Effects of Educational Technology Applications on Reading Outcomes for Struggling Readers: A Best-Evidence Synthesis

Alan C.K. Cheung
The Chinese University of Hong Kong, Shatin, New Territories

Robert E. Slavin
Johns Hopkins University, Towson, Maryland, USA

ABSTRACT
This review examines the effectiveness of educational technology applications in improving the reading achievement of struggling readers in elementary schools. The review applies consistent inclusion standards to focus on studies that met high methodological standards. A total of 20 studies based on about 7,000 students in grades 1–6 were included in the final analysis. Findings indicate that educational technology applications produced a positive but small effect on the reading skills of struggling readers (ES = .14) in comparison with “business as usual” methods. Among four types of educational technology applications, small-group integrated applications such as Read, Write & Type and the Lindamood Phoneme Sequence Program produced the largest effect sizes (ES = .32). These are tutorial educational technology applications that use small-group interaction tightly integrated with reading curriculum. Supplementary models, such as Jostens and Lexia, had a larger number of studies (N = 12) and a more modest effect size (ES = .18). Comprehensive models, such as READ 180 and ReadAbout (ES = .04) and Fast ForWord (ES = .06), did not produce meaningful positive effect sizes. However, the results of these two categories of programs should be interpreted with extreme caution due to the small number of studies involved. More studies are required to validate the effectiveness of all technology applications. Policy implications are discussed.

Despite substantial investments in reading instruction over the past two decades, far too many U.S. students remain poor readers, which has profound implications for these children and for the nation. According to the most recent National Assessment of Educational Progress (NAEP; National Center for Education Statistics, 2011), fewer than half of fourth-grade students (42%) scored at or above the proficient level in reading. The results were more troubling for minorities and English learners. Whereas 55% of white students achieved at or above the proficient level on the NAEP, only 19% of African Americans, 21% of Hispanics, and 3% of English learners scored at this level. Similar patterns were found for eighth graders’ NAEP scores.

Students who cannot read well in the early grades tend to be at higher risk of performing poorly in later grades and other subjects, having emotional and behavioral problems, and dropping out of school (Lesnick, George, Smithgall, & Gwynne, 2010). Concerted efforts have been made over the past 20 years among practitioners, researchers, and policymakers to develop policy and identify effective interventions to help struggling readers succeed in reading. For example, approaches such as improved initial teaching of reading, one-on-one tutoring, small-group tutorials, comprehensive school reform, and technology...
applications have been used for struggling readers in many schools across the country.

Among these approaches, educational technology applications have become one of the most popular (Cheung & Slavin, 2012; Kulik, 2003). With more struggling readers being integrated into general classrooms and the increasingly prevalent use of educational technology in today’s classrooms, it is important that teachers, schools, and districts understand the effectiveness of various types of educational technology applications that are available to them to help improve the reading skills of struggling readers. The purpose of this review is to examine the effects of alternative types of educational technology applications for struggling readers, focusing on high-quality, rigorous evaluations.

Theoretical Background

Learning to read is a complex task in which many things must go right for a student to become a successful, strategic, and motivated reader. Students need to recognize letters and sounds, blend the sounds into syllables and words, comprehend the meanings of words by themselves and in sentences and paragraphs, and get the meaning of texts of all types and genres (Adams, 1990). Students must read with sufficient fluency to maintain motivation and efficiency and not lose comprehension because of slow reading (O’Connor, Swanson, & Geraghty, 2010). They need to understand how word parts, or morphology, contribute to word meaning (Bowers, Kirby, & Deacon, 2010). They also need to have sufficient vocabulary and background knowledge to comprehend texts in many genres (Elleman, Lindo, Morphy, & Compton, 2009) and adequate metacognitive skills to extract meanings from text (Block & Duffy, 2008; Pressley, 2003). If anything goes wrong in any of these processes, students will struggle to learn to read (Snow, Burns, & Griffin, 1998).

The National Reading Panel (National Institute of Child Health and Human Development, 2000) identified five elements that should be at the heart of any early reading approach: phonemic awareness, phonics, reading comprehension, vocabulary, and fluency. This list serves as a guide to reading teachers in the early grades for students in general. All students need those skills, and successful readers are likely to develop them all quite rapidly in the primary grades. For struggling readers, however, the story is quite different. Different students may be failing to learn to read adequately for different reasons. One student may recognize every letter and sound but be slow and uncertain in blending them into words. Another may be proficient in reading words but does not comprehend them or the sentences in which they appear. Yet another may lack vocabulary needed to comprehend texts.

Ideally, struggling readers may receive one-on-one tutoring capable of adapting to their unique needs. One-on-one tutoring is indeed very effective for struggling readers in the early grades, especially if tutors use structured, phonetic approaches (D’Agostino & Murphy, 2004; Elbaum, Vaughn, Hughes, & Moody, 2000; Slavin, Lake, Davis, & Madden, 2011). However, tutoring is very expensive and may not be feasible when there are large numbers of struggling readers in a given school.

Technology has often been proposed as a solution for the needs of struggling readers (e.g., Anderson-Inman & Horney, 2007; Boone & Higgins, 2007; Curry, 2003; Roblyer & Doering, 2013; Silver-Pacuilla & Fleischman, 2006; Stetter & Hughes, 2010). In theory, computers can adapt to the individual needs of struggling readers, building on what they can do and filling in gaps. Computers are clearly motivating for most students (Kamil, Intrator, & Kim, 2000; Leu, 2000), and they can mimic some of the behaviors of expert human tutors (Lever-Duffy & McDonald, 2008). Programs that combine computer-assisted instruction (CAI) with teacher instruction and cooperative learning may use these different settings to get the best of each.

The potential of technology applications of all sorts to enhance learning, and specifically to help students who are struggling to learn to read, have been anticipated for many years (Kamil et al., 2000; Lever-Duffy & McDonald, 2008; Roblyer & Doering, 2013). Today, however, enough high-quality research has been done to evaluate various technology applications with struggling readers to begin to reach some conclusions about effective and less effective approaches and to identify promising directions for future development and research. The purpose of this article is to review research on technology applications with struggling readers that meets high standards of methodological rigor, using systematic review procedures appropriate for synthesizing quantitative, experimental research.

Previous Reviews on Educational Technology Applications for Struggling Readers

Although research reviews on reading interventions for struggling readers have been abundant (Boardman et al., 2008; Edmonds et al., 2009; L.A. Hall, 2004; T.E. Hall, Hughes, & Filbert, 2000; Jiendera, Edwards, Sacks, & Jacobson, 2004; MacArthur, Ferretti, Okolo, & Cavalier, 2001; Okolo & Bouck, 2010; Scammacca et al., 2007; Slavin et al., 2011; Stetter & Hughes, 2010; Vaughn et al., 2008; Wanzek, Wexler, Vaughn, & Ciullo, 2010), none of these reviews focused exclusively on the use of educational technology applications to enhance reading achievement for struggling readers in the elementary grades. In addition, many of these reviews included...
students with serious deficiencies, such as a lack of a control group, brief duration, and use of measures that were closely aligned with content taught to experimental but not control treatments.

For example, in their review, Scammacca et al. (2007) examined effective interventions for adolescent struggling readers in grades 4–12. A total of 31 studies were included, and the overall effect size was .05. However, over 60% of the studies included researcher-developed measures that were closely aligned with the treatment. Such measures greatly overstate program outcomes (Slavin & Smith, 2009). The effect size was significantly lower (.46) when studies with these questionable measures were excluded. Jitendra et al. (2004) carried out a review on vocabulary instruction for students with learning disabilities. Overall, results from the six CAI studies were mixed, with an overall effect size of .16. Many studies in this review had very brief durations (a few weeks or less).

A review carried out by Stetter and Hughes (2010) examined the impacts of CAI on reading comprehension for struggling readers. The review covered three main areas: computerized versus printed reading materials, computerized readers to compensate for reading difficulties, and research on a variety of tools. The findings indicated that “some interventions have had at least a somewhat positive effect on student comprehension, while other efforts have shown less positive effects with more limited teacher involvement” (p. 8). Like the two previous reviews, many of the included studies, as acknowledged by the authors, had “a weak or absent comparison group, insufficient information about the sample and outcome measures, as well as small sample sizes that made it difficult to generalize the findings” (p. 8).

The review by Slavin et al. (2011) applied consistent inclusion criteria to focus on studies that met high methodological standards. In the review, the authors identified a total of 97 studies that compared various approaches to helping struggling readers, including one-on-one tutoring, small-group tutorials, classroom process approaches (e.g., cooperative learning), comprehensive school reform, and technology. Fourteen of the 97 studies were evaluations of educational technology applications in reading for elementary and secondary students. The conclusion was that educational technology had a minimal impact on the reading achievement of struggling readers, with an overall sample size-weighted mean effect size of .09 across all studies. Lexia and Jostens were the only two programs that had promising effects. Since the publication of the review, several additional studies meeting high methodological standards have become available.

The purpose of this review is to examine the research up to the present on using educational technology applications to help teach struggling readers in elementary schools. Only studies that met our strict inclusion criteria were included. In addition to the overall effects, we were interested in exploring the differential impacts of moderator variables, such as type of interventions, grade level, program intensity, research design, and recency of educational technology applications. It is important to note that this review does not attempt to determine the unique contribution of technology itself but rather the effectiveness of programs that incorporate use of educational technology. Technological components are often conflated with curricular contents, instructional strategies, and other elements (Clark, 1983, 1985a, 1985b), making it difficult or impossible to identify the unique contributions of the technology.

**Working Definition of Educational Technology**

It is important to define the term *educational technology* because it has been used broadly in the literature. In this meta-analysis, *educational technology* is defined as a variety of electronic tools and applications that help deliver learning content and support the learning process, in this case for elementary struggling readers. Examples include CAI, integrated learning systems, and the use of video or embedded multimedia as components of reading instruction.

In this review, we identified four major types of educational technology applications: traditional supplemental CAI, comprehensive models, small-group integrated supplemental programs, and Fast ForWord (a distinct approach emphasizing teaching of auditory discriminations). Supplemental CAI programs, such as Destination Reading, PLATO Focus, Waterford, and WICAT (World Institute for Computer-Assisted Teaching), provide additional instruction at students’ assessed levels of need to supplement traditional classroom instruction. Comprehensive models, including READ 180 and ReadAbout, use CAI along with noncomputer activities as students’ core reading approach. Small-group integrated models (e.g., Failure Free Reading; Read, Write & Type [RWT]; Lindamood Phoneme Sequence Program [LIPS]), are tutorial educational technology applications that use small-group interaction tightly integrated with the reading curriculum. Fast ForWord supplements traditional CAI with software designed to retrain the brain to process information more effectively through a set of computer games that slow and magnify the acoustic changes in normal speech (Macaruso & Hook, 2001). Each of these applications is discussed in greater detail in later sections.

**Research Questions**

1. Among rigorous studies of technology applications with struggling readers in elementary
schools, what are the average effects on reading achievement measures?

2. How do substantive characteristics of studies, such as the types of programs implemented, grade level, and program intensity, affect the estimated effects?

3. How do methodological characteristics of studies, such as research design and sample size, affect the estimated effects?

Method

The review methods used here are similar to those used by Slavin, Lake, Chambers, Cheung, and Davis (2009), who adapted a technique called best-evidence synthesis (Slavin, 1986). Best-evidence syntheses seek to apply consistent, well-justified standards to identify unbiased, meaningful information from experimental studies, discussing each study in some detail and pooling effect sizes across studies in substantively justified categories. The method is very similar to meta-analysis (Cooper, 1998; Lipsey & Wilson, 2001), adding an emphasis on narrative description of each study's contribution. The narrative descriptions provide context for each study and allow for discussions of methodological and substantive issues, outcomes other than the main ones, and other information that goes beyond what is in the numerical tables.

The quantitative review methods are similar to those used by the What Works Clearinghouse (2012), with a few important exceptions. First, measures (usually made by experimenters) that focused on content taught in the experimental group but not the control group were excluded, as findings from such measures greatly overstate program effects (Slavin et al., 2011). Second, outcomes are pooled for all studies of a given program, weighting for sample size. Studies with small sample sizes also greatly overstate program effects (Slavin & Smith, 2009). (See Slavin, 2008, for an extended discussion and rationale for the procedures used in this series of best-evidence reviews.)

Comprehensive Meta Analysis Version 2 software (Borenstein, Hedges, Higgins, & Rothstein, 2005) was used to calculate effect sizes and to carry out various meta-analytical tests, such as Q statistics and sensitivity analyses. As in most systematic research reviews, this study followed five key steps:

1. Locating all possible studies
2. Using preset criteria to screen potential studies for inclusion
3. Coding all qualified studies based on their methodological and substantive features
4. Calculating effect sizes for all qualified studies for further combined analyses
5. Carrying out comprehensive statistical analyses covering both average effect sizes and the relationships between effect sizes and study features

Literature Search Procedures

In an attempt to locate every study that could possibly meet the inclusion criteria, a literature search of articles written between 1980 and 2012 was carried out. Electronic searches were made of educational databases (e.g., ERIC, PsycINFO, Dissertation Abstracts International), Web-based repositories (e.g., Google Scholar), and educational technology publishers’ websites, using different combinations of keywords (e.g., educational technology, instructional technology, computer-assisted instruction, interactive whiteboards, multimedia, reading interventions). We also 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. References from other reviews of educational technology programs and programs for struggling readers were further investigated. We also conducted searches of recent tables of contents of key reading journals for the past two years (2010–2012): Journal of Educational Technology & Society, Computers & Education, American Educational Research Journal, Reading Research Quarterly, The Journal of Educational Research, Journal of Adolescent and Adult Literacy, Journal of Educational Psychology, and Reading and Writing. Citations in the articles from these and other current sources were located. The searches resulted in the location of over 250 document abstracts that were reviewed independently by two researchers. Approximately 120 articles were retrieved for full-text review.

Criteria for Inclusion

To be included in this review, studies had to meet the following inclusion criteria (see Slavin, 2008, for more detailed rationales).

1. The studies evaluated applications in grades 1–6 incorporating any type of educational technology, including computers, multimedia, and interactive whiteboards.
2. The studies involved students who were having difficulties with learning to read in the elementary grades. These are defined as students with reading disabilities, students in the lowest 33% (or lower) of their classes, or any student receiving tutoring, Title I, special education, or other intensive services to prevent or remediate serious reading problems. Students identified only as low in socioeconomic status or as limited English proficient were not included unless they were also low in reading performance.
3. The studies compared students taught in classes using a given technology-assisted reading program with those in control classes using standard methods.

4. The studies could have taken place in any country, but the report had to be available in English.

5. 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 with expected scores, were excluded. Studies in which students self-selected themselves into treatments (e.g., chose to attend an after-school program) or were specially selected into treatments (e.g., special education programs) were excluded.

6. Pretest data had to be provided unless the 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 because underlying distributions may be fundamentally different (Shadish, Cook, & Campbell, 2002).

7. The dependent measures were quantitative measures of reading performance, such as standardized reading measures. Experimenter-made measures were accepted if they were comprehensive measures of reading, which would be fair to the control groups, but measures of reading objectives inherent to the program (but unlikely to be emphasized in control groups) were excluded. (As noted earlier, studies using measures made by experimenters and aligned with content taught in the experimental but not the control group greatly overstate effects; Slavin et al., 2011.) Measures of skills that do not require interpretation of print, such as phonemic awareness, oral vocabulary, spelling, or writing, were excluded.

8. A minimum study duration of 12 weeks was required to focus the review on practical programs intended for use for the whole year, rather than brief investigations, and to exclude brief experiments likely to use artificial procedures. The 12-week criterion has been used in all previous syntheses carried out by the authors (Cheung & Slavin, 2012; Slavin, Cheung, Groff, & Lake, 2008; Slavin et al., 2011; Slavin, Lake, Chambers, Cheung, & Davis, 2009) and was originally chosen because it represents the length of a school term in the United Kingdom and is similar to a semester in the United States. Brief studies often advantage experimental groups that either use procedures that could not be maintained over a school year or focus on a particular set of objectives during a limited time period while control groups spread that topic over a longer period. 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.

9. The studies had to have at least two teachers in each treatment group to avoid confounding of treatment effects with teacher effect.

10. The studied programs had to be replicable in realistic school settings. Studies providing experimental classes with extraordinary amounts of assistance (e.g., additional staff in each classroom to ensure proper implementation) that could not be provided in ordinary applications were excluded. This criterion was only applied in situations in which it was clear to all raters that the intervention was not intended to be replicable under the conditions in which it was evaluated.

---

**Study Coding**

To examine the relationship between effects and studies’ methodological and substantive features, the studies were coded. Methodological features included research design, sample size, and year of publication. Substantive features included type of education technology application, grade level, and program intensity. The study features were categorized in the following way:

1. *Types of publication:* Published and unpublished
2. *Decade of publication:* 1980s, 1990s, 2000s, and 2010s
3. *Research design:* Randomized design or quasi-experiment
4. *Sample size:* Small ($N < 250$) and large ($N \geq 250$)
5. *Grade level:* Primary (grades 1–3) and upper elementary (grades 4–6)
6. *Program types:* Supplemental CAI, comprehensive models, small-group integrated programs, and Fast ForWord (See Cheung & Slavin, 2012, for details.)
7. *Program intensity:* Low ($\leq 75$ minutes per week) and high ($> 75$ minutes per week). This number of minutes was around the median for CAI interventions. These times included both time students were working with technology and time they were doing other closely associated off-line activities in comprehensive programs (see Cheung & Slavin, 2012, for details).
Both authors examined each potential study independently according to the inclusion criteria and study coding. The inter-rater agreement was 95%. When disagreements arose, the authors reexamined the studies in question together and came to a final agreement.

**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) 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. Because these outcome measures were not independent, we produced a synthetic effect size for each study, defined as the mean effect size in that study. The mean effect size is based on a variance that takes account of the correlation among the different outcomes. We then used this mean effect size and variance to compute a summary effect size across studies using a random-effects procedure, in which each study was weighted by the inverse of its variance (for detail, see Borenstein, Hedges, Higgins, & Rothstein, 2009).

**Results**

**Study Characteristics**

Twenty studies based on a total of about 7,000 students in grades 1–6 met the inclusion standards. The main features and findings of the qualifying studies are summarized in Table 1. Of these, 11 were published articles, and nine were unpublished reports. Only two were published in the 1980s, four in the 1990s, seven in the 2000s, and seven in the 2010s. Thirteen studies used randomized experimental designs, whereas the other seven were quasi-experiments. The program intensity varied from 25 to 450 minutes per week, with a mean of 150 minutes and a standard deviation of 112.

**Overall Effects**

The overall findings, summarized in Table 2, suggest that educational technology applications produced a positive but small effect size (.14) in comparison with traditional methods. Note that if we had used a fixed-effects weighting model, which gives greater weight to large studies, the mean effect size would have been only .08. The large Q value ($Q_B = 38.13, \text{df} = 19, p < .006$) suggests that there is substantial variation in this collective set of studies. Both substantive and methodological variables are used to model some of these variations.

In the following section, we first present the results of substantive and methodological features of the entire set of studies, followed by a more detailed description of each program and study.

**Substantive Features of the Studies**

Three substantive features of the studies were used to model variations in outcomes: types of interventions, grade levels, and program intensity.

**Types of Interventions**

Outcomes varied substantially according to types of interventions. The four studies of small-group integrated applications, Failure Free Reading, RWT, and LIPS produced the largest effect sizes (.32). The 12 studies of supplemental programs, such as Jostens and Lexia, generated an effect size of .18. The mean effect size from the two qualifying studies of comprehensive models, represented by READ 180 and ReadAbout, was .04, and for the Fast ForWord program, two qualifying studies had an average effect size of .06. A narrative description of each study and further discussion of the strength of evidence supporting each type of intervention are presented later.

**Grade Levels**

Studies were organized in two grade levels: primary (grades 1–3) and upper elementary (grades 4–6). Two of the studies examined outcomes across grades but did not provide disaggregated data. Our findings indicate that the mean effect size for primary grades (.36) was much larger than that for upper elementary grades (.07). The mean effect size for the two mixed-grades studies was .25. The between-group difference was marginally significant ($Q_B = 4.66, \text{df} = 2, p < .09$).

**Program Intensity**

Program intensity was grouped into two categories: low intensity (the use of technology interventions, including any associated off-line activities, ≤75 minutes a week) and high intensity (>75 minutes a week). The effect sizes for low- and high-intensity program were .08 and .19, respectively, but the difference was not statistically significant due to low power ($Q_B = 1.20, p < .27$).
### TABLE 1
Educational Technology Applications

<table>
<thead>
<tr>
<th>Study</th>
<th>Design (large or small)</th>
<th>Duration</th>
<th>N</th>
<th>Grade</th>
<th>Sample characteristics</th>
<th>Posttest</th>
<th>Effect size by subgroup/measure</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Supplemental computer-assisted instruction applications</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lexia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Captain’s Log</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
### TABLE 1
Educational Technology Applications (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Design (large or small)</th>
<th>Duration</th>
<th>N</th>
<th>Grade</th>
<th>Sample characteristics</th>
<th>Posttest</th>
<th>Effect size by subgroup/measure</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Destination Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabiner et al. (2010)</td>
<td>Randomized (S)</td>
<td>1 year</td>
<td>50 students (25 E, 25 C)</td>
<td>1</td>
<td>Low-SES students with attention difficulties</td>
<td>DIBELS fluency</td>
<td>.10</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Woodcock-Johnson III reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Thinking Reader</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gates-MacGinitie comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other supplemental computer-assisted instruction applications</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bass, G., Ries, R., &amp; Sharpe, W. (1986). Teaching basic skills through microcomputer assisted instruction. Journal of Educational Computing Research, 2(2), 207–219.</td>
<td>Matched (S)</td>
<td>1 year</td>
<td>2 schools (1 E, 1 C), 145 students (73 E, 72 C)</td>
<td>5, 6</td>
<td>Title I, low-SES students from 16 districts in Connecticut, Massachusetts, Rhode Island</td>
<td>SRA</td>
<td>.22</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Virginia Basic Learning Skills Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Becker (1994)</td>
<td>Randomized (S)</td>
<td>1 year</td>
<td>60 students</td>
<td>2–5</td>
<td>Low achievers in low-SES schools in Baltimore, 50% receiving free or reduced-price lunch</td>
<td>California Achievement Test reading</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Coomes, P. (1985). The effects of computer assisted instruction on the development of reading and language skills. Unpublished doctoral dissertation, North Texas State University, Denton.</td>
<td>Matched (S)</td>
<td>1 year</td>
<td>4 schools, 36 students (18 E, 18 C)</td>
<td>4</td>
<td>Low achievers in middle class schools in Texas, 90% white</td>
<td>Comprehensive Test of Basic Skills</td>
<td>.30</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Design (large or small)</th>
<th>Duration</th>
<th>N</th>
<th>Grade</th>
<th>Sample characteristics</th>
<th>Posttest</th>
<th>Effect size by subgroup/measure</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campuzano, L., Dynarski, M., Agodini, R., &amp; Rall, K. (2009). Effectiveness of reading and mathematics software products: Findings from two student cohorts (NCEE 2009-4041). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education: • Destination Reading • Waterford • Headsprout • PLATO Focus • Academy of Reading</td>
<td>Randomized (L)</td>
<td>1 year</td>
<td>Cohort 2: 232 students (130 E, 102 C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynarski et al. (2007)</td>
<td>Randomized (L)</td>
<td>1 year</td>
<td>Cohort 1: 755 students (410 E, 345 C)</td>
<td>4</td>
<td>National, 64% receiving free or reduced-price lunch, 17% white, 57% African American, 23% Hispanic, lowest third of students</td>
<td>SAT-10</td>
<td>Cohorts 1 and 2</td>
<td>.04</td>
</tr>
<tr>
<td>Campuzano et al. (2009)</td>
<td></td>
<td></td>
<td>Cohort 2: 95 students (52 E, 43 C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Design (large or small)</th>
<th>Duration</th>
<th>N</th>
<th>Grade</th>
<th>Sample characteristics</th>
<th>Posttest</th>
<th>Effect size by subgroup/measure</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comprehensive models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>READ 180</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim, J.S., Samson, J.F., Fitzgerald, R., &amp; Hartry, A. (2010). A randomized experiment of a mixed-methods literacy intervention for struggling readers in grades 4–6: Effects on word reading efficiency, reading comprehension and vocabulary, and oral reading fluency. Reading and Writing, 23(9), 1109–1129.</td>
<td>Randomized (L)</td>
<td>23 weeks</td>
<td>264 students (133 E, 131 C)</td>
<td>4–6</td>
<td>Struggling readers from three high-poverty schools in southeastern Massachusetts</td>
<td>TOWRE (Test of Word Reading Efficiency)</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GRADE (Group Reading Assessment and Diagnostic Evaluation) reading vocabulary and comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DORF Oral Reading Fluency</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SAT–10 reading comprehension</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DIBELS Oral Reading Fluency</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td><strong>ReadAbout</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GRADE</td>
<td>-.02</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 1
Educational Technology Applications (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Design (large or small)</th>
<th>Duration</th>
<th>N</th>
<th>Grade</th>
<th>Sample characteristics</th>
<th>Posttest</th>
<th>Effect size by subgroup/measure</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Small-group integrated supplemental programs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Failure Free Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Torgesen, J., Schirm, A., Castner, L.,</td>
<td>Randomized (S)</td>
<td>1 year</td>
<td>16 schools, 219 students, (113 E, 104 C)</td>
<td>3, 5</td>
<td>Struggling readers in schools around Pittsburgh, 44% receiving free or reduced-price lunch, 80% white, 20% African American</td>
<td></td>
<td>Average of Woodcock–Johnson III, TOWRE, AIMSweb, and GRADE: Grade 3</td>
<td>.19 .05</td>
</tr>
<tr>
<td>Vartivarian, S., Mansfield, W., Myers, D.,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Same: Grade 5</td>
<td>−.05</td>
</tr>
<tr>
<td><strong>Read, Write &amp; Type (small group)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Woodcock–Johnson III word attack</td>
<td>.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Woodcock–Johnson III passage comprehension</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TOWRE nonword</td>
<td>.26</td>
</tr>
</tbody>
</table>
|                                             |                        |          |            |       |                                                            |          | TOWRE word                     | .22              | (continued)
### TABLE 1
Educational Technology Applications (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Design (large or small)</th>
<th>Duration</th>
<th>N</th>
<th>Grade</th>
<th>Sample characteristics</th>
<th>Posttest</th>
<th>Effect size by subgroup/measure</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lindamood Phoneme Sequence Program (small group)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Torgesen et al. (2010)</td>
<td>Randomized (S)</td>
<td>1 year</td>
<td>74 students (35 E, 39 C)</td>
<td>1</td>
<td>Struggling readers in Florida schools</td>
<td>Woodcock–Johnson III word identification</td>
<td>.63</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Woodcock–Johnson III word attack</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Woodcock–Johnson III passage comprehension</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TOWRE nonword</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TOWRE word</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>Fast ForWord</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marion, G.G. (2004). An examination of the relationship between students’ use of the Fast ForWord reading program and their performance on standardized assessments in elementary schools. Unpublished doctoral dissertation, East Tennessee State University, Johnson City.</td>
<td>Matched (S)</td>
<td>1 year</td>
<td>63 students (34 E, 29 C)</td>
<td>5, 6</td>
<td>Lowest 25% of students in schools in Appalachian Tennessee, 52% receiving free or reduced-price lunch, 100% white</td>
<td>TerraNova</td>
<td></td>
<td>.15</td>
</tr>
</tbody>
</table>

Note. C = control. E = experimental. SES = socioeconomic status.
### 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>20</td>
<td>.14</td>
<td>.06</td>
<td>.22</td>
<td>.00</td>
<td></td>
<td>38.13*</td>
</tr>
<tr>
<td><strong>Type of program</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplemental</td>
<td>12</td>
<td>.18</td>
<td>.04</td>
<td>.28</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensive</td>
<td>3</td>
<td>.04</td>
<td>-.09</td>
<td>.17</td>
<td>.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small group</td>
<td>3</td>
<td>.32</td>
<td>-.05</td>
<td>.69</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fast ForWord</td>
<td>2</td>
<td>.06</td>
<td>-.11</td>
<td>.24</td>
<td>.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grade level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginning</td>
<td>8</td>
<td>.36</td>
<td>.11</td>
<td>.60</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper elementary</td>
<td>10</td>
<td>.07</td>
<td>.00</td>
<td>.13</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>2</td>
<td>.25</td>
<td>-.12</td>
<td>.61</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Program intensity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (&gt;75 minutes a week)</td>
<td>13</td>
<td>.19</td>
<td>.06</td>
<td>.28</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (≤75 minutes a week)</td>
<td>7</td>
<td>.08</td>
<td>-.04</td>
<td>.20</td>
<td>.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type of publication</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published</td>
<td>11</td>
<td>.25</td>
<td>.11</td>
<td>.40</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unpublished</td>
<td>9</td>
<td>.04</td>
<td>-.03</td>
<td>.11</td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Year of publication</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980s</td>
<td>2</td>
<td>.20</td>
<td>-.08</td>
<td>.50</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990s</td>
<td>4</td>
<td>.18</td>
<td>.01</td>
<td>.34</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000s</td>
<td>7</td>
<td>.08</td>
<td>-.05</td>
<td>.28</td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010s</td>
<td>7</td>
<td>.22</td>
<td>.08</td>
<td>.36</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experimental design</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized</td>
<td>13</td>
<td>.08</td>
<td>.01</td>
<td>.16</td>
<td>.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quasi-experiment</td>
<td>7</td>
<td>.28</td>
<td>.09</td>
<td>.48</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small (N &lt; 250)</td>
<td>12</td>
<td>.32</td>
<td>.17</td>
<td>.46</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large (N ≥ 250)</td>
<td>8</td>
<td>.04</td>
<td>-.02</td>
<td>.10</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Design and size</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large randomized</td>
<td>7</td>
<td>.03</td>
<td>-.03</td>
<td>.10</td>
<td>.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small randomized</td>
<td>6</td>
<td>.28</td>
<td>.08</td>
<td>.48</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large matched control</td>
<td>1</td>
<td>.12</td>
<td>-.07</td>
<td>.31</td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small matched control</td>
<td>6</td>
<td>.34</td>
<td>.13</td>
<td>.56</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .10, *p < .05.
Methodological Features of the Studies

Sensitivity Analysis
To check whether there are any outliers in this collection of studies that might skew the overall findings, a sensitivity analysis was performed (Borenstein et al., 2009). The analysis indicated that the removal of any one effect size does not substantially affect the overall effect size.

Publication Bias
To assess the possible impact of publication bias, two statistical analyses were performed: classic fail-safe N and Orwin’s fail-safe N. The classic test estimates that to nullify the effect, a total of 121 studies with null results would be needed. Similarly, Orwin’s test indicates that the number of missing null studies to bring the existing overall mean effect size to .01 would be 157. Given the results of these tests, there is no reason to believe that publication bias could account for the positive effect size.

As an additional test of the possibility of publication bias, we used a mixed-effects model to examine whether there was a significant difference between published journal articles and unpublished sources, such as technical reports and dissertations. The overall effect sizes for published articles and unpublished publications were .25 and .04, respectively. The Q-value ($Q_B = 6.47$, $df = 1$, $p < .01$) indicates substantial publication bias in this collection of studies. In other words, the overall effect sizes from the published journal articles were significantly larger than those in unpublished publications, a difference that is very typical in meta-analysis (Lipsey & Wilson, 2001).

Decade of Publication
The results indicated no trend toward more positive results in recent decades. The effect sizes for studies in the 1980s, 1990s, 2000s, and 2010s were .20, .18, .08, and .22, respectively. No statistically significant differences ($Q_B = 1.72$, $p < .63$) were found among different years of publications.

Methodological Features
To understand possible reasons for variations among these studies, we examined methodological features of the studies, such as research design and sample size, to see how they affect reading outcomes.

Research Design
One potential source of variation is the presence of different research designs (e.g., Abrami & Bernard, 2006). In this collection of studies, we identified two main categories of research designs: randomized ($N = 13$) and matched control studies ($N = 7$). Randomized experiments 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 were ones in which experimental and control groups were matched on key variables at pretest, before posttest scores were known. The average effect sizes for randomized experimental studies and matched control studies were .08 and .28, respectively. The mean effect size for quasi-experimental studies was about 3 times the size of that for randomized studies ($p < .06$).

Sample Size
Studies with small sample sizes typically produce much larger effect sizes than do larger studies (Borenstein et al., 2009; Liao, 1999; Slavin & Smith, 2009). In this collection of studies, there were a total of eight large studies with sample sizes of more than 250 students and 12 small studies with fewer than 250 students. A statistically significant difference was found between large studies and small studies ($Q_B = 11.84$, $df = 1$, $p < .00$). The mean effect size for the 12 small studies (.32) was much larger than that of large studies (.04). This suggests the possibility that the small studies create unrealistic conditions of implementation or that small studies with null results are less likely to be reported.

Design and Size
After examining the effect of research design and sample sizes separately, we looked at the combined effect of these two moderator variables. The difference among the four groups was significant ($Q_B = 11.46$, $p < .00$). Small matched control studies produced the largest effect size (.34), followed by small randomized studies (.28), large matched control studies (.12), and large randomized studies (.03). Within each research design, the effect sizes of small studies were much larger than those of large studies.

Narrative Descriptions of Qualifying Studies
To help the reader understand more about various types of interventions included in this collection of studies, we present a narrative description of these applications, along with the context, design, and findings of each study, in the following section.

Supplemental CAI
Supplemental applications of CAI are by far the most common applications of technology in reading. CAI
usually consists of drill, practice, and self-tutorial materials with regular assessments and assignment of students to appropriate materials based on their unique performance levels. Students typically work on CAI in a lab or at the back of the class, usually in two or three 30-minute sessions each week.

**Jostens (earlier version of Compass Learning)**

Jostens is an earlier form of an integrated learning system now called Compass Learning. The system is designed to provide an extensive set of assessments, which place students in an individualized instructional sequence. Students work individually on exercises designed to fill in gaps in their skills. Jostens/Compass Learning integrated learning systems programs are typically used 15–30 minutes per day, two to five days per week. Three qualifying studies examined the effectiveness of Jostens in the 1990s.

The first qualifying study was carried out by Sinkis (1993) to evaluate Jostens with Title I students in a pull-out program in eight schools in a Northeast urban district. Four schools used Jostens, and four served as matched controls. Students in grades 2–6 were involved, but second- and fourth-grade pretests were more than 50% of a standard deviation apart. Among third graders (71 experimental [E], 63 control [C]), MAT Reading Comprehension posttests adjusted for pretests had an effect size of .14 (n.s.). Corresponding effect sizes for fifth graders (83 E, 61 C) were .22 (n.s.), and for sixth graders (74 E, 70 C), the effect size was −.01 (n.s.), for a mean across grades of .12.

Becker (1994) evaluated Jostens with students in grades 2–5 in a high-poverty school in Baltimore, Maryland. A total of 56 low-achieving students were matched and then randomly assigned to use the Jostens integrated learning system in either reading or math. The Jostens group achieved nonsignificantly better scores on the California Achievement Test than did students who did not use the reading software (effect size [ES] = .41).

Another small Jostens study was conducted by Standish (1996) with second graders in two suburban Delaware schools. The Jostens school had four teachers and 56 students, whereas the control school had five teachers and 83 students. The schools were well matched on cognitive ability tests and demographics. On MAT6 Reading posttests, adjusted for cognitive ability tests and demographic variables, the effect size for a Title I subgroup (22 E, 21 C) was .55.

Across the three studies of Jostens, the weighted mean effect size was .19.

**Lexia**

Lexia Learning Systems has two supplemental CAI programs: Phonics Based Reading and Strategies for Older Students. They consist of various activities that teach phonetic word attack strategies to promote automaticity in word recognition. One hundred and sixty students typically participate in two to four 20–30-minute sessions each week. Fararusco, Hook, and McCabe (2006) evaluated the Lexia programs in a yearlong study in 10 first-grade classes in five Boston schools. One class in each school was assigned to the experimental group and another to the control group (n = 83 E, 84 C). Over 50% of all students were eligible for free or reduced-price lunch. After adjusting for initial pretest differences, the mean effect size for Title I students was .67 (p < .02) on the Gates–MacGinitie Reading Test.

**Captain’s Log (BrainTrain) and Destination Reading**

Rabiner, Murray, Skinner, and Malone (2010) carried out a randomized trial to examine the effectiveness of two computer-based interventions for students with attention difficulties: Captain’s Log (BrainTrain) and Destination Reading. Captain’s Log is a commercially available product that provides structured opportunities for exercising attention. Destination Reading is a popular computer-assisted program that targets five key skills: phonemic awareness, phonics, fluency, vocabulary, and comprehension.

Seventy-seven first graders from five low-socioeconomic status (SES) public schools in the southeastern United States were randomly assigned to one of three conditions: Captain’s Log (n = 25), Destination Reading (n = 27), and control (n = 25). Participants were well matched on pretests and demographics. Students in the Captain’s Log group scored higher than did the controls on two reading outcomes measures: DIBELS fluency (ES = .69) and Woodcock–Johnson III reading (ES = .10), with a median effect size of .40. In contrast, the Destination Reading group scored only slightly higher than the controls did: DIBELS fluency (ES = .10) and Woodcock–Johnson III reading (ES = .13), with a median effect size of .12.

**Thinking Reader**

A randomized study was conducted by Drummond et al. (2011) to examine the effectiveness of Thinking Reader, a software program designed to help improve the reading vocabulary and comprehension of students in grades 5–8 using a reciprocal teaching approach (Brown & Palincsar, 1989). Thinking Reader is intended to be integrated with classroom discussion and peer interaction. A total of 2,407 sixth-grade students (1,286 E, 1,121 C) with low-SES backgrounds from 16 school districts in Connecticut, Massachusetts, and Rhode Island participated in a yearlong study. At the end of the study, treatment students in the lowest-achieving group (n = 425)
scored nonsignificantly higher than did their counterparts in the control group \((n = 383)\) on both the Gates–MacGinitie vocabulary and comprehension subtests, with effect sizes of .14 and .13, respectively.

**Other Supplemental CAI**

The two largest randomized studies of supplemental CAI applications were carried out by Dynarski et al. (2007) and Campuzano, Dynarski, Agodini, and Rall (2009). Dynarski et al. evaluated the use in the first-grade classrooms of five CAI reading programs: Destination Reading, Waterford, Headsprout, PLATO Focus, and Academy of Reading. Outcomes for individual programs were not reported, so this is an evaluation of modern uses of technology in first-grade reading in general, not of any particular approach. The study involved 43 schools in 11 districts. A total of 158 teachers (89 E, 69 C) and their 2,619 students (1,516 E, 1,103 C) were randomly assigned within schools to CAI or control conditions.

CAI students used the programs 94 minutes per week, on average. Control classes also often had computers and used them for purposes such as reading assessment and practice, averaging 18 minutes per week. Schools involved in the study were very diverse and located throughout the United States. However, they were relatively disadvantaged, with 49% of students eligible for free or reduced-price lunch and 76% of schools receiving Title I funds. Overall, 44% of students were white, 31% African American, and 22% Hispanic. Students were pre- and posttested on the SAT–9. There were no differences for students in general. \(n\)s for the lowest 33% of students were 505E and 367C. An analysis of effects on the number of students who had posttests below the 33rd percentile found no treatment effects (ES = .02, n.s.).

Dynarski et al. (2007) also evaluated four CAI programs at the fourth-grade level: Leapfrog, READ 180, Academy of Reading, and KnowledgeBox, used an average of 98 minutes per week. Overall, 64% of these students were eligible for free or reduced-price lunch, 57% were African American, 23% were Hispanic, and 17% were white. In all, 118 classrooms (63 E, 55 C) were randomly assigned to treatments, with 2,265 total students (1,231 E, 1,034 C). \(n\)s for the lowest 33% were 410 E and 345 C. On the SAT–10, there were no differences in the proportions of students scoring below the 33rd percentile (ES = −.01).

Campuzano et al. (2009) reported outcomes for a smaller, second cohort of first graders, most of whom were taught by a subset of the same teachers as those in the first cohort, whose outcomes were reported by Dynarski et al. (2007). Four of the five programs remained in use: Destination Reading, Waterford, Headsprout, and PLATO Focus. The numbers of first graders in the lowest third of their classes was 130 E and 102 C. The technology products were used less than half as often in the second year (19.7 hours) as in the first (42.6 hours). Controlling for pretests, the effect size for the proportion of students scoring below the 33rd percentile was −.39. A weighted mean effect size for first graders across the two cohorts was −.07.

Campuzano et al. (2009) also reported second-cohort data for fourth graders taught by a subset of the teachers who taught the first cohort. Two of the four first-cohort programs remained in use: LeapTrack and Academy of Reading. The \(n\)s (of teachers) were 52E and 43C. The programs were used somewhat more often in the second year (16 hours) than in the first (12 hours). Effects on the number of students scoring below the 33rd percentile were nonsignificantly positive (ES = .48). A weighted average effect size for fourth graders across the two cohorts was .04.

**Multiple CAI Programs**

Coomes (1985) evaluated the use of a variety of drill-and-practice software programs (e.g., Fundamental Punctuation Practice, MicroRead, Spelling Program, Word Attack Program). The participants were 112 students from 16 fourth-grade classrooms in four schools in Texas. The software chosen for the study was evaluated to coordinate with the basic fourth-grade curriculum guide and the Macmillan basal series on each reading level. Students in the two treatment schools used the software programs for 30 minutes per week, whereas students in the other two control schools received 30 minutes per week on the computer using mathematics software during the mathematics instructional period. At the end of the fourth-grade year, the students in all four schools were administered the Comprehensive Test of Basic Skills. The effect size for the 36 low achievers \((n = 18\ E, 18\ C)\) was nonsignificant but positive (.30).

In a small study in two Virginia Title I schools, Bass, Ries, and Sharpe (1986) evaluated the use of a variety of software programs (e.g., Alpine Skier, Tank Tactic, Big Door Deal) in grades 5 and 6. Both groups received regular classroom instruction in reading and mathematics. Students in the treatment school using CAI for 25 minutes weekly as part of their Title I instruction \((n = 73)\) were compared with those in a matched school \((n = 72)\) using conventional, supplementary Title I instruction. Students were pre- and posttested on the SRA and the Virginia Basic Learning Skills Test. Averaging fifth- and sixth-grade scores, effect sizes were .22 and .13, respectively, for a median effect size of .18.

**Computer Networking Specialist**

Becker (1994) reported a randomized evaluation of an integrated learning systems program called Computer Networking Specialist, which incorporates a variety of
April 2014 | 401

Comprehensive Models

READ 180

READ 180 is one of the most widely used approaches for adolescent struggling readers. The model is intended to serve as a comprehensive literacy intervention, combining computer and noncomputer instruction in the classroom, with the support of extensive professional development for teachers. 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 five, including CAI 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).

Numerous READ 180 studies have been conducted in the past decade. However, the majority of them were at the secondary level. Slavin et al. (2008) found positive effects for READ 180 in middle schools (with a weighted mean effect size of .24 across eight studies). Two recent randomized studies with struggling readers at the elementary level were included in this review. The first qualifying READ 180 study at the elementary level was carried out by Kim, Samson, Fitzgerald, and Hartry (2010). Approximately 300 fourth to sixth graders who scored below proficiency on the Massachusetts Comprehensive Assessment System test were randomly assigned to either a modified version of READ 180 or to the district’s regular after-school program. To fit the after-school program schedule, the modified READ 180 was shortened to 60 minutes and included only three key components: individualized computer-assisted reading instruction, independent and modeled reading practice with leveled text, and teacher-directed reading lessons tailored to the reading level of small groups but without a teacher-led vocabulary session. Effects were only found on TOWRE (Test of Word Reading Efficiency) reading fluency with fourth graders. No other significant effects were found on other measures or grades, with an overall effect size of .03.

The second qualifying study conducted by Kim and his colleagues (Kim, Capotosto, Hartry, & Fitzgerald, 2011) built on their previous work. Unlike the modified READ 180 version used in the first study, the later study used a READ 180 Enterprise version that was designed to conclude with a teacher-directed whole-group wrap-up lesson to review key objectives. The control group was not given whole-group instruction, individualized computer-assisted reading instruction, or independent and modeled reading practice with leveled text, so it is important to note that the control group spent much less time in reading.

The participants were 312 fourth to sixth graders who scored below proficiency on the Massachusetts Comprehensive Assessment System test from four elementary schools in a midsize urban district in southeastern Massachusetts. Students within each grade and school were randomly assigned to either READ 180 Enterprise or the district’s after-school program. The treatment students outperformed the controls on SAT–10 vocabulary and reading comprehension, with effect sizes of .23 and .31, respectively. No significant differences were found on the DIBELS Oral Reading Fluency scores (ES = .10). The overall effect size across three measures was .21.

ReadAbout

A large-scale randomized study was conducted by James-Burdumy et al. (2009) to evaluate four reading comprehension interventions: Project CRISS, ReadAbout, Read for Real, and Reading for Knowledge. ReadAbout, developed by Scholastic, was the only educational technology program in this study. Students using ReadAbout are taught reading comprehension skills, vocabulary, and content knowledge through an adaptive computer program three times a week for 20 minutes. In addition, students use offline materials once per week for 20 minutes. Offline materials include whole-class or small-group lessons on comprehension skills, vocabulary strategies, text types, or writing skills. Students rotate among computer, teacher-led, and independent reading groups. Teacher materials include suggestions for English learners and differentiated instruction.

Over 2,600 fifth graders from low-SES schools participated in the study. The number of struggling readers (the bottom third of students) were 415 and 456 for the treatment and control groups, respectively. No significant differences were found between the treatment and
control groups after the one-year study period. The overall median effect size across TOSCRF (Test of Silent Contextual Reading Fluency) composite test scores and GRADE (Group Reading Assessment and Diagnostic Evaluation) scores was −0.03. Similar results were found for the other three nontechnology programs. As a group, the combined treatment group scored lower than the control group, with an effect size of −0.08 across the two measures.

Across three studies, the weighted effect size for comprehensive models was near 0 (0.04, n.s.).

**Small-Group Integrated Programs**

**Failure Free Reading**

Torgesen et al. (2007) carried out a large randomized study to examine the effectiveness of four widely used remedial reading instructional programs for struggling readers in 16 elementary schools: Corrective Reading, Failure Free Reading, SpellRead P.A.T. (Phonological Auditory Training), and Wilson Reading. All four interventions delivered instruction to groups of three students pulled out of their regular classroom activities. Failure Free Reading was the only program that had a technology component. It combines computer-based lessons, workbook exercises, and teacher-led instruction to teach sight vocabulary, fluency, and comprehension. In addition, it was the only program that emphasized building students’ vocabulary of sight words rather than phonemic decoding strategies.

The participating schools were first randomly assigned to one of the four interventions. Within each school and grade, students were then randomly assigned to either the treatment or control condition. Within the Failure Free Reading schools, 51 third graders and 62 fifth graders were assigned to receive the treatment, and 38 third graders and 66 fifth graders were in the control groups. Students were pre- and posttested on a battery, including the Woodcock–Johnson III, TOWRE, AIMSweb, and GRADE. The third graders in the treatment group outperformed the controls on all five Woodcock–Johnson reading, AIMSweb, and TOWRE reading outcome measures, with a combined median effect size of 0.36 and 0.66, respectively. There was no significant difference between RWT and LIPS, however.

Across three studies of small-group integrated programs, the overall effect size was 0.32 (p < 0.09). This was by far the largest mean effect size among the four categories of programs, but it is important to note that the larger positive effects were all from very small studies.

**Fast ForWord**

Fast ForWord, published by Scientific Learning, is a computerized program designed around the theory that many students 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 students 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. Students participate in Fast ForWord 90–100 minutes per day, five days a week, for six to eight weeks, so it is intended to make a substantial difference in a relatively short time.

Although many studies of Fast ForWord have been done, most did not qualify for the current review. Most
were too brief (less than 12 weeks), and most used measures of language rather than reading. The most rigorous of the brief studies, an eight-week randomized evaluation by Borman and Rachuba (2009), found no differences between Fast ForWord and control students on reading measures.

The one randomized study of Fast ForWord that met the 12-week duration criterion was an evaluation by Rouse and Krueger (2004), involving four schools in a Northeastern city. All schools were implementing Success for All (Slavin, Madden, Chambers, & Haxby, 2009). About 66% of students were Hispanic and 27% African American, 59% qualified for free or reduced-price lunch, and 61% came from homes in which a language other than English was spoken. Students in grades 3–6 who were in the bottom 20% on the state’s standardized test and had parental permission were randomly assigned to the Fast ForWord (n = 237) or control (n = 217) conditions.

Students in the Fast ForWord group participated in one of two eight-week flights in the spring of 2001. Students in grades 3 and 5 received an average of 35 days of treatment from January to March, and those in grades 4 and 6 received an average of 28 days from March to June. A variety of measures was given just before and just after treatment and thus did not meet the duration requirement of 12 weeks. (They did not show any significant differences on reading outcomes, in any case.) However, the study analyzed state reading test data from fall 2000 and fall 2001. On posttests adjusted for pretests, there were no differences between Fast ForWord and control students (ES = .05, n.s.). Subanalyses of data for students who received the full treatment also showed no differences.

The second qualifying study was conducted by Marion (2004) with a group of fifth- and sixth-grade students in rural Appalachian Grainger County, Tennessee. Almost all of the students were white, and 52% received free or reduced-price lunch. Students who received Fast ForWord (N = 215) were matched with those who did not (N = 134) on TerraNova pretests. On TerraNova posttests, adjusted for pretests, Fast ForWord students in the lowest quartile (34 E, 29 C) scored nonsignificantly higher (ES = .15).

The two studies of Fast ForWord had a weighted mean effect size near 0 (.06, n.s.).

**Discussion**

The purpose of this review was to examine the overall effectiveness of educational technology applications on reading outcomes for struggling readers. We identified a total of 20 high-quality studies that met our inclusion criteria. Our findings indicate that educational technology applications had a small impact on reading achievement of struggling readers, with an overall weighted mean effect size of .14. The effect size is similar to the effect size of .16 reported in a review carried out by Cheung and Slavin (2012) for all students in K–12 classrooms.

Among the four types of educational technology applications, small-group integrated applications such as RWT and LIPS produced the largest effect sizes (.32), but these were mostly small studies, which tend to overstate program impacts. Supplementary models, such as Jostens, had a larger number of studies and a more modest effect size (.16). Comprehensive models (.04) and Fast ForWord (.06) did not produce meaningful positive effect sizes. However, the results of these two categories of programs should be interpreted with extreme caution due to the small number of studies involved.

It is unsurprising that the largest effects were found in the small-group integrated supplemental programs. Previous studies have found that small-group tutorials were effective for struggling readers (Slavin et al., 2011). Unlike traditional supplemental programs, these small-group integrated programs were tightly integrated with the existing curriculum in small-group settings. For example, RWT and LIPS were designed to provide explicit and systematic support of the development of phonemic awareness, phonetic decoding, and text reading accuracy in small-group settings. The findings from these experimental studies provide additional evidence that small-group integrated supplemental programs have a greater impact on reading outcomes for struggling readers than do traditional methods.

In contrast to reviews of applications in secondary schools, comprehensive models such as READ 180 and ReadAbout did not generate meaningful effects on reading outcomes for struggling readers in elementary schools. For example, in their review, Slavin et al. (2008) identified a total of eight qualifying READ 180 studies in middle schools and concluded that there was moderate evidence of a positive impact on reading comprehension (ES = .24). However, the results of the two qualifying READ 180 studies in elementary schools were mixed. The first did not find any statistically significant differences between the treatment and control groups in reading outcomes (ES = .03), whereas the second found more positive results (ES = .21). It is important to point out that unlike the full 90-minute version of READ 180, these two studies used a modified 60-minute version that was used in an after-school setting.

As mentioned earlier, a typical 90-minute READ 180 class includes 20 minutes of whole-class teacher-directed instruction of high-utility words that appear frequently across content areas and three 20-minute instructional activities designed to improve reading efficiency, reading comprehension and vocabulary, and oral reading fluency. In contrast, the modified
60-minute version included individualized computer-assisted reading activities, independent and modeled reading of leveled books, and teacher-directed lessons for small groups of students. The researchers suspected that "the absence of 30-minute whole group instruction may have limited vocabulary gains and the overall efficacy of the READ 180 intervention" (Kim et al., 2010, p. 18). Although the second study also used a similar modified 60-minute version of READ 180, it included regular teacher-directed whole-group instruction and a whole-group wrap-up. The results were more in line with those of previous studies.

The findings of these two studies provide some suggestive evidence that teacher-directed whole-group activities that provide students with systematic and explicit instruction in vocabulary may be beneficial in improving reading comprehension. However, given the small number of studies involved, there is a clear need for more studies evaluating READ 180 in elementary schools.

There is some evidence that technology applications for struggling readers may be more effective with younger students than older students (Kim et al., 2010; Torgesen et al., 2007). The eight qualifying studies that took place in the primary grades had an overall effect size of .36, whereas the 10 studies carried out in the upper elementary grades produced an effect size near 0 (.07). This finding provides some evidence that early intervention is essential for struggling readers.

It appears that high-intensity programs had a bigger impact on struggling readers than did low-intensity programs. The effect sizes for low-intensity and high-intensity programs were .08 and .19, respectively. It is important to note that the majority of these high-intensity programs combined technology and nontechnology components in their reading interventions. For example, RWT and LIPS were designed to provide explicit and systematic support for reading development in the small-group setting. Unlike traditional CAI approaches, these programs were well integrated with classroom instruction. Computer and integrated noncomputer activities were taught about 200 minutes per week or 50 minutes daily. These programs become core, daily activities for students, not supplements, which may account for their apparent effectiveness.

In addition to these overall findings, several other interesting findings emerged from this review and warrant a brief mention. First, 13 of the 20 qualifying studies (65%) used randomized experiments to evaluate program effectiveness. Compared with previous reviews, this percentage of randomized studies is surprisingly high. Importantly, we found a significant difference between experimental and quasi-experimental designs. Effect sizes were generally 3 times larger in quasi-experiments than in randomized experiments.

Consistent with previous findings (Cheung & Slavin, 2012; Pearson, Ferdig, Blomeyer, & Moran, 2005; Slavin & Smith, 2009), small studies in this collection of studies had a much larger effect than did larger ones (ES = .32 and .04, respectively). A few possible reasons could explain these findings. First, it is much easier to maintain high implementation fidelity in small-scale studies as compared with large ones. Second, large-scale studies are more likely to use standardized tests, which are often less sensitive to treatment. Finally, small studies with null effects may have never been written or made available in published or report form, whereas large-scale studies, especially those funded by the government or nonprofit organizations or institutions, are more likely required to make the results, be they positive or negative, available in the public domain as technical reports or in published form.

Limitations

As with any research review, the current review has several limitations. First, 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. Second, the review focuses on traditional measures of reading 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. Finally, despite our efforts to locate every qualifying study, only 20 met our standards, making any conclusions tentative.

Implications for Policy and Practice

The most important practical implication of the review presented here is that there is a limited evidence base for the use of technology applications to enhance the reading performance of struggling readers in elementary schools. Only 20 studies met the inclusion standards, and many of these were small experiments; the larger studies, especially those that used random assignment to conditions, reported the smallest effects. Among eight large, randomized evaluations, the weighted mean effect size was essentially 0 (.04, n.s.).

Within the existing literature, however, the largest effect sizes were found for small-group interventions that supplement first-grade instruction with phonetic activities integrating computer and noncomputer activities and occupying substantial time each week. Among
currently available models, these include the LIPS and RWT. Among more traditional supplemental CAI models, there is supportive evidence for Lexia, also used in first grade. However, each of these is supported by a single small study, so none can be confidently prescribed as a broadly applicable solution for struggling readers.

Further, the effect sizes found for the various technology applications for struggling first graders are, at best, similar to those found for similar phonics-focused small-group interventions that do not use technology, and are much less than those associated with phonetic or small-group one-on-one instruction and comprehensive school reform models (see Slavin et al., 2011). For upper elementary students, none of the technology applications had notable impacts, whereas Slavin et al. reported substantial positive effects in grades 3–6 for several whole-class interventions, such as cooperative learning, as well as other nontechnology interventions targeted at struggling readers.

None of this is meant to imply that technology applications do not have a role in improving outcomes for struggling readers, especially first graders. Three studies that directly compared small-group and individualized tutoring with and without well-integrated technology all found that the use of the technology enhanced reading outcomes for struggling first graders. What it does suggest, however, is that there is no magic in the machine. What determines the effectiveness of technology applications for struggling readers is the nature of the software, the role of the teacher, the nature and quality of professional development and follow-up, the amount of time devoted to the technology and nontechnology parts of each approach, how these activities are placed in students’ days and weeks, what activities they replace, and much more. If anyone still imagines that computers will make a difference if they merely arrive in a box, ready to plug in and play with minimal professional development and follow-up, the findings reported here should be sobering.

There is no question that technology will be part of future solutions to the problems of reading difficulties. With further research, many of the programs reviewed here could build a stronger evidence base, and the best of them should serve as a basis for further development of impactful models. Approaches using technologies now becoming commonplace in elementary schools, such as interactive whiteboards, electronic response devices, and laptops or other devices for all students, have not yet been adequately researched for struggling readers but could hold promise. As computers and other electronic devices become ubiquitous in students’ homes, additional possibilities arise in integrating home and school activities. New applications of embedded multimedia, using bits of video to enhance teachers’ lessons, also have promise.

The evidence to date shows promise for some types of technology applications, but much more remains to be done both in research and in the development of more effective solutions. The problems of reading failure in elementary schools are important, and they justify continued efforts to create and validate reliably effective approaches combining the best efforts of teachers and technology.

NOTE

1 A study was defined as a unique comparison of experimental and control treatments. Two articles reported on more than one treatment–control comparison.

REFERENCES

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


Submitted November 26, 2012
Final revision received April 8, 2013
Accepted April 11, 2013

ALAN C.K. CHEUNG is an associate professor in the Department of Educational Administration and Policy at The Chinese University of Hong Kong, Shatin, New Territories; e-mail alancheung@cuhk.edu.hk (corresponding author).

ROBERT E. SLAVIN is the director of the Center for Research and Reform in Education at Johns Hopkins University in Towson, Maryland, USA; e-mail rslavin@jhu.edu.