Effects of Educational Technology Applications on Student Achievement for Disadvantaged Students: What Forty Years of Research Tells Us

Alan Cheung

Department of Educational Administration and Policy the Chinese University of Hong Kong, Shatin, Hong Kong

Abstract

The purpose of this review is to examine the effectiveness of educational technology applications and how features of using technology programs and characteristics of these evaluations affect achievement outcomes for disadvantaged students in grades K-12. The review applies consistent inclusion standards to focus on studies that met high methodological standards. A total of 154 qualifying studies were included in the final analysis. The findings of the review suggest that educational technology applications generally produced a positive, though modest, effect (ES=+0.16) in comparison to traditional methods. A marginally significant difference was found among four types of educational technology applications. Larger effect sizes were found with comprehensive models (ES=+0.23) and innovative technology applications (ES=+0.20). Effect sizes for supplemental programs and computer-managed learning were +0.15 and +0.12, respectively. The findings provide some suggestive evidence that approaches that integrated computer and non-computer instruction in the classrooms and innovative approaches are effective in improving student achievement. Differential impacts were also found by both substantive and methodological features.

Keywords: Educational technology applications, reading and math achievement, K-12, meta-analysis

1. Introduction

The use of educational technology in K-12 classrooms has been gaining traction across the country in the past few decades. Substantial investments have steadily increased at the district,
state, and national levels to equip schools with various types of technology, such as computers, internet access, interactive whiteboards and educational software and to integrate technology into curriculum to improve student learning. Based on a recent survey conducted by the U.S. Department of Education (SETDA, 2010) on the use of educational technology in U.S. public schools, almost all public schools had one or more instructional computers with internet access, and the ratio of students to instructional computers with internet access was 3.1 to 1. In addition, almost all public schools had one or more instructional computers located in classrooms and almost 60% of schools had laptops on carts. A majority of public schools surveyed also indicated that they provided various educational technology devices for instruction such as LCD (liquid crystal display) and DLP (digital light processing) projectors, digital cameras, and interactive whiteboards. To support the use of educational technology, the U.S. Department of Education provides grants to state education agencies. For instance, Congress allocated $650 million in educational technology through the Enhancing Education Through Technology (E2T2) program in 2009 (SETDA, 2010). Given the importance of educational technology, particularly for disadvantaged students, it is important to know how best to use technology to enhance student achievement in K-12 classrooms. The purpose of this review is to examine evidence from rigorous evaluations of alternative technology applications to determine which are best supported by the evidence in grade K-12 student achievement, reading and mathematics in particular.

1.1 Working Definition of Educational Technology

Before going any further, it is important to first provide a working definition of the term “educational technology” since it has been used broadly and loosely in the literature. In this review, educational technology refers to a variety of technology-based programs or applications that help deliver learning materials and support learning in K-12 classrooms to improve academic learning goals. Examples include computer-assisted instruction (CAI), integrated learning systems (ILS), and technology-based curricula.

In this review, we identified four major types of educational technology applications: supplemental CAI programs, computer-managed learning systems, innovative applications, and comprehensive models. Supplemental CAI programs, such as Jostens, PLATO, Larson Pre-Algebra, and SRA Drill and Practice, provide additional instruction at students’ assessed levels of need to supplement traditional classroom instruction. Computer-managed learning systems included only Accelerated Reader and Accelerated Math, which use computers to assess students’ reading and mathematics levels, assign materials at appropriate levels, score tests on this material, and chart students’ progress. Innovative technology applications included Fast ForWord and Reading Reels. Fast ForWord supplements traditional CAI with software designed to help children discriminate sounds whereas Reading Reels provides brief, embedded multimedia in whole-class first grade reading instruction to model letter sounds, sound blending, and vocabulary. Comprehensive models, such as Cognitive Tutor, I Can Learn, Read 180, and Voyager Passport use computer-assisted instruction along with non-computer activities (such as teacher-led instruction and cooperative learning) as the students’ core approach to mathematics.
1.2 Previous Reviews of Educational Technology on Student Achievement

Research on educational technology has been abundant. In the past three decades, over twenty major reviews have been conducted in this area (e.g., Bangert-Drowns, Kulik, & Kulik, 1985; Becker, 1992; Christmann & Badgett, 2003; Fletcher-Finn & Gravatt, 1995; Hartley, 1977; C. L. C. Kulik & Kulik, 1991; J. A. Kulik, 2003; Li & Ma, 2010; Ouyang, 1993; Rakes, Valentine, McGatha, & Ronau, 2010; Slavin & Lake, 2008; Slavin, Lake, & Groff, 2009). The majority of these examined a wide range of subjects (e.g., reading, mathematics, social studies, science) and grades from K to 12. The majority of the reviews concluded that there were positive effects of educational technology applications on student achievement, with effect sizes ranging from +0.10 to +0.50. Table 1 presents a summary of student outcomes of these major reviews. The review carried out by Hartley (1977) was perhaps one of the earliest meta-analyses of computer-based technology on student achievement. Its focus was to examine the effects of individually paced instruction in mathematics with four techniques: computer-assisted instruction, cross-age and peer tutoring, individual learning packets, and programmed instruction. Twenty-two studies involving grade 1 to grade 8 students were included in his review. The overall effect size was +0.42.

Becker (1992) conducted a meta-analysis on the effectiveness for academic achievement of computer-based integrated learning systems in the elementary and middle grades. Similar to Hartley (1977), the weighted effect size of the 11 qualified studies was +0.33. Probably the most often-cited review in educational technology was conducted by Kulik and Kulik (1993), which found computers as valuable tools for teaching and learning. In their review, they claimed that educational technology was capable of producing positive effects on student achievement (ES=+0.30). In addition, educational technology also could produce substantial savings in instruction time and foster positive attitudes toward technology.

Several recent reviews (Li & Ma, 2010; Rakes et al, 2010; Slavin & Lake, 2008; and Slavin et al, 2009) also found positive, though smaller effects of educational technology on student achievement. For example, Slavin et al (2008; 2009) included a total of 38 educational technology studies in their elementary review and 38 in a secondary review and found a modest effect size for elementary schools and a small effect size of +0.10 for secondary schools on mathematics achievement. Rakes et al. (2010), reviewed a total of 15 technology-based curricula and yielded a small effect size of +0.16 for student achievement. Similar to Rakes et al. (2010), Li and Ma (2010) reported an average effect size of +0.28 when examining the impact of computer technology on mathematics achievement. Across all 23 reviews, the overall study-weighted effect size was +0.35.

Problems with Previous Reviews

The findings from previous reviews suggest that there was a positive impact of educational technology on student achievement. However, upon closer examination of the included studies in these reviews, especially those in earlier years, many of them suffered from serious methodological problems such as lack of a control group, measures inherent to experimental treatments, limited evidence of initial equivalence between the treatment and control groups, and large pretest differences. For example, the majority of the studies included in Hartley (1977) and Burn (1981) did not have a traditional control group. Over 20% of the effect sizes included in Ouyang (1993) came from pre and post studies with a control group. In addition, many of these reviews included studies that had brief durations. One of the problems with such
brief studies is that they generally produce larger effects than long-duration studies due to novelty factors and a better controlled environment. In addition, brief studies are more likely to use of measures that are closely aligned with the content taught to the experimental group but not to the control group (e.g. Roschelle et al., 2010). To examine the problem with the use of treatment-inherent measures, Slavin & Madden (2011) identified a total of 17 reading and mathematics studies accepted by the What Works Clearinghouse which reported outcomes on both treatment-inherent and treatment-independent measures. For reading, the effect sizes were +0.45 and -0.03, respectively. A similar pattern was found in math. The effect sizes were +0.51 and +0.06, respectively. The review carried out by Kulik (2003) included many studies that were extremely brief, only 2 weeks or less. Another example could be found in Ryan (1991), who reviews 40 studies that evaluated the effectiveness of microcomputer applications in elementary schools. Across several subject areas, Ryan’s overall effect size was +0.31. However, the majority of the studies included (over 70%) had a duration of less than 12 weeks. Furthermore, although claiming to be studies of technology, many confounded use of technology with one-to-one tutoring, small group tutorials, or other teaching strategies known to be effective without technology (e.g. Barker & Torgesen, 1995; Ehri, Dryer, Flugman, & Gross, 2007). Finally, few of these reviews examined how methodological and substantive features affect student achievement. It is against this background that the current review was carried out.

**Rationale for Present Review**

The present review was designed to overcome the major problems of previous meta-analyses by applying rigorous, consistent inclusion criteria to identify high-quality studies. In addition, we examine how substantive and methodological features affect the overall outcome of educational technology applications on student achievement. Two key research questions were addressed:

1. Do education technology applications improve student achievement in K-12 classrooms as compared to traditional teaching methods without educational technology?
2. What study and research features moderate the effects of educational technology applications on student achievement?

**2. Method**

The current review adopted the method of meta-analyses method proposed by Glass, McGaw and Smith (1981) and Lipsey and Wilson (2001). Comprehensive Meta-analysis Software Version 2 (Borenstein, Hedges, Higgins, & Rothstein, 2009) was used to calculate effect sizes and to carry out various meta-analytical tests, such as Q statistics and sensitivity analyses. Like any other meta-analyses, this review followed five key steps: 1) Locate all possible studies; 2) screen potential studies for inclusion using preset criteria; 3) code all qualified studies based on their methodological and substantive features; 4) calculate effect sizes for all qualified studies for further combined analyses; and 5) carry out comprehensive statistical analyses covering both average effects and the relationships between effects and study features.

**2.1 Literature Search Procedures**

All the qualifying studies were from four major sources. References from studies cited in previous reviews provided the first source. The second source was from a comprehensive
literature search of articles written between 1980 and 2012. Electronic searches were made of educational databases (e.g., JSTOR, ERIC, EBSCO, Psych INFO, Dissertation Abstracts), web-based repositories (e.g., Google Scholar), and educational technology publishers’ websites, using different combinations of key words (e.g., educational technology, instructional technology, computer-assisted instruction, interactive whiteboards, multimedia, reading and mathematics interventions, etc.). In addition, we conducted searches by program name. We attempted to contact producers and developers of educational technology programs to check whether they knew of studies that we had missed. We also conducted searches of recent tables of contents of key journals from the past five years: Educational Technology and Society, Computers and Education, American Educational Research Journal, Journal of Educational Research, Journal of Research on Mathematics Education, and Journal of Educational Psychology. Furthermore, we sought papers presented at AERA, SREE, and other conferences. Citations in the articles from these and other current sources were located. Over 1,000 potential studies were generated for preliminary review as a result of the literature search procedures.

2.2 Criteria for Inclusion

To be included in this review, the following inclusion criteria were established. The studies evaluated any type of educational technology applications, including computer-assisted instruction, integrated learning systems, and other technology-based programs, used to improve reading and mathematics achievement. The studies involved students in grades K-12. The studies compared students taught in classes using a given technology-assisted reading and mathematics program to those in control classes using an alternative program or standard methods. Studies could have taken place in any country, but the report had to be available in English. Random assignment or matching with appropriate adjustments for any pretest differences (e.g., analyses of covariance) had to be used. Studies without control groups, such as pre-post comparisons and comparisons to “expected” scores, were excluded.

Pretest data had to be provided, unless studies used random assignment of at least 30 units (individuals, classes, or schools), and there were no indications of initial inequality. Studies with pretest differences of more than 50% of a standard deviation were excluded because, even with analyses of covariance, large pretest differences cannot be adequately controlled for as underlying distributions may be fundamentally different (Shadish, Cook, & Campbell, 2002).

The dependent measures included quantitative measures of student performance, such as standardized reading and mathematics measures. Experimenter-made measures were accepted if they were comprehensive measures of reading and mathematics, which would be fair to the control groups, but measures of reading and mathematics objectives inherent to the program (but unlikely to be emphasized in control groups) were excluded.

A minimum study duration of 12 weeks was required. This requirement is intended to focus the review on practical programs intended for use for the whole year, rather than brief investigations. Studies with brief treatment durations that measured outcomes over periods of more than 12 weeks were included, however, on the basis that if a brief treatment has lasting effects, it should be of interest to educators.
Studies had to have at least two teachers in each treatment group to avoid the confounding of treatment effects with teacher effects.

Programs had to be replicable in realistic school settings. Studies providing experimental classes with extraordinary amounts of assistance that could not be provided in ordinary applications were excluded. Studies had to have taken place from 1980 to 2012. Participants were disadvantaged students.

2.3 Study Coding

In an effort to examine the relationship between effects and the studies’ methodological and substantive features, all studies were coded. Methodological features included research design and sample size. Substantive features included grade levels, types of educational technology programs, program intensity, level of implementation, and socio-economic status. The study features were categorized in the following way:

1. Types of publication: Published or unpublished.
2. Decade of publication: 1980s and before, 1990s, or 2000s and later.
3. Research design: Randomized and matched control.
4. Sample size: Small (N ≤250 students) or large (N>250).
5. Grade level: Elementary (Grade 1-6), or secondary (Grade 7-12).
6. Program types: Computer-managed learning (CML), integrated, innovative or supplemental.
7. Program intensity: Low (≤30 minutes per week), medium (between 30 and 75 minutes per week), or high (>75 minutes per week).
8. Implementation: Low, medium, or high (as rated by study authors).

2.4 Effect Size Calculation and Statistical Analyses

In general, effect sizes were computed as the difference between experimental and control individual student posttests after adjustment for pretests and other covariates, divided by the unadjusted posttest pooled standard deviation. Procedures described by Lipsey and Wilson (2001) and Sedlmeier and Gigerenzer (1989) were used to estimate effect sizes when unadjusted standard deviations were not available, as when the only standard deviation presented was already adjusted for covariates or when only gain score standard deviations were available. If pretest and posttest means and standard deviations were presented but adjusted means were not, effect sizes for pretests were subtracted from effect sizes for posttests. Studies often reported more than one outcome measure. Since these outcome measures were not independent, we produced an overall average effect size for each study. After calculating individual effect sizes for all 154 qualifying studies, Comprehensive meta-analysis software was used to carry out all statistical analyses, such as Q statistics and overall effect sizes.

2.5 Limitations

Before presenting our findings and conclusions, it is important to mention several limitations in this review. First, due to the scope of this review, only studies with quantitative measures of reading and mathematics were included. There is much to be learned from other non-experimental studies, such as qualitative and correlational research, that can add depth and insight to understanding the effects of these educational technology programs. Second, the review focuses on replicable programs used in realistic school settings over periods of at least 12 weeks, but it does not attend to shorter, more theoretically-driven studies that may also provide
useful information, especially to researchers. Finally, the review focuses on traditional measures of reading and math performance, primarily standardized tests. These are useful in assessing the practical outcomes of various programs and are fair to control as well as experimental teachers, who are equally likely to be trying to help their students do well on these assessments. However, the review does not report on experimenter-made measures of content taught in the experimental group but not the control group, although results on such measures may also be of importance to researchers or educators.

3. Findings

A total of 154 qualifying studies were included in our final analysis with a sample size of approximately 120,000 K-12 students. Of these, 39 were published articles and 115 unpublished reports. Of these, 35 were published in the 1980s, 33 in the 1990s, 76 in 2000s, and 10 in the 2010s. In terms of research design, 48 studies used an experimental design, whereas the other 106 were quasi-experiments. There were 71 studies focusing on mathematics achievement and 83 on reading achievement. Almost three-fourths (72%) of the studies (N=111) were with elementary grades and 28% (N=28) of the studies were with secondary grades. The great majority of the participants in these studies were disadvantaged children in urban areas.

As indicated in Table 2, the overall weighted effect size is +0.16. The large Q value indicated that the distribution of effect sizes in this collection of studies is highly heterogeneous ($Q_b=695$, df=153, $p<0.00$). In other words, the variance of study effect sizes is larger than can be explained by simple sampling error. In order to explain this variance, program, design, and methodological characteristics were used to model some of the variation.

3.1 Program Characteristics

Type of intervention. With regards to intervention types, the studies were divided into four major categories: computer-managed learning (CML) (N=14), comprehensive models (N=27), innovation applications (N=6), and supplemental programs (N=107). Approximately 70% of all studies fell into the supplemental CAI program category.

A marginally significant difference ($Q_b=3.29$, df=3, $p<0.09$) was found among types of programs. The effect sizes for comprehensive models, innovative applications, supplemental programs, and CML were +0.23, +0.20, +0.15, and +0.12, respectively. The results of the analyses of innovative applications must be interpreted with caution due to the small number of studies in the category, however.

Program intensity. Program intensity was grouped into three major categories: low intensity (the use of technology less than 30 minutes a week), medium intensity (between 30 and 75 minutes a week), and high intensity (over 75 minutes a week). Analyzing the use of technology as a moderator variable, a statistically significant difference was found between the three intensity categories ($Q_b=7.32$, df=2, $p=0.03$). The effect sizes for low, medium, and high intensity were +0.10, +0.21, and +0.17, respectively. In general, programs that were used more than 30 minutes a week had a bigger effect than those that were used less than 30 minutes a week.

Level of implementation. We also found significant differences among low, medium, and high levels of implementation. It is important to note that almost half of the studies (47%) did
not provide sufficient information about implementation, and levels of program implementation were estimated by the authors. The average effect size of studies with high levels of implementation (ES=+0.23) was significantly greater than those of low implementers (ES=+0.05). The effect size of studies with a medium level of implementation was +0.16. However, the implementation ratings must be considered cautiously because researchers who knew that there were no experimental-control differences may have described poor implementation as the reason, while those with positive effects might be less likely to describe implementation as poor.

**TABLE 2**: Mixed effects moderator analyses examining effect sizes by methodological and substantive features

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>k</th>
<th>d</th>
<th>95% CI</th>
<th>p</th>
<th>Q-value</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall effect size</td>
<td>154</td>
<td>0.16</td>
<td>0.13</td>
<td>0.20</td>
<td>0.00</td>
<td>695.60*</td>
<td>153</td>
</tr>
<tr>
<td>Type of program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.29*</td>
<td>3</td>
<td>0.09</td>
</tr>
<tr>
<td>Computer Managed Learning</td>
<td>14</td>
<td>0.12</td>
<td>0.04</td>
<td>0.21</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innovative</td>
<td>6</td>
<td>0.20</td>
<td>0.10</td>
<td>0.29</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensive</td>
<td>27</td>
<td>0.23</td>
<td>0.11</td>
<td>0.31</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplemental</td>
<td>107</td>
<td>0.15</td>
<td>0.12</td>
<td>0.20</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program intensity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.34*</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td>High (&gt;75min a week)</td>
<td>82</td>
<td>0.17</td>
<td>0.13</td>
<td>0.22</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium (30-75 min a week)</td>
<td>31</td>
<td>0.21</td>
<td>0.13</td>
<td>0.29</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (&lt;30min a week)</td>
<td>39</td>
<td>0.10</td>
<td>0.05</td>
<td>0.15</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26.24*</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>High</td>
<td>23</td>
<td>0.23</td>
<td>0.12</td>
<td>0.33</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>45</td>
<td>0.16</td>
<td>0.10</td>
<td>0.22</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>13</td>
<td>0.05</td>
<td>0.01</td>
<td>0.08</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>73</td>
<td>0.17</td>
<td>0.13</td>
<td>0.22</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.28*</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Elementary</td>
<td>111</td>
<td>0.14</td>
<td>0.11</td>
<td>0.17</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>43</td>
<td>0.22</td>
<td>0.15</td>
<td>0.30</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decade of publication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.72*</td>
<td>3</td>
<td>0.05</td>
</tr>
<tr>
<td>1980s</td>
<td>35</td>
<td>0.21</td>
<td>0.13</td>
<td>0.28</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990s</td>
<td>33</td>
<td>0.11</td>
<td>0.06</td>
<td>0.15</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000s</td>
<td>76</td>
<td>0.14</td>
<td>0.10</td>
<td>0.19</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010s</td>
<td>10</td>
<td>0.29</td>
<td>0.09</td>
<td>0.50</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of publication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.95</td>
<td>1</td>
<td>0.16</td>
</tr>
<tr>
<td>Published</td>
<td>39</td>
<td>0.20</td>
<td>0.14</td>
<td>0.26</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unpublished</td>
<td>115</td>
<td>0.15</td>
<td>0.11</td>
<td>0.19</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.42*</td>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>Randomized</td>
<td>48</td>
<td>0.09</td>
<td>0.05</td>
<td>0.13</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quasi-Experiment</td>
<td>106</td>
<td>0.19</td>
<td>0.15</td>
<td>0.23</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.02*</td>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>Small (N&lt;250)</td>
<td>65</td>
<td>0.25</td>
<td>0.19</td>
<td>0.32</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large (N≥250)</td>
<td>89</td>
<td>0.13</td>
<td>0.10</td>
<td>0.17</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design &amp; size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19.28*</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>Large Randomized</td>
<td>29</td>
<td>0.08</td>
<td>0.03</td>
<td>0.12</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Randomized</td>
<td>18</td>
<td>0.18</td>
<td>0.10</td>
<td>0.27</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large Matched Control</td>
<td>59</td>
<td>0.16</td>
<td>0.11</td>
<td>0.21</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Matched Control</td>
<td>48</td>
<td>0.27</td>
<td>0.19</td>
<td>0.35</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.2 Study Characteristics

Grade level. The effect size for secondary school studies (ES=+0.22) was significantly higher than that for elementary school studies (ES=+0.14) (p<0.05). The direction of the difference is consistent with previous reviews (Ouyang, 1993), suggesting that educational technology had a more positive effect on secondary students than primary students.

Decade of publication. One might expect that the overall effectiveness of educational technology applications would be improving over time as technology becomes more advanced and sophisticated. However, the evidence does not support this expectation. Fletcher-Finn andvGravatt (1995) and Liao (1998) did not find a consistent upward pattern. Christmann and Badgett (2003) found a negative trend over a 14 year time span, with effect sizes dropping from +0.73 in 1969 to +0.36 in 1998. Our present review found no trend toward more positive results in recent years. The mean effect sizes for studies in the 80s, 90s, 2000s, and 2010s were +0.21, +0.11, +0.14, +0.30, respectively.

Type of publication. We used a mixed-effects model to test whether there was a significant difference between published journal articles and unpublished publications, such as conference papers, technical reports, and dissertations. Effect sizes for published articles and unpublished articles were +0.20 and +0.15, respectively. However, the difference was not statistically significant (p<0.16).

In addition, to check whether there was a significant number of studies with null or negative results that have not been uncovered in the literature search which might nullify the effects found in the meta-analysis, classic fail-safe N and Orwin’s fail-safe N analyses were performed. The classic fail-safe N test determined that a total of 4505 studies with null results would be needed in order to nullify the overall effect size estimate of +0.16. The Orwin’s test estimates the number of missing null studies that would be required to bring the mean effect size to a trivial level. We set 0.01 as the trivial value. The result indicated that the number of missing null studies to bring the existing overall mean effect size to 0.01 was 1563. Both tests suggest that publication bias could not account for the significant positive effects observed across all studies.

Furthermore, to avoid the impact of potential outliers that might skew the overall results, a sensitivity analysis was conducted to check for extreme positive as well as negative effect sizes for the overall effect size estimate of +0.16. Using a “one-study removal” analysis (Borenstein et al., 2009), the range of effect sizes still falls within the 95% confidence interval (0.11 to 0.20). In other words, the removal of any one effect size does not substantially affect the overall effect sizes.

3.3 Methodological Characteristics

Research design. One potential source of variation may lie in the research design of the different studies (e.g., Abrami & Bernard, 2006). There were two main types of research designs in this review: randomized experiments and matched control studies. Randomized experiments (N=48) were those in which students, classes, or schools were randomly assigned to conditions...
and the unit of analysis was at the level of the random assignment. Matched control studies (N=106) were ones in which experimental and control groups were matched on key variables at pretesting, before posttests were known. The average effect sizes for randomized experimental studies, and matched control studies were +0.09 and +0.19, respectively. Thus, the mean effect size for quasi-experimental studies was more than twice that for randomized studies.

Sample size. Another potential source of variation may be study sample size (Slavin & Smith, 2008). Previous studies suggest that studies with small sample sizes are likely to produce much larger effect sizes than do large studies (Cheung & Slavin, 2011; Liao, 1999). In this collection of studies, there were a total of 89 large studies with sample sizes greater than 250 and 65 small studies with fewer than 250 students. We found a statistically significant difference between large studies and small studies. The mean effect size for the 65 small-sample studies (ES=+0.25) was about twice that of large-sample studies (ES=+0.13, p<0.00).

Design/size. Within each research design (randomized and matched control), the effect sizes of the small studies were about twice as large as those of the large studies. Large matched control studies produced an effect size of ES=+0.16, as compared to +0.27 for small matched control studies. A similar pattern was also found within the randomized group. Large randomized studies had an effect size of +0.08, whereas small randomized studies had an effect size that was more than twice as large (ES=+0.18).

4. Discussion

The main objective of this review was to examine the overall effectiveness of educational technology applications on student achievement in K-12 classrooms. The findings of this review indicate that educational technology applications produce a positive but modest effect (ES=+0.16) on reading and mathematics achievement. The overall effect size was smaller than that of previous reviews (ES=+0.35). The difference could be explained by two possible reasons. First, many of the previous reviews included studies of marginal quality, which often inflate effect size estimates. In particular, studies using measures inherent to the experimental treatment, very small studies, and matched (rather than randomized) studies all tend to report larger effect sizes. In this review, we applied strict inclusion criteria to select our studies. As a result, many studies included in other reviews were not included in the present review. Second, none of the previous reviews included the two most recent large-scale third-party evaluation reports by Campuzzano et al. (2009) and Dynarski et al. (2007), which found minimal effects of educational technology in middle and high schools on reading and math achievement. Since these two reports contained studies that had large sample sizes, including them has a negative effect on the overall effect size.

Differential impact was found among four types of educational technology applications. Larger effect sizes were found with comprehensive models (+0.23) and innovative technology applications (+0.20). The supplemental program had an effect size of +0.15. Computer-management learning (CML) had the smallest effect size of +0.12. The effect size of CML is similar to that reported in reviews by Kulik et al. (1985) and Niemiec et al. (1987), who also found CML to have a minimal effect on student achievement. The findings provide some suggestive evidence that integrated approaches such as Read 180 and Voyager Passport, which integrate computer and non-computer instruction in the classroom and innovative approaches like Fast ForWord are effective in improving student achievement. Since the majority of these
studies used a quasi-experimental design, more randomized studies are needed to validate the results.

In addition to these overall findings, this review also examined the differential impact of educational technology applications on student achievement by various study and methodological features. It is worth mentioning some of the key findings generated from these variables and how they might impact student outcomes.

First, one might expect that with technological advancement, these newly developed or more sophisticated applications would have an effect on increasing student achievement. However, we failed to find improvements over time in effects of technology on student achievement. Our findings are consistent with some previous reviews (Liao, 1998; Christmann & Badgett, 2003) that also found no positive trend in outcomes for recent studies.

Second, we found a statistically significant difference among the three categories of program intensity. Applications that required computer use of 30 minutes or more had a larger effect than those that required less than 30 minutes a week. The results should come as no surprise. Programs that require 30 minutes or more each week are generally more comprehensive and integrated with the curriculum. For example, in a typical READ 180 classroom, students are provided with a daily 90-minute reading lesson in a group of no more than 15 students. The lesson consists of 20 minutes of whole-class teaching followed by three 20-minute rotation activities in groups of 5, including computer-assisted instruction in reading, modeled or independent reading, and small-group instruction with the teacher. The class then ends with a whole-group wrap-up for 10 minutes. Teachers are given materials and professional development to support instruction in reading strategies, comprehension, word study, and vocabulary (Davidson & Miller, 2002).

Like READ 180, Fast ForWord, published by Scientific Learning, is also a high intensity program, over 90 minutes per day. Fast ForWord is a computerized program designed on the theory that many children with reading and language delays have auditory processing disorders. It uses computer games that slow and magnify acoustic changes within normal speech to “retrain the brain” to process information more effectively. The program was developed by neuroscientists who demonstrated that children using computer games of this type showed improvements in “temporal processing” skills (Merzenich et al., 1996; Tallal et al., 1996). The initial model was expanded into software for use in schools, adding exercises on reading skills such as word recognition, decoding, fluency, spelling, and vocabulary. Children participate in Fast ForWord 90-100 minutes per day, 5 days a week, for 6-8 weeks, so it is intended to make a substantial difference in a relatively short time. Future studies should investigate more closely the impact of the time and integration factors for various grade levels.

Third, a significant difference was found among low, medium, and high levels of implementation. Studies with lower levels of implementation had a significantly smaller effect size than those with higher levels of implementation. It is important to note that about half of the studies did not provide information with regards to implementation. As previously suggested, the results should be interpreted with caution because authors might tend to describe their studies as poorly implemented if they found no significant difference between the treatment and control group. On the other hand, studies with positive results are less likely to describe the program implementation as poor.
Fourth, there is clear evidence to suggest that study outcomes varied as a function of the design variables. We found that both research design and sample size had an impact on the effect sizes. For example, the effect sizes of the quasi-experimental studies (+0.19) were about twice the size of the randomized studies (+0.09). Also studies with small sample sizes produce, on average, twice the effect sizes of those with large sample sizes. Similar results were also found within each research design. The results support the findings of other research studies that made similar comparisons (Slavin et al, 2011; Pearson, Ferdig, Blomeyer, & Moran, 2005; Slavin & Smith, 2008). There are three possible reasons to explain these patterns. For one thing, small-scale studies are often more tightly controlled and have higher implementation fidelity than large-scale studies and, therefore, are more likely to produce bigger effect. In addition, standardized outcome measures are more often used in large scale studies, and these are usually less sensitive to treatments. Finally, the file-drawer effect, in which nonsignificant studies are not published or disseminated, is more likely to apply to small-scale studies with null effects than to large-scale studies. In their review, Niemiec et al. (1987) found that “methodologically weaker studies produced different results than strong studies ... [and] the results of quasi-experimental studies have larger variances.” Unequal variances may produce results that could be potentially unreliable and misleading (Hedges, 1984). The present findings point to an urgent need for more practical randomized studies in the area of educational technology applications for mathematics.

Fifth, in contrast to some previous reviews (e.g. Kulik et al, 1983, 1985 & Lee, 1990) we found no publication bias in this collection of studies. The effect sizes for published articles and unpublished reports were not statistically significant. The results from the Classic fail-safe N and Orwin’s fail-safe N tests confirmed such findings.

Finally, more recent educational technology such as interactive whiteboards, digital games, and hand-held learning response devices has become increasingly popular in classrooms today. Since these technology applications and devices are relatively new, few studies have been conducted to examine the effectiveness of these applications. High quality evaluations are needed for these increasingly popular applications in today’s classrooms.

5. Conclusions

According to a recently released 2011 report by the National Assessment of Educational Progress (NAEP, 2012), only 17% of fourth graders eligible for free lunch scored at proficient or better in reading, while 48% of middle class students scored this well. Among African American students, only 17% scored proficient or better, and the percentages were 18% for Hispanics, compared to 44% for Whites. Similar patterns were found in reading for eighth graders and in math for both fourth and eighth graders. Though all of these scores have been improving over time, the gaps between disadvantaged children and their middle class counterparts remain.

Our findings provides some suggestive evidence that alternative uses of technology applications such as integrated and innovative approaches are effective in improving achievements for disadvantaged children. Schools and districts should make concerted efforts to identify and adopt research-proven education technology program to improve student academic achievement and to close the ability and racial gaps in their schools. In addition,
government and foundation funders should continue to invest in evaluation of these innovative programs and in creation of new technology applications.

Educational technology has been gradually altering K-12 curriculum in the past two decades and will no doubt continue to play an increasingly important role in education in the years to come. So the question is no longer whether teachers should use educational technology or not, but rather how best to integrate educational technology tools into instruction effectively to improve student learning, especially for disadvantaged children.

References


