to examining the overall mean study-weighted effect size, we also examined the effect sizes for each of the three types of outcomes. The mean study-weighted effect size for the 13 comparisons containing cognitive outcomes was .39, \( (p > .05) \), with a 95-percent confidence interval (CI) of -.050 - .830. This result indicates that teaching and learning with technology has a small, positive effect on students’ cognitive outcomes when compared to traditional instruction. The mean study-weighted effect size for the 60 comparisons that focused on student affective outcomes was .208, \( (p > .05) \), with a 95-percent confidence interval (CI) of -.305 - .862. This result indicates that teaching and learning with technology has a small, positive, significant \( (p > .05) \) effect on students’ affective outcomes when compared to traditional instruction. On the other hand, the mean study-weighted effect size for the 30 comparisons that contained behavioral outcomes was -.154, \( (p > .05) \), with a 95-percent confidence interval (CI) of -.2.503 - .2.195, indicating that technology had a small, negative effect on students’ behavioral outcomes. The mean of the study-weighted effect sizes is .24, suggesting that the average student in the technology-based group scores at the 59th percentile of control-group distributions. The mean study-weighted effect size for the 41 comparisons containing cognitive outcomes was .27, suggesting percentiles on student cognitive outcomes of 61 and 50 for the experimental and control groups, respectively.

The unweighted effect sizes are similar to the study-weighted results. Of the 138 effect sizes that were examined in the 20 studies, about 70 percent were positive. The overall, unweighted effect size was .22, \( (p < .05) \), with a 95-percent confidence interval (CI) of .146 - .229. The unweighted effect sizes was .33 for cognitive outcomes, .21 for affective outcomes, and .10 for behavioral outcomes.

The standard deviations for both the study-weighted and unweighted effect sizes are quite large, indicating a great deal of variation among the studies. One effect size was found to be an outlier because it was more than three standard deviations beyond the overall mean. This outlier was windsorized and recoded to the next highest value. The confidence intervals reported in Table 2a and Table 2b describe the precision of the estimate of the mean effect size by indicating the range within which the population mean is likely to be, given the observed data (Lipsey & Wilson, 2001).

The relationship of each of the 56 conditioning (i.e., independent) variables to the mean study-weighted effect size was tested for significance using ANOVA. The results indicate that none of the variables had a statistically significant \( (p < .01) \) impact on the study-weighted effect size. In other words, the overall findings suggest that the results do not differ significantly across categories of technology, instructional characteristics, methodological rigor, characteristics of the study, and subject characteristics.
Discussion

The results of this quantitative synthesis show a modest, positive effect of teaching and learning with technology on student outcomes. The mean effect size of .30 is quite similar to the median of other recent meta-analyses in the area of instructional technology in education (see Table 1 on page 5). Furthermore, the findings from the present meta-analysis revealed no significant differences across the contextual categories of study quality, teaching, and technology characteristics. In other words, the results can be generalized across a wide variety of conditions that have been investigated as well as across student, school, and study characteristics.

Research Quality Issues

One of the most important issues related to teaching and learning with technology that needs to be addressed is the soundness of the research for the implementation and improvement of technology programs. First, there were few quantitative studies published in the last five years that included relevant data to permit a meta-analysis and calculation of effect sizes. Scientific journals that use independent peer review in deciding what research merits publication are generally considered to be the highest standard of research, yet much of the work in the field of teaching and learning with technology does not meet that standard. The lack of quality, refereed quantitative studies points to a serious problem of research in the field.

Second, there were few studies that used a randomized, experimental design. Only 20 percent of the studies included in the meta-analysis used randomized, experimental designs. Furthermore, it is somewhat surprising that there are still many recent articles published in technology journals that are merely descriptive in nature and just report anecdotes from “selected” teachers or students who enjoy using the technology application. Other published studies explicitly state that their work is “exploratory” in nature, which might explain why they do not report specific findings.

A final concern regarding the quality of research in the field pertains to the lack of details that were included in many of the published articles included in this meta-analysis. Many of the studies lacked the specificity that was needed for us (and potentially others) to code all of the teaching and technology characteristics that we were specifically interested in. About 20 percent of the studies, for example, did not even specify what software was being used in their study. Researchers and journal editors need to make sure that all the relevant details about the classroom processes (e.g., teaching and technology components) are included in articles. Without that explicit information, we will return to the past decades of research on instructional technology, where we were considered to be in a “black box” stage in which we had no idea why instructional technology was effective (Waxman & Bright, 1993).

Limitations of the Present Study

The present meta-analysis, like most others, has several limitations. First, meta-analysis findings are correlational in nature and, therefore, do not warrant strong causal inferences. Second, meta-analysts do not have experimental control over data that reduces the sensitivity of the analysis.
Third, the overall findings from the meta-analysis often are limited by the quality of the primary studies, a problem we have previously discussed.

Another perceived limitation of this meta-analysis may be that we included only published articles in refereed journals. Our justification for doing that is threefold. First, one of the critical scientific principles of educational research is that “scientific studies do not contribute to a larger body of knowledge until they are widely disseminated and subjected to professional scrutiny by peers” (National Research Council, 2002, p. 5). In recent years, a growing number of educators and researchers have become concerned about the quality of work that is posted and disseminated on the Internet. The Committee on Scientific Principles for Education Research (National Research Council, 2002), for example, maintains that the “extent to which the principles of sciences are met in some electronically posted work is often unclear” (p. 72). In this era of evidence-based and scientifically based research, one of the critical characteristics of a study is that it is refereed (i.e., approved for publication) by a panel of independent reviewers (International Reading Association, 2002).

A second explanation of why we excluded Web-based reports is that they often are too broad in nature yet not specific enough to allow meaningful coding. For example, we carefully examined about 20 potential sources from the major national research labs, regional support services, policy institutes, and government institutions. Some of those reports were quantitative studies and even included effect sizes; but they covered, for example, statewide programs over a 10-year intervention period. This report and others clearly do not feature the singularity and clarity of focus that one needs for inclusion in a meta-analysis.

We excluded books, chapters, dissertations, conference proceedings, and technical reports because they are unevenly reviewed. Furthermore, a final explanation of why we excluded nonpublished reports is that there is some evidence that nonpublished Web-based reports in technology have dramatically higher effect sizes than published reports (Niemiec, Sikorski, & Walberg, 1996). Also, there is evidence that many Web-based technical reports are sponsored by agencies that have obvious conflict of interests associated with the results (Wilson, Floden, & Ferrini-Mundy, 2001). For example, in a recent meta-analysis investigating the use of technology to enhance connections between home and school, Penuel et al. (2002) examined the relationship between the researchers who conducted the evaluation studies included in the meta-analysis and the programs they evaluated. They found that in more than half of the studies, the researcher was hired to do the research by either the vendor or school district involved in the study.

A final limitation of our study relates to the recentness of the review. Although we were interested in recent applications of technology, a few of the studies included in the meta-analysis stated that their article was based on projects completed in the early 1990s. In other words, even though we chose only articles that were published within the last five years, some of those articles are still based on technology (i.e., software and hardware) that is nearly a decade old. Future research syntheses may want to include either more rigorous criteria to ensure that only recent technology projects are included or expand the criteria to include older studies in order to examine if there are any secular trends.
Conclusions

The results of this meta-analysis are both discouraging and encouraging. The discouraging aspect is that the overall effects are quite modest, although similar to other recent meta-analyses conducted in the area of instructional technology. In addition, the criticisms of the quality of the research are similar to previous concerns that have been raised by other researchers in the field of instructional technology (Waxman & Bright, 1993).

The aspect of the present study that is encouraging and that may stimulate future research lies in the comprehensive list of variables included in the meta-analysis (see Appendix). This conceptualization suggests that teaching and technology processes either may directly impact student outcomes or may interact with technology features and indirectly impact outcomes. We also believe that the coding procedures effectively captured the essential features of the original research we synthesized. The final list of variables and specific codes included in the Appendix reflects a collaborative process among researchers and practitioners that evolved over time. The high inter-rater agreement we obtained in coding the studies supports our claims of the viability of the process.

There are, of course, many unanswered questions about the effects of teaching and learning with technology on students’ outcomes. We maintain, however, that research can play a critical role in answering some of these questions. Policymakers, however, will need to invest more money on research in technology. The findings from this research synthesis suggest that more and better research needs to be funded and conducted by researchers in this area. Although recognition of the uniqueness of each school and classroom situation will always need to be considered, the accumulation of research evidence over time and across studies may provide consistent findings that enhance our understandings of the role of teaching and learning with technology.
References

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


## APPENDIX

### Information Coded for Each Study

### Study Characteristics

**Author** (Report last name, first; e.g., Doe, John).

**Year of Study** (Report year of study; e.g., 2000).

**Number of Comparisons Within Study** (Report number; e.g., 1 or 2 or 3).

**Student Sex** (Males = 1; Females = 2; Mixed or not specified = 3).

**Grade Level** (Unspecified = 00; 1st grade = 01; Other grades 2–12 use 02 to 12; Mixed primary [K–3] = 24; Mixed middle [4–6] = 25; Mixed upper [7–8] = 26; Mixed high school [9–12] = 27; K–12 = 28).

**Unit of Analysis** (Unspecified = 0; Individual = 1; Class = 2; School = 3; District = 4; State = 5; mixed = 6).

**Student Sample Size** (Report actual sample size; e.g., 4,024).

**School Sample Size** (Report actual sample size; e.g., 4,024).

**Publication Features** (Technology journal = 1; Other educational journal = 2).

**Students’ Ethnicity** (Unspecified = 1; Black = 2; Hispanic = 3; Asian = 4; White = 5; Mixed = 6; Other = 7).

**Students’ Socioeconomic Status** (Unspecified = 0; Lower = 1; Lower middle = 2; Middle = 3; Upper middle = 4; Upper = 5; Mixed = 6).

**Geographical Region** (Northeast = 1; Southeast = 2; Midwest = 3; South Central = 4; Southwest = 5; Northwest = 6; Mixed = 7; Other = 8).

**School Type** (Unspecified = 0; Public = 1; Private = 2; Special school = 3; Mixed = 4; Other = 5).

**Community Type** (Unspecified = 0; Urban = 1; Rural = 2; Suburban = 3; Mixed = 4; Other = 5).

**Content Area** (Content area where technology is used. Unspecified = 0; Reading = 2; Mathematics = 3; Social studies = 4; Science = 5; Reading and math = 6; Language arts = 7; Foreign language = 8; Mixed = 9; Other = 10).

### Quality of Study Indicators

**Method of Observation of Independent Variable** (i.e., technology use. Unspecified = 0; Systematic observation = 1; Informal observation = 2; Student survey or interview = 3;...
Teacher survey or interview = 3; Administrator survey or interview = 4; Computer logs = 5; Multiple methods = 6; Other = 7).

**Pretest Equivalency** (Has the initial differences between the two groups been accounted for?
Unspecified = 0; Statistical Control (e.g., ANCOVA, regression) = 1; Random Assignment = 2; Statistical Control and Random Assignment = 3; Gain Scores = 4; Matching = 5; Other = 6).

**Reported Reliability of Measures** (Unspecified = 00; Actual reliability statistic (e.g., 70 or 83).

**Manner in Which Outcome Scores Are Reported** (Unspecified = 0; Standard scores = 1; Raw scores = 2; Percentile ranks = 3; Gain scores = 4; Other = 5).

**Duration of Study** (Unspecified = 00; List the number of months that the implementation of the technology occurred).

**Cognitive Outcomes** (Unspecified = 0; Testing company standardized achievement test = 1; Federal/national standardized test = 2; State-level achievement test = 3; District-level achievement test = 4; School-level test = 5; Grade-level test = 6; Teacher-made test = 7; Researcher-developed test = 8; Authentic assessment = 9; Creativity test = 10; Higher-level thinking test = 11; Other = 12).

**Affective Outcomes** (Unspecified = 0; Student attitudes toward computers, content areas, anxiety, or instruction = 1; Academic self-concept or motivation = 2; Other = 3).

**Behavioral Outcomes** (Unspecified = 0; Student time-on-task = 1; student perseverance = 2; Tasks attempted = 3; Tasks completed = 3; Success rate = 4; Positive peer interaction = 5; Interactivity with computers = 6; Other = 7).

**Effect Size Coefficient** (actual coefficient)

**Statistics** (Statistics used in determining effect size; Means = 1; t-value = 2; F-value = 3; Chi-square = 4; Other = 5).

**Weight** (One divided by the actual number of comparisons in the study, e.g., 3 comparisons = 1/3 or .333).

**Sources of Invalidity**

**Type of Design** (Quasi-experimental/nonrandomized one group pretest-posttest = 1; Nonrandomized static-group comparison = 2; Nonrandomized pre-post control group = 3; Time series = 4; Randomized posttest-only control group = 5; Randomized pre-post control group = 6; Other = 7).

**History** (Have specific events occurred between the first and second measurement in addition to the experimental variable? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Maturation** (Are there processes within the participants operating as a function of the passage of time [e.g., growing older, more tired] that might account for changes in the dependent
measure? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Testing** (Is there an effect of taking a test upon the scores of a second testing? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Instrumentation** (Do changes in calibration or observers’ scores produce changes in the obtained measurement? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Statistical Regression** (Have groups been selected on the basis of their extreme scores? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Selection Bias** (Have biases resulted in the differential selection of comparison groups? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Mortality** (Has there been a differential loss of participants from the experimental and control groups? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Selection-Maturation Interaction** (Is there an interaction between extraneous factors such as history, maturation, or testing and the specific selection differences that distinguish the experimental and control groups? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Reactive or Interaction Effect of Testing** (Does pretesting influence the participants’ responsiveness to the experimental variable, making the results for a pretested population unrepresentative of the effects of the experimental variable for the unpretested universe from which the participants were selected? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Interaction of Selection Biases and Treatment** (Are there selective factors upon which sampling was based which interact differentially with the experimental variable? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Reactive Effects of Experimental Arrangements** (Are there effects of the experimental setting that would preclude generalizing about the effect of the experimental variable upon persons being exposed to it in nonexperimental settings? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).

**Multiple-Treatment Interference** (Are there nonerasable effects of previous treatments applied to the same participants? Adequately controlled by design = 1; Definite weakness of design = 2; Possible source of concern = 3; Not a relevant factor = 4).
Statistical Power (Is the sample size large enough to reject the null hypothesis at a given level of probability, or are the estimate coefficients within reasonably small margins of error? [a sample > 60 for groups such as classes, schools, or districts; a sample > 100 for individuals]. Probable threat [< 60 for groups or < 100 for individuals as the unit of analysis] = 1; Adequately minimized [> 60 for groups; > 100 for individuals] = 2).

Technology Characteristics

Type of Technology (Unspecified = 0; PCs = 1; Laptops = 2; Networked labs = 3; HP calculators = 4; Multimedia = 5; Other = 6).

Software (Unspecified = 0; Tutorial = 1; Drill-and-practice = 2; Exploratory environment [e.g., simulations, microworlds, hypermedia, and hypertext] = 3; Tools for other tasks [e.g., word processor for writing, e-mail, or computer-conference for course assignments] = 4; Programming language = 5; Other = 6).

Technology Resources/Support Available (Unspecified = 0; No resources = 1; Minimal resources = 2; Adequate resources = 3; Ample resources = 4; Other = 5).

Role/Focus of Technology (Unspecified = 0; Productivity = 1; Delivery system [e.g., ILS] = 2; Resource [e.g., Internet] = 3; Other = 4).

Quantity of Technology (Unspecified = 0; Few [< 3 per classroom] = 1; Average [4–8 per classroom] = 2; Ample [> 9 per classroom] = 3; Other = 4).

Number of Computer Sessions (Unspecified = 0; List number of sessions [e.g., 12]).

Duration of Computer Sessions (Unspecified = 0; List number of average minutes per sessions [e.g., 40]).

Teachers’ Experience with Technology (Unspecified = 0; None = 1; Minimal experience = 2; Average = 3; Experienced = 4; Very experienced = 5).

Students’ Experience with Technology (Unspecified = 0; None = 1; Minimal experience = 2; Average = 3; Experienced = 4; Very experienced = 5).

Teacher Training in Technology (Unspecified = 0; List hours of training [e.g., 15]).

Feedback and Assessment Practices (Unspecified = 0; No feedback = 1; Minimal feedback = 2; Elaborate feedback = 3; Other = 4).

Learning Responsibility (Unspecified = 0; Student controlled = 1; Teacher directed = 2; System directed = 3; Mixed = 4; Other = 5).

Task Difficulty (Unspecified = 0; Difficult = 1; Moderately difficult = 2; Not difficult = 3; Mixed levels of difficulty = 4; Other = 5).

Type of Learning Task (Unspecified = 0; Basic skills/factual learning = 1; Problem solving = 2; Inquiry/investigation = 3; Project-based = 4; Mixed types = 5; Other = 6).
Type of Technology Program (Unspecified = 0; Basic skills/factual learning = 1; Problem solving = 2; Inquiry = 3; Mixed types = 4; Other = 5).

Pattern of Student Computer Use (Unspecified = 0; Teacher use only = 1; Presentation station = 2; One student per computer = 3; Two students per computer = 4; 3–5 students per computer = 4; > 5 students per computer = 6; Mixed pattern = 7; Other = 8).

Percentage of Students Using Computers (Unspecified = 0; > 10% = 1; 10–25% = 2; 26–50% = 3; 51–75% = 4; 76–90% = 5; > 90% = 6).

Objectives of Computer Use (Unspecified = 0; Remediation of skills not learned = 1; Expressing themselves in writing = 2; Communicating electronically with other people = 3; Finding out about ideas and information = 4; Analyzing information = 5; Presenting information to an audience = 6; Improving computer skills = 7; Learning to work collaboratively = 8; Learning to work independently = 9; Multiple Objectives = 10; Other = 11).

Instructional/Teaching Characteristics

Joint Productive Activity/Collaboration (e.g., Designs instructional activities requiring student collaboration to accomplish a joint product; monitors and supports students collaboration in positive ways. No evidence = 1; Some evidence = 2; Extensive evidence = 3).

Language and Literacy Development (e.g., Connects student language with literacy and content-area knowledge through speaking, listening, reading, and writing activities; encourages students to use content vocabulary to express their understanding. No evidence = 1; Some evidence = 2; Extensive evidence = 3).

Contextualization/Making Meaning (e.g., Begins activities with what students already know from home, community, and school; encourages students to use content vocabulary to express their understanding. No evidence = 1; Some evidence = 2; Extensive evidence = 3).

Challenging Activities (e.g., Designs instructional tasks that advance students’ understanding to more complex levels. Assures that students—for each instructional topic—see the whole picture as a basis for understanding the parts. No evidence = 1; Some evidence = 2; Extensive evidence = 3).

Instructional Conversation (e.g., Arranges the classroom to accommodate conversational between the teacher and a small group of students on a regular and frequent basis. Guides conversation to include students’ views, judgments, and rationales using text evidence and other substantive support. No evidence = 1; Some evidence = 2; Extensive evidence = 3).

Setting (Unspecified = 0; Classroom = 1; Networked lab within class = 2; Computer lab in school = 3; Other = 4).
**Mode of Instruction** (Unspecified = 0; Whole-group instruction = 1; Paired = 2; Small-group instruction [3–5 members] = 3; Individualized = 4; Mixed = 5; Other = 6).

**Role of Teacher** (Unspecified = 0; Deliverer of knowledge = 1; Facilitator of groups/student learning = 2; Modeling processes [e.g., problem solving] = 3; Mixed = 4; Other = 5).

**Teacher Qualifications** (Unspecified = 0; Alternatively certified or provisional certificate = 1; Certified in content area = 2; Not certified in content area = 3; Other = 4).

**Policy**

**Level** (Unspecified = 0; School = 1; District = 2; State = 3; Federal = 4; Other = 5).

**Focus** (Unspecified = 0; Reducing achievement gaps = 1; Increased use of technology = 2; Increased specific type of use = 3; Other = 4).