Effective Reading Programs for Middle and High Schools: A Best-Evidence Synthesis

Robert E. Slavin
Johns Hopkins University, Baltimore, MD, USA
University of York, England

Alan Cheung
Hong Kong Institute of Education

Cynthia Groff
University of Pennsylvania, Philadelphia, USA

Cynthia Lake
Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

This article systematically reviews research on the achievement outcomes of four types of approaches to improving the reading of middle and high school students: (1) reading curricula, (2) mixed-method models (methods that combine large- and small-group instruction with computer activities), (3) computer-assisted instruction, and (4) instructional-process programs (methods that focus on providing teachers with extensive professional development to implement specific instructional methods). Criteria for inclusion in the study were use of randomized or matched control groups, a study duration of at least 12 weeks, and valid achievement measures that were independent of the experimental treatments. A total of 33 studies met these criteria. The review concludes that programs designed to change daily teaching practices have substantially greater research support than those focused on curriculum or technology alone. Positive achievement effects were found for instructional-process programs, especially for those involving cooperative learning, and for mixed-method programs. The effective approaches provided extensive professional development and significantly affected teaching practices. In contrast, no studies of reading curricula met the inclusion criteria, and the effects of supplementary computer-assisted instruction were small.

Students who enter high school with poor literacy skills face long odds against graduating and going on to postsecondary education or satisfying careers. Joftus and Maddox-Dolan (2003) reported that in the United States, roughly 6 million secondary students read far below grade level and that approximately 3,000 students drop out of U.S. high schools every day. The secondary years provide a last chance for many students to build sufficient reading skills to succeed in their demanding courses (Biancarosa & Snow, 2006; Joftus, 2002).

Even among students who do graduate from high school, inadequate reading skills are a key impediment to success in postsecondary education (American Diploma Project, 2004). Students who struggle with reading often lack the prerequisites to take academically challenging coursework that could lead to more wide reading and thus exposure to advanced vocabulary and content ideas (Au, 2000). The 2006 report by ACT, Inc., *Reading Between the Lines: What the ACT Reveals About College Readiness in Reading*, describes even more troubling

trends. Only 51% of students who took the ACT test in 2004 were ready for college-level reading demands (ACT, Inc., 2006).

Students who read at low levels often have difficulty understanding the increasingly complex narrative and expository texts that they encounter in high school and beyond. For example, one of the major hurdles in acquiring science literacy is the conceptual density of math and science materials (Barton, Heidema, & Jordan, 2002). Students' performance on these more difficult texts, which include context-dependent vocabulary, concept development, and graphical information, provides the strongest indication as to whether or not they are prepared to succeed in college and the workplace (ACT, Inc., 2006). Clearly, well-evaluated programs capable of enabling middle and high school students with poor reading skills to meet the demands of complex texts are needed to ensure that these students not only succeed in their high school coursework but also graduate ready for college and work-related reading tasks.

Due in large part to accountability programs focusing on reading, U.S. schools are increasingly providing instruction in reading to a large proportion of middle and high school students (Deshler, Palincsar, Biancarosa, & Nair, 2007). Once seen only in remedial or special education programs, reading courses are now common in middle schools, and remedial reading courses are becoming more widespread in high schools. Yet, there is little understanding of which particular programs are likely to be effective in middle and high schools. Remarkably, a systematic, comprehensive review of the research on middle and high school reading programs has never been done. The federal What Works Clearinghouse (2007) has completed a review of research on elementary school reading programs but does not even have a review of research on secondary reading programs in its long-term plans. Published by Deshler et al. (2007), *Informed Choices for Struggling Adolescent Readers: A Research-Based Guide to Instructional Programs and Practices* contains brief discussions of the research evidence supporting each of 48 widely used programs for adolescent readers, as well as lists of articles about each program; however, it does not attempt to synthesize or compare the evidence for these programs.

The purpose of the present article is to review research on middle and high school reading programs, applying consistent methodological standards. This review is intended both to provide fair comparisons among the achievement effects of the full range of approaches available to educators and policymakers and to summarize the current state of the art in secondary reading programs. The scope of the review comprises all of the types of programs that teachers, principals, and superintendents might consider as a means of solving their secondary students' reading problems.

The present review uses a form of best-evidence synthesis (Slavin, 1986) that has been adapted for use in reviews of "what works" literatures where there are usually only a few studies evaluating each of many programs (see Slavin, 2008). Similar methods have been used to review research on elementary math programs (Slavin & Lake, in press), middle and high school math programs (Slavin, Lake, & Groff, 2007), and reading programs for English-language learners (ELLs; Cheung & Slavin, 2005).

Even though the two math reviews (Slavin & Lake, in press; Slavin et al., 2007) involved a subject other than reading, they provide important background for the current review. In the case of both of these previous reviews, median effect sizes across many qualifying studies were quite low for math curricula as diverse as the constructivist programs funded by the National Science Foundation (e.g., Everyday Mathematics) and the algorithmic Saxon Math. Median effect sizes for studies evaluating innovative math curricula were +0.05 for elementary school studies and +0.07 for middle and high school studies. Both reviews found larger but still modest effects for computer-assisted instruction (CAI) programs such as Jostens and SuccessMaker. Median effect sizes for these programs were +0.19 for elementary school studies and +0.16 for middle and high school studies. The largest effects were for instructional-process programs such as cooperative learning and classroom motivation and management programs and other approaches that focused on changing teacher and student behaviors during daily lessons. For example, median effect sizes for cooperative learning programs were +0.29 for elementary school studies and +0.32 for middle and high school studies. Studies of these instructional-process programs were also more likely to have used random assignment to treatments.

The Cheung and Slavin (2005) review of research on (mostly elementary school) studies of reading programs for ELLs also found that the most effective programs were those that emphasized professional development and changed classroom practices, such as cooperative learning and comprehensive school reform. Recognizing that reading is not the same as math and that secondary reading is not the same as reading at the elementary level, we nevertheless hypothesized that secondary reading programs focused on reforming daily instruction would have stronger impacts on student achievement than would programs focused on innovative curricula or CAI alone.

Focus of the Current Review

Using procedures similar to those employed in the previously discussed math reviews, the present review examines research on reading programs designed for use in middle and high schools with students in grades 6–12.

(Data on sixth graders appear in the current review if the middle school included this grade.) The purpose of this review is to place results of all types of programs intended to enhance the reading achievement of middle and high school students on a common scale and to provide educators and policymakers with meaningful, unbiased information that they can use to select programs most likely to make a difference with their students. To maximize the usefulness of the review for educators, it emphasizes practical programs that are or could be used at scale. The review therefore focuses on large studies that were completed over significant periods of time and that used standard measures.

This review also seeks to identify common characteristics of programs likely to make a difference in student reading achievement. Intended to include all kinds of approaches to reading instruction, the review groups these approaches into four categories: (1) reading curricula, (2) mixed-method models, (3) CAI, and (4) instructional-process programs. The reading-curricula category primarily encompasses innovative textbooks and curricula such as McDougal Littell and LANGUAGE! Mixed-method models, represented in the review by READ 180 and Voyager Passport, are those that combine large- and small-group instruction, computer activities, and other elements to create a complete instructional approach. CAI refers to programs that use technology to enhance reading achievement. CAI programs are usually supplementary, as when students are sent to computer labs for additional practice. A related category is computer-managed instruction, represented in the review by Accelerated Reader, which uses computers to assign readings and assess progress. CAI is the one category of secondary reading programs that has been reviewed in the past. A few secondary reading studies were included in reviews by Kulik (2003), Murphy et al. (2002), and Chambers (2003). The fourth category, instructional-process programs, is the most diverse. All programs in this category rely primarily on professional development to give teachers effective strategies for teaching reading. These include programs that focus on cooperative learning and strategy instruction. Comprehensive school reform programs were included in the present review only if they involved specific middle or high school reading programs. (For a broader review of outcomes of secondary comprehensive school reform models, see Comprehensive School Reform Quality Center, 2006, and Borman, Hewes, Overman, & Brown, 2003.)

Review Methods

The methods used in the current review are similar to those used by Slavin and Lake (in press) and Slavin et al. (2007), 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); however, it also includes a narrative description of each study's contribution. In addition, the methods used in best-evidence syntheses are very similar to the methods used by the What Works Clearinghouse (2007), although a few exceptions are noted in the following sections. (For an extended discussion of and rationale for the methods used in best-evidence syntheses, see Slavin, 2008.)

Criteria for Inclusion

Criteria for inclusion of studies in this review were as follows:

1. Studies had to have evaluated reading programs for middle and high schools. Studies of variables, such as the use of ability grouping, block scheduling, or single-sex classrooms, were not reviewed.

2. Studies had to have involved middle and/or high school students in grades 7–12. Studies involving middle schools that began at grade 6 could also be included.

3. Studies had to have compared children in classes using a given reading program to those in control classes using an alternative program or standard methods.

4. Studies could have taken place in any country, but the report of the study had to be available in English.

5. Studies had to have used random assignment or matching with appropriate adjustments for any pretest differences (e.g., analyses of covariance). Studies without control groups, such as pre–post comparisons and comparisons to expected scores, were excluded. Studies in which students had selected themselves into treatments (e.g., chose to attend an after-school program) or had been selected into treatments by others (e.g., gifted or special education programs) were excluded unless experimental and control groups had been designated after selections were made.

6. Studies had to have provided pretest data, unless random assignment of at least 30 units (individuals, classes, or schools) had been used and no indications of initial inequality had been found. Studies with pretest differences of more than 50% of a standard deviation were excluded. This was done because when underlying distributions are fundamentally different, even analyses of covari-

ance cannot adequately control for large pretest differences (Shadish, Cook, & Campbell, 2002).

7. Studies' dependent measures had to have included quantitative measures of reading performance such as standardized reading measures. Studies involving experimenter-made measures were accepted if there were comprehensive measures of reading that would have been fair to control groups. However, studies involving measures of reading objectives that were inherent to the program (but unlikely to be emphasized in control groups) were excluded. The exclusion of studies with measures inherent to the experimental treatment is a key difference between the procedures used in the present review and those used by the What Works Clearinghouse (2007).

8. Studies had to have had a minimum duration of 12 weeks. This requirement was intended to focus the review on practical programs designed for use throughout an entire year, rather than brief investigations. On the one hand, studies of shorter duration may not allow programs to show their full effect. On the other hand, these studies often advantage experimental groups that focus on a particular set of objectives for a limited time period when compared with control groups that engage with these same objectives less intensely and over a longer period of time. 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. The 12-week criterion has been consistently used in all of the systematic reviews previously completed by the authors of the present review (i.e., Cheung & Slavin, 2005; Slavin & Lake, in press).

9. Studies had to have had at least two teachers and 15 students in each treatment group.

The Appendix lists those studies that were considered germane but that were excluded from the current review according to the criteria for inclusion. The Appendix also gives the reason for each study's exclusion. One of the reasons provided is "no adequate control group," which means that although there was some sort of counterfactual, it did not meet the standards of the review because the control group either was not well matched, studied different content, or did not use standard practices. Another reason given is "inadequate outcome measure." These are nonstandard, experimenter-made measures of unknown validity that were judged to be slanted toward content taught in the experimental but not the control classes.

Literature Search Procedures

A broad literature search was carried out in an attempt to locate every study that might possibly meet the inclusion requirements. Electronic searches were conducted of educational databases (JSTOR, ERIC [Education Resources Information Center], EBSCO, PsycINFO, and Dissertation Abstracts International) using different combinations of key words (e.g., "secondary students," "reading," and "achievement"). Search results were limited to studies published between 1970 and 2007. Results were then narrowed by subject area (e.g, "reading intervention," "educational software," "academic achievement," and "instructional strategies"). In addition to searching for studies using key terms and subject areas, we conducted searches by program name. We also looked for studies using Internet search engines, examined the websites of educational publishers, and attempted to contact producers and developers of reading programs to find out whether they knew of studies that we had missed. Further, we investigated citations from previous reviews of research on reading programs (e.g., Deshler et al., 2007) and other potentially related topics such as technology (Chambers, 2003; Murphy et al., 2002). We also searched the following journals' tables of contents from 2000 to 2007 to locate additional citations: *American Educational Research Journal*, *Reading Research Quarterly*, *Journal of Educational Research*, *Journal of Adolescent & Adult Literacy*, *Journal of Educational Psychology*, and *Reading & Writing Quarterly*. The citations appearing in those studies found during the first wave of searches were investigated as well. Unlike the What Works Clearinghouse, which excluded studies that were more than 20 years old, studies meeting the selection criteria were included in the current review if they were published from 1970 to the present. This enabled us to include a few high-quality studies completed in the 1970s and the early 1980s that are of direct relevance to today's schools.

Effect Sizes

In general, effect sizes were computed as the difference between the posttest scores for individual students in the experimental and control groups after adjustment for pretests and other covariates, divided by the unadjusted standard deviation of the control group's posttest scores. If a standard deviation was not available for the control group, then a pooled standard deviation was used. 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. This occurred 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, then the effect sizes for pretests were subtracted from the effect sizes for posttests.

Effect sizes were pooled across studies for each program and for various categories of programs. This pooling used means weighted by the final sample sizes. The use of weighted means was the only important methodological difference between the present review and those previously completed by Slavin and Lake (in press) and Slavin et al. (2007), which used medians to pool effect sizes. Weighted means were used to maximize the importance of large studies since these earlier reviews, among many others, found that small studies tend to overstate effect sizes (see Rothstein, Sutton, & Borenstein, 2005; Slavin, 2008). A cap weight of 2,500 students was used to avoid having studies that were very large dominate the means.

Limitations

It is important to note several limitations of the current review. First, the review focuses on quantitative measures of reading. There is much to be learned from qualitative and correlational research, which can provide new insights about and deepen our understanding of the effects of secondary reading programs. Second, the review focuses on replicable programs used in school settings over periods of at least 12 weeks. This emphasis is consistent with the review's purpose in providing educators with useful information about the strength of evidence supporting various practical programs; however, the review 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 performance, primarily standardized tests. In addition to being useful in assessing the practical outcomes of various programs, these measures are fair to both control and experimental classes where the teachers are equally likely to be trying to help their students perform well on the assessments. However, the review does not report on experimenter-made measures of content that was taught in the experimental group but not in the control group, although the results from such measures may also be of importance to researchers and/or educators.

Categories of Research Design

Four categories of research design were identified. Randomized experiments were those in which students, classes, or schools were randomly assigned to treatments, and data analyses were at the level of random assignment. When schools or classes were randomly assigned but there were too few schools or classes to justify analysis at the level of random assignment, the study was categorized as a randomized quasi-experiment (Slavin, 2008). Several studies claimed to use random assignment because students were assigned to classes by a computerized scheduling system, but scheduling constraints (such as conflicts with advanced or remedial courses taught during the same period) can greatly affect such assignments. In addition, routine scheduling done by school officials often changes students' schedules after initial assignments have been made by a computerized scheduling system. Studies using computerized scheduling systems or other random-appearing assignment methods under the control of school administrators were categorized as matched, not random. Matched studies were those in which experimental and control groups were matched on key variables at pretest, before posttests were known, while matched post-hoc studies were those in which groups were matched retrospectively, after posttests were known. For reasons described by Slavin (2008), studies using fully randomized designs are preferable to randomized quasi-experiments, but all randomized experiments are less subject to bias than matched studies. Among matched designs, we gave preference to prospective designs over post-hoc or retrospective designs. In the subsequent descriptions of the studies under review and in the accompanying tables, studies of each type of program are addressed according to their research design in the following order: (1) randomized experiment, (2) randomized quasi-experiment, (3) matched, and (4) matched post-hoc. Within these categories of research design, studies with larger sample sizes are described first. Therefore, studies discussed earlier in each descriptive section should be given greater weight than those that appear later, all other things being equal.

Results

Reading Curricula

No studies of secondary reading curricula met the criteria for this review. This is surprising in light of the widespread use of such programs in middle and high schools throughout North America. It is not the case that the inclusion standards applied in the present review excluded many studies. Despite an extensive search, only 14 studies of reading curricula were located (see the Appendix). No studies were found, for example, of McDougal Littell, and only two studies of LANGUAGE! were retrieved, neither of which had control groups. Corrective Reading was the only textbook program found that has been the focus of many studies; however, none of these studies met the criteria for inclusion in the present review. The lack of research evaluating common secondary reading textbooks does not, of course, mean that these textbooks are ineffective, but it does indicate that there is little evidence for using any one of these pro-

grams in preference to any other if enhancing achievement is the goal.

Mixed-Method Models

Two widely used secondary reading programs, READ 180 and Voyager Passport, were categorized as mixed-method models. These programs combine large-group, small-group, and computer-assisted, individualized instruction. Unlike supplemental CAI models, mixed-method models are intended to serve as complete literacy interventions. Descriptions and outcomes of all studies of mixed-method models in secondary reading that met the inclusion criteria appear in Table 1.

READ 180

READ 180 is an intervention program for upper-elementary, middle, and high school students who are struggling with reading. The program was originally developed by Hasselbring and Goin (2004) at Vanderbilt University and is currently marketed by Scholastic. Stage B of the program, which is designed for students in grade 6 and above who are reading at grade levels from 1.5 to 8, provides groups of 15 students with 90 minutes of instruction per day. Each period of instruction begins with a 20-minute shared-reading and skills lesson. Students then rotate among three activities in groups of five: (1) computer-assisted instructional reading, (2) modeled or independent reading, and (3) small-group instruction with the teacher. The READ 180 software includes videos, mostly about science and social studies topics, and students read about the video content and engage in comprehension, vocabulary, fluency, and word-study activities around this content. In addition, audiobooks model comprehension, vocabulary, and self-monitoring strategies used by good readers, and students read leveled paperbacks in many genres. Teachers are given materials, and they attend workshops to support instruction in reading strategies, comprehension, word study, and vocabulary. A key methodological problem in studies of READ 180 is that many students in READ 180 classes received considerably more instructional time in reading than did their counterparts in control classes. In these cases, the instructional time was confounded with the effects of the program itself.

White, Haslam, and Hewes (2006) and Johnson, Haslam, and White (2006), under contract to the publisher of READ 180, carried out a large-scale evaluation of the program in the Phoenix Union High School District in Phoenix, Arizona, USA. Low-achieving students engaged with READ 180 across the district were matched with low-achieving nonparticipants using propensity matching. The two groups were nearly identical on pretest measures (the Stanford Achievement Test, ninth edition; SAT-9). There were three cohorts that had control groups: (1) students ($n = 1,652$) who were in ninth grade during the 2003–2004 academic year, (2) students ($n = 1,630$) who were in ninth grade during the 2004–2005 academic year, and (3) students ($n = 2,058$) who were in the ninth grade during the 2005–2006 academic year. Experimental groups in all three cohorts used READ 180 for a full year. At the end of the 2003–2004 school year, students who experienced READ 180 scored 1.3 normal curve equivalents (NCE) higher on the SAT-9 than the control group (effect size [ES] = +0.12, $p < .05$). Larger positive effects were obtained for ELLs (ES = +0.32). However, after a one-year follow-up, the 2003–2004 cohort had scores identical to those of nonparticipants on the AIMS (Arizona's Instrument to Measure Standards) reading test (ES = 0.00). Ninth graders in the 2004–2005 cohort scored 2.9 NCEs higher than the control group on the Terra Nova (ES = +0.24, $p < .05$). Once again, positive effects were found for ELLs (ES = +0.41). Students from the 2004–2005 cohort also scored nearly identical to nonparticipants on the AIMS reading test (ES = 0.00) at the end of tenth grade. Ninth graders in the 2005–2006 cohort scored 0.9 NCEs higher than the control group on the Terra Nova (ES = +0.04, $p < .05$). Positive effects were found for ELLs (ES = +0.23). Averaging effect sizes across the SAT-9 outcomes for the 2003–2004 cohort and the Terra Nova outcomes for the 2004–2005 and the 2005–2006 cohorts yielded a mean effect size of +0.13 overall and a mean effect size of +0.32 for ELLs.

Papalewis (2004) carried out a study of 1,073 low-achieving, mostly Hispanic eighth graders in a large urban district in Los Angeles, California, USA. Most students were retained and about half were ELLs. The study compared 537 students enrolled in schools throughout the district who were using READ 180 to 536 well-matched comparison students from other schools across the district. Students who used READ 180 made substantially greater gains on the reading portion of the SAT-9 (ES = +0.68, $p < .05$).

Mims, Lowther, Strahl, and Nunnery (2006), who were third-party evaluators, carried out a large matched evaluation of READ 180 in middle and high schools in Little Rock, Arkansas, USA. Approximately 1,000 mostly African American students in five middle schools and five high schools used READ 180. Using the scores on the reading portion of the 2005 Iowa Tests of Basic Skills (ITBS) and demographic information, each student was individually matched with a student in the same school and grade level who was not using READ 180. Scores on the reading portion of the Spring 2006 ITBS and the Arkansas Benchmark Exams were used as outcome measures.

On the Spring 2006 ITBS, differences favored the control group at all grade levels (grade 6, ES = –0.15; grade 7, ES = –0.23; grade 8, ES = –0.12; and grade 9, ES = –0.16), for an overall mean effect size of –0.17.

Table 1. Mixed-Method Models: Descriptive Information and Effect Sizes for Qualifying Studies

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Posttest	Effect size	Mean effect size
READ 180									
White, Haslam, & Hewes (2006); Johnson, Haslam, & White (2006)	Matched (L)	1 year	1,652 students (826T, 826C)	9th	Students with low reading scores in Phoenix, AZ	Well matched on pretest	SAT-9 ELLs AIMS 1-year follow-up	+0.12 +0.32 0.00	+0.13
		1 year	1,630 students (815T, 815C)				Terra Nova ELLs AIMS 1-year follow-up	+0.24 +0.41 0.00	
		1 year	2,058 students (1,029 T; 1,029 C)				Terra Nova ELLs	+0.04 +0.23	
Papalewis (2004)	Matched (L)	1 year	1,073 students (537T, 536C)	8th (mostly)	Low-performing students in Los Angeles, CA	Well matched on pretest, demographics, and language proficiency	SAT-9 Reading	+0.68	+0.68
Mims, Lowther, Strahl, & Nunnery (2006)	Matched (L)	1 year	1,000 students	6th–9th	Mostly African American students in Little Rock, AR	Well matched on pretests and demographics	ITBS Arkansas Benchmark	−0.17 −0.07	−0.12
Interactive, Inc. (2002)	Matched (L)	1 year	700 students (387T, 323C)	6th–8th	Two middle schools each in Boston, MA and Houston and Dallas, TX	Well matched on pretest and demographics	SAT-9	+0.24	+0.24
Haslam, White, & Klinge (2006)	Matched (L)	1 year	614 students (307T, 307C)	7th–8th	Low-performing students in Austin, TX	Well matched on pretest and demographics	TAKS	+0.18	+0.18
Woods (2007)	Matched (L)	1 year	268 students (134T, 134C)	6th–8th	Low-performing mostly African American students in southeastern Virginia	Well matched on pretest and demographics	Cohort 1: DRP Cohort 2: STAR Reading	+0.05 +0.81	+0.43
Caggiano (2007)	Matched (S)	1 year	120 students (60T, 60C)	6th–8th	Low-performing mostly African American students in southeastern Virginia	Well matched on pretest and demographics	Virginia SOL Grade 6 Grade 7 Grade 8	+0.64 −0.29 −0.31	+0.01
Nave (2007)	Matched post-hoc (S)	1 year	110 students (80T, 30C)	7th	At-risk students in Sevier County, TN	Well matched on pretest and demographics	TCAP	+1.58	+1.58
Voyager Passport									
Shneyderman (2006)	Matched (L)	1 year	8 schools (4T, 4C) 847 students (453T, 394C)	9th–10th	Mostly Hispanic ESL students in Miami, FL	Schools matched on pretest, demographics, and LEP	FCAT Grade 9 Grade 10	+0.22 +0.12	+0.17

Note. L = large study with at least 250 students or at least 10 classes or schools; S = small study with less than 250 students or less than 10 classes or schools; T = treatment; C = control; ESL = English as a second language; LEP = limited English proficiency; AIMS = Arizona's Instrument to Measure Standards; SAT-9 = Stanford Achievement Test, 9th edition; ELLs = English-language learners; ITBS = Iowa Tests of Basic Skills; TAKS = Texas Assessment of Knowledge and Skills; DRP = Degrees of Reading Power; SOL = Standards of Learning; TCAP = Tennessee Comprehensive Assessment Program; FCAT = Florida's Comprehensive Assessment Test.

However, differences were only statistically significant at grades 7 and 9. On the Arkansas Benchmark Exams, patterns were similar. Effect sizes were –0.19 at grade 6, –0.05 at grade 7, and +0.02 at grade 8, for an overall mean effect size of –0.07. Averaging effect sizes for the 2006 ITBS and the benchmark exams gave a mean effect size of –0.12.

The Council of the Great City Schools and Scholastic commissioned an evaluation of READ 180 in three urban districts located in three major U.S. cities (Interactive, Inc., 2002). The study focused on grade 6 in Boston, Massachusetts; grade 8 in Dallas, Texas; and grades 7 and 8 in Houston, Texas. In each case, the SAT-9 was administered as a pre- and posttest. Students in schools using READ 180 were compared to those in schools that were not using the program. Students were matched on pretests and demographic factors. Across the three cities, there were 387 students in the cohort using READ 180 and 323 in the control group. On adjusted posttests, effect sizes averaged +0.24, $p < .001$.

Haslam, White, and Klinge (2006) evaluated READ 180 in the Austin Independent School District in Austin, Texas. Low-achieving seventh and eighth graders using READ 180 throughout the school district ($n = 307$) were matched with a control group ($n = 307$) on demographic factors and Texas Assessment of Knowledge and Skills pretests. At posttest, adjusting for pretests, students who had used READ 180 gained 1.9 NCEs more than the control group (ES = +0.18, $p < .05$).

Woods (2007) evaluated READ 180 in an urban school located in the southeastern part of the U.S. state of Virginia with two cohorts of reading intervention students. Cohort 1 and Cohort 2 were enrolled in middle school during the 2003–2004 and the 2004–2005 academic years, respectively. Data from a third cohort could not be used because the outcome measure was the Scholastic Reading Inventory (SRI), which is used in the READ 180 program. Students in grades 6–8 who needed additional literacy support ($N = 268$) were assigned to either READ 180 or the traditional reading remediation program based on reading pretests and teacher recommendations. READ 180 and comparison students were well matched on reading pretests and demographic factors. Approximately 57% of students participating in the study received free lunch. Of the participants, 63% were African American, and 32% were white. There were 58 students using the READ 180 program during the 2003–2004 school year and 76 using it during the 2004–2005 school year. An equal number of control students participated in the traditional reading remediation program. Students in the treatment group received 90 minutes of READ 180 every other day for the entire school year, whereas students in the comparison condition received 90 minutes of the traditional reading remediation program every other day for one quarter of the

school year. At the end of the 2003–2004 school year, Cohort 1 students who experienced READ 180 gained slightly more on the Degrees of Reading Power test than the control group (ES = +0.05). The use of this test was discontinued, and comparisons between the students who participated in READ 180 during the 2004–2005 school year and those who experienced the traditional reading remediation program were conducted using the STAR Reading assessment program. READ 180 students in Cohort 2 made substantially greater gains on STAR Reading (ES = +0.81). Combining across the two cohorts, the effect size was +0.43.

Caggiano (2007) carried out a year-long study of 120 mostly African American struggling readers enrolled in grades 6, 7, and 8 of an urban middle school located in southeastern Virginia. Twenty students from each grade participated in the READ 180 program. These 60 students were matched with 60 nonparticipants by grade level, gender, ethnicity, and the SRI pretest. All classes received 75 minutes of language arts instruction each day. The students in the experimental group received an additional 90 minutes of supplementary instruction every other day using READ 180. Students were posttested using both the SRI and the Virginia Standards of Learning test. The SRI was included as an assessment tool in the READ 180 package; therefore, we report only the Virginia Standards of Learning test using SRI pretests as covariates. On adjusted posttests, effect sizes were +0.64 at grade 6, –0.29 at grade 7, and –0.31 at grade 8, for an overall mean effect size of +0.01.

Nave (2007) conducted a small retrospective analysis of READ 180 with 110 seventh graders in Sevier County, Tennessee, USA. The Tennessee Comprehensive Assessment Program (TCAP) was used to compare the performance of academically at-risk students who participated in the READ 180 program ($n = 80$) during the 2004–2005 school year to that of a similar group of at-risk students ($n = 30$) who did not participate in the program. There were substantial positive effects on TCAP Reading–Language Arts scores (ES = +1.58).

Across eight studies of READ 180, the mean effect size weighted by sample size was +0.24.

Voyager Passport

Voyager Passport is a mixed-method model designed to provide intensive assistance to students who are reading below grade level. In addition to whole-group instruction, flexible small-group activities, and partner practice, the program engages students with DVDs; online learning activities; and other instructional strategies focusing on comprehension, vocabulary, fluency, and writing.

Shneyderman (2006) carried out an evaluation of Voyager Passport with ninth and tenth graders of limited English proficiency (LEP) in Miami, Florida, USA. Four schools implemented the Voyager Passport program

with their low-achieving, mostly Hispanic LEP students (n = 453). Four control schools were selected using propensity matching, and individual students from these schools (n = 394) were matched to experimental students based on ESOL (English for Speakers of Other Languages) levels. The schools and the sets of students were well matched on the Florida Comprehensive Assessment Test (FCAT) pretests, ESOL levels, and other variables. The report did not state whether or not the control students received any remedial reading intervention. Hierarchical linear modeling with FCAT pretests as covariates found significant positive effects for ninth graders (ES = +0.22, p < .05) but nonsignificant effects for tenth graders (ES = +0.12, p > .05), for a mean effect size of +0.17.

Conclusions: Mixed-Method Models

Across nine studies involving approximately 10,000 students, the weighted mean effect size for mixed-method models was +0.23.

CAI Programs

The effectiveness of CAI has been extensively debated over the past 20 years, and there is a great deal of research on the topic. Kulik (2003) concluded that research did not support use of CAI in elementary or secondary reading, although Chambers (2003) came to a somewhat more positive conclusion, giving a mean effect size of +0.25. A large study of technology immersion, in which Texas middle schools received laptops for every student, extensive software, and significant amounts of professional development, found no significant effects on reading or math achievement in comparison to schools with ordinary levels of technology (Texas Center for Educational Research, 2007). A large randomized evaluation of various computer software programs by Dynarski et al. (2007) found no effects on the reading achievement of first and fourth graders or on the math achievement of sixth graders or students taking algebra. None of these studies or reviews focused specifically on secondary reading, but they nevertheless provide context for this review of the effects of CAI on reading in middle and high schools.

Eight studies of CAI met the standards for this review. These were divided into two categories: (1) supplemental CAI programs and (2) computer-managed learning systems. Supplemental CAI programs such as Jostens and the Computer Curriculum Corporation's (CCC) integrated learning systems are designed to supplement traditional classroom instruction by providing additional instruction at students' assessed levels of need. The category of computer-managed learning systems included only one program, Accelerated Reader. This program uses computers to assess students' reading levels, to assign reading materials at students' levels, to score tests on those

readings, and to chart students' progress; however, students do not work directly on the computer. Descriptions and outcomes of all studies of CAI in secondary reading that met the inclusion criteria appear in Table 2.

Supplemental CAI

Jostens

Jostens is an earlier version of an integrated learning system now called Compass Learning. It provides an extensive set of assessments, which place students in an individualized instructional sequence, and students work individually on exercises designed to fill in gaps in their skills. Jostens is typically used for 15–30 minutes, two to five days per week.

Two studies in rural schools evaluated the Jostens integrated learning system. Roy (1993) evaluated the program in a junior high and a middle school located in different rural areas of Texas. Both schools served primarily Anglo populations. At Midway Junior High, there were 54 sixth graders using Jostens matched with 54 control students. Adjusting for the Norm-Referenced Assessment Program for Texas (NAPT) pretests, there were significantly positive effects on NAPT Reading (ES = +0.38, p < .05). At Hallsville Middle School, 150 seventh and eighth graders using Jostens were matched with a control group of 150 students. There were nonsignificant effects on the NAPT among seventh (ES = +0.10, p > .05) and eighth graders (ES = +0.04, p > .05), for a mean effect size of +0.07. The weighted mean effect size across the two schools was +0.15.

Hunter (1994) evaluated Jostens's effect on second through eighth graders' performance in reading and math in rural Jefferson County, Georgia, USA. The reading evaluation in grades 6–8 is described here. Students participating in Title I, a program providing financial assistance to high-poverty schools and districts, engaged with Jostens for 30 minutes each day for a total of 28 weeks. These students were compared with a control group that did not receive CAI. Three experimental and three control schools were compared. Fifteen students at each grade level from each of the six schools were randomly selected for measurement. Effect sizes were estimated at +0.37 for sixth grade, +0.37 for seventh grade, and +0.19 for eighth grade, for a mean of +0.31.

Across the two studies of Jostens, the weighted mean effect size was +0.21.

CCC Integrated Learning System

The CCC integrated learning system has students work individually on computers to learn and practice skills appropriate to their assessed needs. In a study by Liston (1991), remedial tenth graders used CCC materials focused on four courses of study: (1) reader's workshop and reading for comprehension, (2) practical reading

Table 2. Computer-Assisted Instruction (CAI): Descriptive Information and Effect Sizes for Qualifying Studies

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Posttest	Effect size	Mean effect size
Jostens									
Roy (1993)	Matched (L)	1 year	408 students (204T, 204C)	6th–8th	Schools in rural northeast and central Texas	Well matched on pretest	NAPT Grade 6 (MJH) Grade 7–8 (HMS)	 +0.38 +0.07	+0.15
Hunter (1994)	Matched (L)	28 weeks	6 schools (3T, 3C) 270 students (135T, 135C)	6th–8th	Schools in rural Jefferson County, GA	ANCOVA was used to adjust for pretest differences	ITBS Grade 6 Grade 7 Grade 8	 +0.37 +0.37 +0.19	+0.31
Computer Curriculum Corporation									
Liston (1991)	Matched post-hoc (L)	1 year	49 schools (26T, 23C) 4,597 students (2,288T; 2,309C) in 2 cohorts	10th	Remedial students in South Carolina; 72% African American and 28% white	Schools matched on pretest and demographics	South Carolina Exit Exams African American White	 +0.09 +0.02	+0.06
Other Supplemental CAI Programs									
Chiang, Stauffer, & Cannara (1978)	Matched (S)	1 year	8 schools (4T, 4C) 168 students (99T, 69C)	Junior high school	Special education students in Cupertino, CA	Schools matched on demographics; students matched on pretests and handicaps	PIAT Reading Recognition Reading Comprehension	 +0.33 -0.05	+0.14
Metrics Associates (1981)	Matched (S)	1 year	105 students (70T, 35C)	7th–9th	Two Massachusetts school districts	Scores were adjusted for pretest differences	MAT Reading	+0.56	+0.56
Computer-managed learning systems									
Accelerated Reader									
Hagerman (2003)	Matched (S)	12 weeks	121 students (64T, 57C)	6th	Two suburban middle schools in Oregon	Well matched on pretest and demographics	TORC-3	+0.53	+0.53
Ross & Nunnery (2005)	Matched post-hoc (L)	1 year	10 schools (5T, 5C) 3,230 students (2,106T; 1,124C)	6th–8th	Schools in southern Mississippi	Schools were matched on pretest and demographics; pretest differences were controlled using ANCOVA	MCT Grade 6 Grade 7 Grade 8	 +0.11 +0.16 +0.12	+0.13
Ross, Nunnery, Avis, & Borek (2005)	Matched post-hoc (L)	1 year	4,085 students (2,419T; 1,666C)	6th–8th	Schools in southern Mississippi	Well matched on pretest and demographics	MCT Grade 6 Grade 7 Grade 8	 -0.04 +0.04 +0.10	+0.03

Note. L = large study with at least 250 students or at least 10 classes or schools; S = small study with less than 250 students or less than 10 classes or schools; T = treatment; C = control; NAPT = Norm-Referenced Assessment Program for Texas; MJH = Midway Junior High; HMS = Hallsville Middle School; ITBS = Iowa Tests of Basic Skills; PIAT = Peabody Individualized Achievement Test; MAT = Metropolitan Achievement Test; TORC-3 = Test of Reading Comprehension, 3rd edition; MCT = Mississippi Curriculum Test.

skills, (3) critical reading skills, and (4) survival skills. After an initial assessment, the students were placed at the appropriate points in the individualized curriculum.

The Liston (1991) study involved tenth graders across the U.S. state of South Carolina who had been identified as being in need of remedial instruction according to state standards. Overall, 72% of the students were African American, and 28% were white. Twenty-six CCC high schools were compared with 23 control schools matched on the Comprehensive Test of Basic Skills (CTBS) pretests and ethnicity in a matched post-hoc design. Two cohorts were studied during the 1988–1989 and 1989–1990 school years, respectively. There were 2,278 students (1,161 treatment students and 1,117 control students) in Cohort 1 and 2,319 students (1,127 treatment students and 1,192 control students) in Cohort 2.

CTBS pretests were nearly identical in CCC and control schools. South Carolina exit exams, which are given each spring, showed nonsignificant differences for the first cohort (ES = +0.02, $p > .05$) and small but significant differences for the second cohort (ES = +0.10, $p < .01$), using analyses of covariance. Effect sizes were +0.09 and +0.02 for African American and white students, respectively. The overall mean effect size was +0.06.

Other Supplemental CAI Programs

In an early study of CAI, Chiang, Stauffer, and Cannara (1978) evaluated the use of teacher-authored reading software among academically handicapped students in eight junior high schools in suburban Cupertino, California (N = 168; 99 treatment students and 69 control students). Students used drill-and-practice software in a computer lab for an average of 33 minutes per week as a supplement to other instruction. Schools were matched according to socioeconomic status and pretests. Students, categorized as educable mentally retarded, learning disabled, or oral-language handicapped, were individually pre- and posttested on the Peabody Individualized Achievement Test. Students who received CAI scored higher on Reading Recognition (ES = +0.33) but slightly lower on Reading Comprehension (ES = –0.05), for a mean effect size of +0.14.

Metrics Associates (1981) carried out a small evaluation of the use of a variety of supplemental CAI programs in six school districts in Massachusetts. Two of the districts that participated in the study, Billerica and Woburn, included junior high schools (grades 7–9). In one junior high school in each district, Title I students in the CAI conditions (n = 70) spent 10 minutes of their daily 30-minute remedial reading period using drill-and-practice software. Matched students (n = 35) participated in daily 30-minute remedial classes without CAI. Students were pre- and posttested on the Metropolitan

Achievement Test. Adjusted posttests indicated an effect size of +0.56, $p < .001$.

Computer-Managed Learning Systems

Accelerated Reader

Accelerated Reader is a supplemental program that assesses students' reading levels using a computer, which then prints out suggestions for reading materials at students' levels. Students read books or other materials and then take tests on the computer to show their comprehension of what they have read. Students can earn recognition or rewards based on the number of tests that they have passed.

A small matched study by Hagerman (2003) evaluated Accelerated Reader with sixth graders in a suburban middle school near Portland, Oregon, USA. After using Accelerated Reader for 12 weeks, the treatment students (n = 64) were compared with matched students who were enrolled in another middle school in the same district (n = 57). Students were pre- and posttested on the Test of Reading Comprehension, third edition. On posttests adjusted for pretests, the Accelerated Reader group scored significantly higher (ES = +0.53, $p < .001$).

The largest evaluations by far of Accelerated Reader in grades 6–8 were carried out in two school districts, Pascagoula and Biloxi, in the U.S. state of Mississippi. Data on two cohorts of students were analyzed by third-party evaluators working under contract to the program's publisher. During the 2002–2003 school year, Ross and Nunnery (2005) compared one-year gains for schools using Accelerated Reader (n = 2,106 students) to those in matched schools using traditional methods (n = 1,124 students). The schools using Accelerated Reader were also using Accelerated Math. During the 2003–2004 school year, the same comparisons were made in the same schools by Ross, Nunnery, Avis, and Borek (2005) with 2,419 students using the Accelerated Reader program and 1,666 students in the control group. Some students were of course in the treatment groups for both years, but the data are presented as two cross-sectional studies, not as a longitudinal study. Effect sizes for the 2002–2003 cohort on the reading portion of the Mississippi Curriculum Test, adjusted for pretests, were +0.11 for sixth grade, +0.16 for seventh grade, and +0.12 for eighth grade, for a mean of +0.13, $p < .05$. For the 2003–2004 cohort, effect sizes were –0.04 for sixth grade, +0.04 for seventh grade, and +0.10 for eighth grade, for a mean of +0.03, $p > .05$. Combining across both cohorts, the mean effect size was +0.08.

The weighted mean effect size across all three qualifying studies of Accelerated Reader was +0.09.

Conclusions: CAI

A total of 8 qualifying studies evaluated various forms of CAI. The studies involved a total of 12,984 students.

Overall, the weighted mean effect size was +0.10. This is less than the median effect size of +0.18 for CAI in secondary math reported by Slavin et al. (2007), but it is in accord with the conclusions drawn from a review of research on CAI by Kulik (2003). (Kulik did not report a mean effect size.)

Instructional-Process Programs

Instructional-process programs are methods that focus on providing teachers with extensive professional development to implement specific instructional methods. These programs fell into three categories: (1) cooperative learning, (2) strategy instruction, and (3) comprehensive school reform. Cooperative learning programs (Slavin, in press) have students work in small groups to help one another master academic content. Strategy instruction programs incorporate methods that teach students to use specific study strategies such as paraphrasing, summarization, and prediction to improve their reading comprehension. Comprehensive school reform programs attend to instruction, curriculum, assessment, classroom management, and parent involvement, among other factors. Only comprehensive school reform programs that incorporate specific reading approaches are reviewed here (for others, see Comprehensive School Reform Quality Center, 2006; Borman et al., 2003). Descriptions and outcomes of all studies of instructional-process programs that met the inclusion criteria appear in Table 3.

Cooperative Learning Programs

Peer-Assisted Learning Strategies

Peer-Assisted Learning Strategies, or PALS, is a cooperative learning program in which students work in pairs, taking turns reading aloud to one another and engaging in summarization and prediction activities. PALS has primarily been used in the early elementary grades, where it has been successfully evaluated (Fuchs, Fuchs, Mathes, & Simmons, 1997); however, it is also used in remedial and special education programs in upper-elementary and secondary grades.

Calhoon (2005) evaluated an application of PALS with students who were enrolled in two middle schools in the southwestern United States and who were reading at or below the third-grade level. The 31-week treatment combined PALS with a training approach that emphasized linguistic skills in which students took turns tutoring each other on specific phonological and spelling skills. Four special education teachers and their classes of students with learning disabilities (N = 38) were randomly assigned to PALS or control conditions, making this a randomized quasi-experiment. Most students were sixth graders; however, a few seventh graders and one eighth grader also participated. Students were pre- and posttested on four scales from the Woodcock-Johnson III.

Adjusting for pretests, there were significant differences on Letter–Word Identification (ES = +0.84, p < .05), Passage Comprehension (ES = +0.66, p < .05), and Word Attack (ES = +0.46, p < .05) but not on Reading Fluency (ES = –0.13, p > .05). The mean effect size was +0.46.

Fuchs, Fuchs, and Kazdan (1999) evaluated PALS among special education and remedial classes in 10 high schools in the southeastern United States (N = 102 students). Eighteen teachers were nonrandomly assigned to PALS or control classes in a 16-week study. The experimental group used PALS procedures on alternating days, averaging 2.5 times per week for the entire study. Students were pre- and posttested on an experimenter-made measure called the Comprehensive Reading Assessment Battery, an oral reading measure not aligned with the PALS intervention. Controlling for pretests, differences were statistically significant on comprehension questions (ES = +0.33, p < .05) but not on words read correctly (ES = +0.04, p > .05), for a mean effect size of +0.19.

Hankinson and Myers (2000) evaluated PALS in a suburban middle school near Pittsburgh, Pennsylvania, USA. A total of 51 eighth graders experienced PALS, and 32 served as a matched control group in a 12-week study. Students were pretested on the Gates–MacGinitie Reading Test (GMRT) and the comprehension measure of the Pennsylvania System of School Assessment (PSSA), and 12 weeks later, they were posttested. Adjusting for pretests, PALS students gained more than controls on GMRT Vocabulary (ES = +0.10) and Comprehension (ES = +0.44), although these gains were nonsignificant, for a mean effect size of +0.27. On the PSSA, students in the control group made nonsignificantly greater gains than the treatment group (ES = –0.34), although the report noted that the control group received special practice on this measure. The mean across the two measures was –0.04.

The weighted mean effect size across the three studies of PALS was +0.15; however, the one randomized quasi-experiment had the strongest positive effects.

Student Team Reading[1]

Student Team Reading (Stevens & Durkin, 1992) is a cooperative learning program for middle schools in which students work in four- or five-member teams to help one another build reading skills. Based on a program called Cooperative Integrated Reading and Composition (Stevens, Madden, Slavin, & Farnish, 1987), which is used in upper-elementary grades, Student Team Reading has students engage in partner reading, story retelling, story-related writing, word mastery, and story-structure activities to prepare them and their teammates for individual assessments that form the basis for team scores. Instruction focuses on explicit teaching of metacognitive strategies.

Stevens and Durkin (1992, Study 1) carried out a large-scale matched evaluation of Student Team Reading in five high-poverty, mostly African American middle

Table 3. Instructional-Process Programs: Descriptive Information and Effect Sizes for Qualifying Studies

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Posttest	Effect size	Mean effect size
Peer-Assisted Learning Strategies (PALS)					Cooperative learning programs				
Calhoon (2005)	Randomized quasi-experiment (S)	31 weeks	38 students	6th–8th	Special education classes taught by 4 teachers in 2 middle schools in the southwestern U.S.	Scores were adjusted for any pretest differences	WJ-III Letter-Word Identification Passage Comprehension Word Attack Reading Fluency	+0.84 +0.66 +0.46 -0.13	+0.46
Fuchs, Fuchs, & Kazdan (1999)	Matched (S)	16 weeks	18 classes (9T, 9C) 102 students (52T, 50C)	High school	Special education and remedial classes in 10 high schools within one metropolitan southeastern U.S. school district	Scores were adjusted for any pretest differences	CRAB Comprehension Correct words read	+0.33 +0.04	+0.19
Hankinson & Myers (2000)	Matched (S)	12 weeks	83 students (51T, 32C)	8th	Suburban middle school near Pittsburgh, PA	Well matched on pretest	GMRT Vocabulary Comprehension PSSA Reading Comprehension	+0.10 +0.44 -0.34	-0.04
Student Team Reading (STR)									
Stevens & Durkin (1992), Study 1	Matched (L)	1 year	3,986 students	6th–8th	5 middle schools in Baltimore, MD	Well matched on demographics; control group's scores were higher than the treatment group's at pretest	CAT Reading Vocabulary Reading Comprehension	+0.46 +0.34	+0.40
Stevens & Durkin (1992), Study 2	Matched (L)	1 year	1,233 students (455T, 768C)	6th	20 classes in 6 middle schools in an urban district in Maryland	Well matched on pretest	CAT Reading Vocabulary (all students) Reading Comprehension (all students) Reading Vocabulary (special education students) Reading Comprehension (special education students)	-0.02 +0.13 +0.28 +0.60	+0.06

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Posttest	Effect size	Mean effect size
The Reading Edge									
Chamberlain, Daniels, Madden, & Slavin (2007); Slavin, Chamberlain, Daniels, & Madden (2008)	Randomized (L)	1 year	788 students (405T, 383C) in 2 cohorts	6th	Two majority-white, high-poverty rural middle schools in West Virginia and Florida	Well matched on pretest, demographics, and special-education status	GMRT Total Vocabulary Comprehension	+0.15 +0.15 +0.12	+0.15
Slavin, Daniels, & Madden (2005)	Matched (L)	3 years	14 schools (7T, 7C) 3,470 students (1,748T; 1,722C)	6th–8th	High-poverty schools throughout the U.S.	Schools were matched on pretest and demographics	State assessments	+0.33	+0.33
Strategy instruction programs									
Reading Apprenticeship									
Kemple et al. (2008)	Randomized (L)	1 year	17 schools 1,140 students (686T, 454C)	9th	Struggling students in schools across the U.S.; mostly African American and Hispanic students	Well matched on pretest	GRADE Comprehension Vocabulary	+0.09 +0.05	+0.07
Xtreme Reading									
Kemple et al. (2008)	Randomized (L)	1 year	17 schools 1,273 students (722T, 551C)	9th	Struggling students in schools across the U.S.; mostly African American and Hispanic students	Well matched on pretest	GRADE Comprehension Vocabulary	+0.09 +0.01	+0.05
Benchmark Detectives Reading Program									
Gaskins (1994)	Matched (S)	1–2 years	83 students (36T, 47C)	6th	Benchmark School in Pennsylvania	ANCOVA was used to adjust for IQ and age	MAT Reading 1 year 2 years	+0.21 +0.52	+0.52
Strategy Intervention Model (SIM)									
Losh (1991)	Matched (S)	1 year	64 Students (32T, 32C)	7th–9th	Students with learning disabilities in a Nebraska junior high school	Students were individually matched on pretest, demographics, grade level, and handicapping condition	CAT Composite Comprehension Vocabulary	+0.11 +0.24 -0.01	+0.11
Mothus (1997)	Matched post-hoc (S)	2 years	67 students (33T, 34C)	8th	Low-performing students at 2 middle class, mostly white junior high schools in central British Columbia, Canada	Students were well matched on pretest	SDRCT	+0.36	+0.36

(continued)

Table 3. Instructional-Process Programs: Descriptive Information and Effect Sizes for Qualifying Studies (continued)

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Posttest	Effect size	Mean effect size
					Comprehensive school reform programs				
Talent Development High School (TDHS)									
Balfanz, Legters, & Jordan (2004)	Matched (L)	1 year	6 schools (3T, 3C) 457 students (257T, 200C)	High school	Inner-city high schools in Baltimore, MD	Scores were adjusted for pretest differences	Terra Nova	+0.17	+0.17
Kemple, Herlihy, & Smith (2005)	Matched (L)	3 years	11 schools (5T, 6C) 399 students	9th–11th	High-poverty, mostly African American schools in Philadelphia, PA	Students were matched on 8th-grade test scores	PSSA-Reading	-0.04	-0.04
Talent Development Middle School (TDMS)									
Herlihy & Kemple (2004, 2005)	Matched (L)	4–6 years	12 schools (6T, 6C)	6th–8th	Middle schools in Philadelphia, PA	TDMS schools were matched with comparison schools on pretest and demographics	PSSA Year 1 Year 2 Year 3 Year 4 Year 5 Year 6	-0.07 +0.16 0.00 -0.06 +0.15 +0.06	+0.04
Mac Iver, Balfanz, Ruby, Byrnes, Lorentz, & Jones (2004)	Matched (L)	3 years	6 schools (3T, 3C) 1,552 students (890T, 662C)	6th–8th	Middle schools in Philadelphia, PA	TDMS schools were matched with comparison schools on pretest and demographics	PSSA	+0.20	+0.20

Note. L = large study with at least 250 students or at least 10 classes or schools; S = small study with less than 250 students or less than 10 classes or schools; T = treatment; C = control; WJ-III = Woodcock Johnson, 3rd edition; CRAB = Comprehensive Reading Assessment Battery; GMRT = Gates–MacGinitie Reading Test; PSSA = Pennsylvania System of School Assessment; CAT = California Achievement Test; GRADE = Group Reading Assessment and Diagnostic Examination; MAT = Metropolitan Achievement Test; SDRCT = Stanford Diagnostic Reading Comprehension Test.

schools in Baltimore, Maryland, USA. Two Student Team Reading schools with 72 reading classes in grades 6–8 were matched on demographic characteristics and California Achievement Test (CAT) pretests with three control schools with 88 reading classes in grades 6–8 (N = 3,986). Students in the Student Team Reading classes also experienced a component called Student Team Writing.

On reading measures, using z-scores to combine across grades 6–8 and adjusting for pretests, Student Team Reading classes scored significantly higher than the control classes on CAT Reading Vocabulary (+0.46, p < .05) and Reading Comprehension (+0.34, p < .05), for a mean effect size of +0.40. There were also positive effects on CAT Language Expression, but this is ascribed to the Student Team Writing component, not Student Team Reading.

In a similar study, Stevens and Durkin (1992, Study 2) evaluated Student Team Reading in six high-poverty, mostly African American middle schools that were also located in Baltimore. Three schools with 20 sixth-grade classes were compared to three schools with 34 sixth-grade classes (N = 1,233; 455 treatment students and 768 control students). On CAT posttests, controlling for CAT pretests, there were small but significant differences favoring Student Team Reading on Reading Comprehension (ES = +0.13, p < .05), but there were no differences on Reading Vocabulary (ES = –0.02, p > .05). The mean effect size was +0.06. Separate analyses for students with special needs found much larger impacts with effect sizes of +0.60 for Reading Comprehension and +0.28 for Reading Vocabulary, for a mean effect size of +0.44.

The Reading Edge[2]

In an adaptation of Student Team Reading, Slavin, Daniels, and Madden (2005) created a program called The Reading Edge to serve as the reading component of the Success for All Middle School program. The Reading Edge uses the same cooperative learning structures and basic lesson design as Student Team Reading but regroups students for reading instruction according to their reading levels across grades and classes.

An evaluation of The Reading Edge by Chamberlain, Daniels, Madden, and Slavin (2007) and Slavin, Chamberlain, Daniels, and Madden (2008) randomly assigned two successive cohorts of sixth graders within two high-poverty, majority-white middle schools to treatment or control classes. One of the middle schools was located in a rural area of the U.S. state of West Virginia, the other in a rural area of Florida. Combining across cohorts, there was a total of 788 students (405 treatment students and 383 control students). On GMRT posttests, controlling for pretests, students in The Reading Edge classes scored significantly higher than those in the control classes on Reading Total (ES = +0.15,

p < .01). On subtests, students in The Reading Edge classes scored significantly higher on Vocabulary (ES = +0.15, p < .01), and there were smaller significant differences on Comprehension (ES = +0.12, p < .05). There were no significant differences in outcomes between the two cohorts.

A large-scale matched study of The Reading Edge was carried out by Slavin et al. (2005). Seven high-poverty schools in six U.S. states implemented The Reading Edge over a three-year period. Each of the seven schools was matched on prior achievement and demographic factors with a control school in the same state (usually in the same district), and state test scores (percent scoring proficient or better) were compared at pre- and posttest. A total of 3,470 students (1,748 treatment students and 1,722 control students) were involved. Using arcsine transformations to analyze data on the proportions of experimental and control students who passed their state tests at pre- and posttest (Lipsey & Wilson, 2001), effect sizes were estimated for each pair of schools. One of the schools, located on an American Indian reservation in the U.S. state of Washington, made extraordinary gains, going from a zero to a 96% passing rate on the Washington Assessment of Student Learning, while its control school, which was also on a reservation, gained 18 percentage points, for an effect size of +2.29. Because of this positive outlier, a median rather than a mean was computed across all seven school pairs on their respective state tests, yielding a median effect size of +0.33.

Across seven qualifying studies of cooperative learning approaches to middle school reading, the weighted mean effect size was +0.28. The four studies of the similar Student Team Reading and The Reading Edge approaches had a weighted mean effect size of +0.29.

Strategy Instruction Programs

Strategy instruction programs are reading approaches that emphasize the teaching of cognitive and metacognitive reading strategies such as summarization, use of graphic organizers, and previewing.

Reading Apprenticeship and Xtreme Reading

Both Reading Apprenticeship and Xtreme Reading are supplemental literacy programs designed to help struggling high school readers improve their reading skills. Reading Apprenticeship was designed by WestEd, an educational laboratory. Through teaching strategies based on "cognitive apprenticeship" (gradually passing responsibility from teacher to students), this program emphasizes the development of metacognitive skills, sustained silent reading, language study, and writing. Xtreme Reading was developed by the Center for Research on Learning at the University of Kansas and emphasizes teaching of cognitive and metacognitive skills, vocabulary, and word identification. Teachers and

students follow a regular routine of modeling, practice, paired practice, independent practice, differentiated instruction, and integration and generalization.

As part of a recent initiative of the U.S. Institute of Education Sciences, Kemple et al. (2008) evaluated these two promising approaches to reading instruction. Kemple et al. (2008) randomly assigned 34 high schools in 10 districts across the United States to use either Reading Apprenticeship or Xtreme Reading. Within schools, entering ninth graders reading two to four grades below level were randomly assigned to treatment (686 Reading Apprenticeship students; 722 Reading Xtreme students) or control conditions (454 students in Reading Apprenticeship control group; 551 students in Xtreme Reading control group). Overall, the students were 45% African American, 32% Hispanic, 18% white, and 5% other. Students were pre- and posttested on the Group Reading Assessment and Diagnostic Evaluation. Controlling for pretests, the Reading Apprenticeship outcomes for comprehension (ES = +0.09, $p > .05$) and vocabulary (ES = +0.05, $p > .05$) resulted in a mean effect size of +0.07. For Xtreme Reading, there were few differences in reading comprehension (ES = +0.09, $p > .05$) or reading vocabulary (ES = +0.01, $p > .05$), for a mean effect size of +0.05.

The Benchmark Detectives Reading Program

Gaskins (1994) evaluated a form of strategy instruction for struggling readers of normal or superior intelligence called the Benchmark Detectives Reading Program. This program was used in the Benchmark School, a Pennsylvania middle school where teachers were given professional development in the use of cognitive and metacognitive reading strategies across the curriculum (N = 83 students). In monthly inservice sessions taught by a variety of national experts on the use of cognitive strategy instruction, as well as within-school coaching, coteaching, and conference attendance, the teachers learned several comprehension strategies and methods for introducing these strategies to their students. An evaluation compared students in three cohorts entering the middle grades to those in a previous cohort that did not experience strategy instruction. The cohorts were similar on IQ measures from the revised Wechsler Intelligence Scale for Children (WISC-R). On the reading portion of the Metropolitan Achievement Test, adjusted for WISC-R, the strategy group had scores that were higher but not significantly higher than the baseline group after one year (ES = +0.21, $p > .05$) and scores that were significantly higher after two years (ES = +0.52, $p < .01$).

Strategy Intervention Model

The Strategy Intervention Model, also known as the Strategic Instruction Model (SIM; Schumaker, Denton, & Deshler, 1984), is a method in which low-achieving sec-ondary students are taught metacognitive reading strategies, especially paraphrasing, to help them comprehend text.

A small study of SIM by Losh (1991) involved students with learning disabilities in a junior high school located in the U.S. state of Nebraska. Students in a SIM group (n = 32) were individually matched with students in a control group (n = 32) based on CAT reading scores, handicapping condition, gender, and grade level. On the Spring 1990 CAT scores, controlling for prior scores on the 1989 CAT, SIM students scored higher on the CAT Composite (ES = +0.11, $p > .05$), although these scores were nonsignificant. There were positive effects for Comprehension (ES = +0.24, $p > .05$) but not Vocabulary (ES = –0.01, $p > .05$).

Mothus (1997) carried out a small matched post-hoc evaluation of SIM in two middle class, mostly white junior high schools in central British Columbia, Canada. One school had used SIM for two years with two cohorts of low-achieving eighth graders (n = 33). These students were compared to students in the same school and in a neighboring school (n = 34) who received conventional learning assistance and were well matched on the Stanford Diagnostic Reading Comprehension Tests (SDRCT) given at the beginning of eighth grade. The students in the SIM treatment group were also compared to matched low achievers in both schools who received neither SIM nor conventional learning assistance but were similarly low achieving. On SDRCT posttests at the end of the two years of treatment, SIM students scored significantly higher than both the learning-assistance group (ES = +0.39, $p < .05$) and the unserved group (ES = +0.32, $p < .05$), for a mean effect size of +0.36.

Comprehensive School Reform Programs

Comprehensive school reform programs are whole-school models that include extensive professional development in instructional methods, curriculum, school organization, classroom management, parent involvement, and other issues. As noted earlier, only comprehensive school reform models with specific approaches to reading were included.

Talent Development High School[3]

Talent Development High School, or TDHS, is a comprehensive reform model that focuses on improving students' reading and math performance in high-poverty high schools. A key element of the approach is a ninth-grade academy, which provides a "double dose" of reading and math instruction (90 minutes of each per day). The reading program, called Strategic Reading, is used in the first semester. It emphasizes teacher modeling of comprehension processes, minilessons on comprehension strategies and writing, cooperative learning with paired reading and discussion groups, and self-selected

reading. In the second semester, students experience the district's English I curriculum, supported by TDHS discussion guides and writing supplements that combine Strategic Reading methods with the district curriculum.

Balfanz, Legters, and Jordan (2004) evaluated the TDHS Strategic Reading approach in three inner-city, very low-achieving high schools in Baltimore with mostly African American student populations. The three TDHS schools, which had 20 general-education reading classes taught by eight teachers (n = 257 students), were compared to three control schools (n = 200 students) that were well matched on pretest scores and demographic factors. The control schools also provided a double dose of reading and math instruction (90 minutes of each per day); thus, instructional time was similar for students in both the treatment and control schools. At the end of one year, TDHS students scored significantly better than students in the control group on the district-administered Terra Nova scores, after adjusting for pretests (ES = +0.17, $p < .01$).

A third-party evaluation of the TDHS model was carried out in five high-poverty, mostly African American schools in the U.S. city of Philadelphia by Kemple, Herlihy, and Smith (2005). Six high schools matched on eighth-grade PSSA scores served as controls. Eleventh-grade PSSA-Reading scores served as posttests. Due to high mobility over the course of the three-year experiment, only 399 students from the original sample were still present at posttest, but the rate of attrition was similar for the two groups. Among this subsample, effect sizes were estimated at –0.04, $p > .05$.

Talent Development Middle School[4]

Talent Development Middle School (TDMS) is a comprehensive reform model designed to help high-poverty urban middle schools improve outcomes for their students. It organizes schools into small, interdisciplinary learning communities and introduces teaching methods in language arts, math, science, and U.S. history that emphasize cooperative learning. Remedial courses in reading and math are provided for struggling students, and extensive professional development and coaching are given to all teachers. For reading, TDMS uses an adaptation of Student Team Reading called Student Team Literature, which also incorporates a focus on classic books, more high-level questions, and additional background information for students.

A third-party evaluation of TDMS was carried out by Herlihy and Kemple (2004, 2005). Using a comparative interrupted time-series design, six middle schools in Philadelphia were compared to six matched comparison schools in the same district over three baseline years and four to six implementation years. For reading, eighth-grade scores on the PSSA for successive cohorts of students were compared in terms of each school's deviation from its own three-year baseline average. The comparisons in gains were made across experimental and control groups. Different schools had different numbers of follow-up years, but differences in scores on the PSSA were small in all years (Year 1, ES = –0.07, $p > .05$; Year 2, ES = +0.16, $p < .01$; Year 3, ES = 0.00, $p > .05$; Year 4, ES = –0.06, $p > .05$; Year 5, ES = +0.15, $p > .05$; Year 6, ES = +0.06, $p > .05$). The mean effect size across all years was +0.04.

Mac Iver et al. (2004) reported a three-year evaluation of TDMS in the first three Philadelphia schools to use the program involving cohorts overlapping those in the Herlihy and Kemple (2004, 2005) study. The TDMS schools (n = 890 students) were compared to three matched control schools (n = 662). Overall, the schools were approximately 42% African American, 41% Hispanic, 9% white, and 8% Asian American and served impoverished neighborhoods. Controlling for fifth-grade PSSA scores, eighth-grade PSSA scores for students who had been in their respective schools throughout the study favored the TDMS schools by 4.3 NCEs (ES = +0.20, $p < .001$).

Averaging across the two evaluations of TDMS, the mean effect size was +0.12.

Conclusions: Instructional-Process Programs

As was true in the Slavin and Lake (in press) elementary math review and the Slavin et al. (2007) secondary math review, the largest numbers of rigorous studies that met the inclusion criteria for the present review were those that evaluated instructional-process programs. Across 16 studies, involving approximately 15,000 students, the weighted mean effect size was +0.21. The three randomized studies had a weighted mean effect size of +0.08.

Seven of the studies (two of which used randomized designs) evaluated various forms of cooperative learning with 9,700 students. These had a weighted mean effect size of +0.28. This corresponds with findings from the math reviews, which for cooperative learning reported median effect sizes of +0.29 at the elementary level (Slavin & Lake, in press) and +0.32 at the middle and high school level (Slavin et al., 2007). The weighted mean effect size across the four studies of the two similar programs Student Team Reading and The Reading Edge was +0.29; these studies involved 9,477 students. Two large randomized studies and three small matched studies found small positive effects for programs that teach cognitive and metacognitive strategies to students, with a weighted mean effect size of +0.09.

Overall Patterns of Outcomes

Across all categories, there were 33 qualifying studies of middle and high school reading programs involving a total of nearly 39,000 students. Four of the qualifying studies used random assignment. The mean effect size

weighted by sample size across all 33 studies was +0.17. These studies were identified from among more than 300 studies initially reviewed and represent those that used rigorous experimental procedures.

The most surprising finding is the fact that no studies of secondary reading textbooks met the inclusion criteria. Widely used programs such as McDougal Littell and LANGUAGE! have not been studied in experimental-control comparisons that met the standards of this review. This contrasts with the situation in secondary math, where Slavin et al. (2007) found 38 qualifying studies of math curricula and 100 qualifying studies overall. Of course, reading traditionally has not been taught in middle and high schools except to students in remedial and special education programs, but it is distressing, nevertheless, to find so little evidence behind the curricula used with hundreds of thousands of secondary students who struggle with reading.

The three categories in which qualifying studies did exist were mixed-method models, CAI, and instructional-process programs. There were robust positive effects on achievement in mostly matched quasi-experiments for mixed-method models such as READ 180 and Voyager Passport (weighted mean effect size of +0.23 across nine studies) and for instructional-process programs using cooperative learning (weighted mean effect size of +0.28 across seven studies). However, effects for CAI programs were small (weighted mean effect size of +0.10 across eight studies), as were effects for reading strategy programs that did not emphasize cooperative learning (weighted mean effect size of +0.09 across five studies).

The mean effect sizes reported for programs categorized as having moderate evidence of effectiveness range from +0.20 to +0.35 and are similar to those found in previous reviews of research on math programs. Such effects are modest compared to those often reported for brief experiments or studies that use measures closely aligned with treatments, but they are important given that they come from large, realistic studies mostly using the kinds of standardized tests for which schools are held accountable. In addition, these standardized tests probably underestimate the true impact of experimental treatments as the tests are unlikely to be sensitive to the specific content being taught. The importance of effect sizes of this magnitude becomes clear in light of the fact that an effect size of +0.25 represents about half of the minority–white achievement gap on the National Assessment of Educational Progress (Lee, Grigg, & Donahue, 2007). The large, extended studies with standard measures that form the core of the present review illustrate what could be accomplished at the policy level if schools widely adopted and implemented effective programs, not what could theoretically be gained under ideal, hothouse conditions.

Sample Size Matters

One factor that did differentiate among studies was sample size. Studies with total sample sizes of 250 or more students (125 students per treatment), or 10 or more classes, were considered "large." Previous research (e.g., Rothstein et al., 2005; Slavin, 2008; Sterne, Gavaghan, & Egger, 2000; Taylor & Tweedie, 1998) has shown that studies with small sample sizes report larger effect sizes than studies with large samples. This is due primarily to the fact that small studies produce much more variable outcomes than large studies. In addition, small, underpowered studies that produce zero or negative effects are less likely to be published or locatable in any format; thus, these studies are rarely available for review. Moreover, authors are reluctant even to write up the results of small studies that find zero or negative effects, and journal editors are unlikely to publish such studies. As a result, reports of small studies are likely to be available only when their effects are so large that they are statistically significant despite their small sample sizes. In contrast, large studies finding zero or negative effects are more likely to be published, and because large studies are likely to have been funded or completed as part of a scholar's doctoral work, they are more likely to be reported, even if the report is not published. In addition, studies with statistically significant differences are more likely to be published or otherwise reported, and small studies only have significant differences if effect sizes are large (Rothstein et al., 2005).

In the present review, large studies clearly produced lower effect sizes than small studies. For the 22 large studies, the median effect size was +0.15, while the 11 small studies had a median effect size of +0.36. Because of these differences, the present study used mean effect sizes weighted by sample size (up to a cap of 2,500 students) in pooling effect sizes across studies.

Summarizing Evidence of Effectiveness for Current Programs

For many audiences, it is useful to have summaries of the strength of the evidence supporting achievement effects for programs that educators might select to improve student outcomes. Slavin (2008) proposed a rating system for such programs that is intended to balance methodological quality, weighted mean effect sizes, sample sizes, and other factors, and this system was applied by Slavin and Lake (in press) and Slavin et al. (2007). Using the same rating system and drawing on the results of the present review, secondary reading programs were categorized as follows: strong evidence of effectiveness, moderate evidence of effectiveness, limited evidence of effectiveness, insufficient evidence of effectiveness, and no qualifying studies. Programs with strong evidence of effectiveness

had at least two large studies, one of which was a large randomized or randomized quasi-experimental study, or multiple smaller studies, with an effect size weighted by sample size of at least +0.20. A large study was defined as one in which at least 10 classes or schools, or 250 students, were assigned to treatments. Smaller studies were counted as equivalent to a large study if their collective sample sizes were at least 250 students. Effect sizes from randomized studies took precedence over those from matched studies. Programs with moderate evidence of effectiveness had at least two studies of any design, each with a collective sample size of 250 students, with a weighted mean effect size of at least +0.20. Programs with limited evidence of effectiveness had at least one qualifying study of any design with a weighted mean effect size of at least +0.10. Those programs categorized as having insufficient evidence of effectiveness had one or more qualifying study of any design with nonsignificant outcomes and a weighted mean effect size of less than +0.10.

Table 4 summarizes currently available programs falling into each of these categories. (Within categories, programs are listed in alphabetical order.)

None of the programs qualified for the strong evidence of effectiveness category; however, four programs met the criteria for moderate evidence of effectiveness. Two of these were the cooperative learning programs The Reading Edge and Student Team Reading. READ 180, a mixed-method approach that uses computers in a broader comprehensive model, also fell into this category, as did the early CAI program, Jostens.

Six programs fell into the limited evidence of effectiveness category. These were SIM and the Benchmark Detectives Reading Program, both of which provide strategy instruction to students, as well as Voyager Passport, PALS, Accelerated Reader, and TDMS.

Discussion

The most important conclusion of the research reviewed in this article is that there are fewer large, high-quality studies of middle and high school reading programs than one would wish. There were no methodologically adequate studies comparing different reading texts or curricula. Although 33 studies (involving nearly 39,000 students) did qualify for inclusion, there were only a small number of studies of any particular program, and only four studies involved random assignment to conditions. Further, causal claims cannot be made with confidence in systematic reviews, which can only examine existing studies.

Keeping these limitations in mind, there are several important patterns in the findings that are worthy of note. First, this review found that most of the programs with good evidence of effectiveness have cooperative learning at their core. These programs all rely on a form of cooperative learning in which students work in small groups to help one another master reading skills and in which the success of the team depends on the individual learning of each team member. Both of these elements have been identified by previous reviewers (e.g., Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003; Slavin, 1995, in press; Webb & Palincsar, 1996) as essential to the effectiveness of cooperative learning. The finding of positive effects for cooperative learning programs is consistent with the findings of reviews of elementary and secondary math programs (Slavin & Lake, in press; Slavin et al., 2007).

Positive effects were also seen for other programs designed to improve the core of classroom practice. Mixed-method models, which combine large-group, small-group, and CAI, provide extensive professional development to teachers, as do strategy instruction programs such as SIMS and the Benchmark Detectives Reading Program. Like cooperative learning programs, these approaches focus on improving classroom teaching, and have good evidence of effectiveness.

Also consistent with previous research is the finding in the present study that forms of CAI generally produced small effects. An earlier review of CAI in math and reading by Kulik (2003) found similarly few positive effects for reading.

The findings of this review add to a growing body of evidence to the effect that what matters for student achievement are approaches that fundamentally change what teachers and students do every day (such as cooperative learning and mixed-method models). In earlier reviews, these strategies had outcomes that were clearly and consistently more positive than those found for curricula or CAI alone. More research and development of reading programs for secondary students is clearly needed, but we already know enough to take action, to use what we know now to improve reading outcomes for students with reading difficulties in their critical secondary years.

Table 4. Strength of Evidence for Secondary Reading Programs

Strength of evidence	Program
Strong	None
Moderate	Jostens The Reading Edge READ 180 Student Team Reading
Limited	Accelerated Reader Benchmark Detectives Peer-Assisted Learning Strategies (PALS) Strategy Intervention Model Talent Development Middle School Voyager Passport
Insufficient	Computer Curriculum Corporation (CCC) Reading Apprenticeship Talent Development High School Xtreme Reading

No qualifying studies

100 Book Challenge
ABC's of Reading
Academy of Reading
Achieve 3000
Achieving Maximum Potential
Advancement Via Individual Determination (AVID)
AfterSchool KidzLit
Alphabetic Phonics
America's Choice Ramp-Up Literacy
AMP Reading System
Barton Reading & Spelling System
Be a Better Reader
BOLD
Breaking the Code
Bridges to Literacy
Caught Reading
Charlesbridge Reading Fluency
Classworks
Compass Learning
Comprehension Upgrade
Concept-Oriented Reading Instruction (CORI)
Corrective Reading
Creating Independence Through Student-Owned Strategies (CRISS/Project CRISS)
Cross-Aged Literacy Program
Direct Instruction
Disciplinary Literacy
Electronic Bookshelf
Essential Learning Systems
Exemplary Center for Reading Instruction (ECRI)
Failure Free Reading
Fast ForWord
Fast Track Reading
First Steps
Fluent Reader
Glass-Analysis Method
Glencoe
Great Leaps
Harcourt
HOSTS
Houghton Mifflin
IMPACT
IndiVisual Reading

In Step Readers
Intensive Supplemental Reading
Jamestown Education
Junior Great Books
Kaplan SpellRead
Knowledge Box
K-W-L Strategy
LANGUAGE!
Learning Experience Approach
Learning Upgrade
Lexia Strategies for Older Students
Like to Read
Lindamood-Bell
LitART
Literacy First
MacMillan
McDougal-Littell
Merit Software
Multicultural Reading and Thinking (McRAT)
My Reading Coach
OnRamp Approach
Open Book Anywhere
Open Court
Pathway Project
Phonics for Reading
Phono-Graphix
PLATO
Prentice Hall Literature
Project Read
Puente
Questioning the Author
QuickReads-Secondary
Quicktionary Reading Pen II
Ramp-Up Literacy
Rave-O
REACH
ReadAbout
Read Naturally
Read Now
Read On!
READ RIGHT
Read XL
Reader's Choice
Reader's Journey
Reading Excellence: Word Attack and Development Strategies (REWARDS)

Reading Horizons
Reading in the Content Areas
Reading is FAME
Reading Power in the Content Areas
Reading Plus
Reading with Purpose
Reciprocal Teaching
Rosetta Stone Literacy
Saxon Phonics
Scaffolded Reading Experience
Scott Foresman
Second Chance at Literacy Learning
Second Chance Reading
Slingerland
Soar to Success
Soliloquy Reading Assistant
Sound Sheet
Spalding Method
Spell Read P.A.T.
Strategic Literacy Initiative
SuccessMaker
Supported Literacy Approach
Text Mapping Strategy
Thinking Reader
Thinking Works
Transactional Strategies Instruction
Vocabulary Improvement Program
Voyager TimeWarp Plus
Wilson Reading System
Wisconsin Design for Reading Skills Development (WDRSD)
Write to Learn

References

ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City, IA: Author.

American Diploma Project. (2004). *Ready or not: Creating a high school diploma that counts*. Washington, DC: Achieve, Inc.

Au, K.H. (2000). A multicultural perspective on policies for improving literacy achievement: Equity and excellence. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 835–851). Mahwah, NJ: Erlbaum.

Balfanz, R., Legters, N., & Jordan, W. (2004, April). *Catching Up: Impact of the Talent Development ninth grade instructional interventions in reading and mathematics in high-poverty high schools* (Tech. Rep. No. 69). Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Barton, M.L., Heidema, C., & Jordan, D. (2002). Teaching reading in mathematics and science. *Educational Leadership*, *60*(3), 24–28.

Biancarosa, G., & Snow, C.E. (2006). *Reading next: A vision for action and research in middle and high school literacy. A report from Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Borman, G.D., Hewes, G.M., Overman, L.T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*(2), 125–230.

Caggiano, J.A. (2007). *Addressing the learning needs of struggling adolescent readers: The impact of a reading intervention program on students in a middle school setting*. Unpublished doctoral dissertation, The College of William and Mary, Williamsburg, VA.

Calhoon, M.B. (2005). Effects of a peer-mediated phonological skill and reading comprehension program on reading skill acquisition for middle school students with reading disabilities. *Journal of Learning Disabilities*, *38*(5), 424–433.

Chamberlain, A., Daniels, C., Madden, N.A., & Slavin, R.E. (2007). A randomized evaluation of the Success for All Middle School reading program. *Middle Grades Research Journal*, *2*(1), 1–21.

Chambers, E.A. (2003). *Efficacy of educational technology in elementary and secondary classrooms: A meta-analysis of the research literature from 1992–2002*. Unpublished doctoral dissertation, Southern Illinois University at Carbondale.

Cheung, A., & Slavin, R.E. (2005). Effective reading programs for English language learners and other language-minority students. *Bilingual Research Journal*, *29*(2), 241–267.

Chiang, A., Stauffer, C., & Cannara, A. (1978). *Demonstration of the use of computer-assisted instruction with handicapped children: Final report*. (ERIC Document Reproduction Service No. ED166913).

Comprehensive School Reform Quality Center (2006, October). *CSRQ Center report on middle and high school comprehensive school reform models*. Washington, DC: American Institutes for Research.

Cooper, H. (1998). *Synthesizing research: A Guide for Literature Reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Deshler, D.D., Palincsar, A.S., Biancarosa, G., & Nair, M. (2007). *Informed choices for struggling adolescent readers: A research-based guide to instructional programs and practices*. Newark, DE: International Reading Association.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007, March). *Effectiveness of reading and mathematics software products: Findings from the first student cohort* (NCEE Rep. No. 2007-4005). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Fuchs, D., Fuchs, L.S., Mathes, P.G., & Simmons, D.C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*(1), 174–206.

Fuchs, L.S., Fuchs, D., & Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial and Special Education*, *20*(5), 309–318.

Gaskins, I.W. (1994). Classroom applications of cognitive science: Teaching poor readers how to learn, think, and problem solve. In K.

McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 129–154). Cambridge, MA: MIT Press.

Hagerman, T.E. (2003). *A quasi-experimental study on the effects of Accelerated Reader at middle school*. Unpublished doctoral dissertation, University of Oregon, Eugene.

Hankinson, R.D., & Myers, D.L. (2000). *Effectiveness of the Middle School PALS in Reading program on the reading comprehension of middle school students*. Unpublished doctoral dissertation, Duquesne University, Pittsburgh, PA.

Haslam, M.B., White, R.N., & Klinge, A. (2006, May). *Improving student literacy: READ 180 in the Austin Independent School District 2004–05*. Washington, DC: Policy Studies Associates.

Hasselbring, T.S., & Goin, L.I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading & Writing Quarterly*, *20*(2), 123–144.

Herlihy, C.M., & Kemple, J.J. (2004, December). *The Talent Development Middle School model: Context, components, and initial impacts on students' performance and attendance*. New York: MDRC.

Herlihy, C.M., & Kemple, J.J. (2005, August). *The Talent Development Middle School model: Impacts through the 2002–2003 school year. An update to the December 2004 report*. New York: MDRC.

Hunter, C.T.L. (1994). *A study of the effect of instructional method on the reading and mathematics achievement of Chapter One students in rural Georgia*. Unpublished doctoral dissertation, South Carolina State University, Orangeburg.

Interactive, Inc. (2002, January). *Final report: Study of READ 180 in the Council of Great City Schools*. New York: Author.

Joftus, S. (2002, September). *Every child a graduate: A framework for an excellent education for all middle and high school students*. Washington, DC: Alliance for Excellent Education.

Joftus, S., & Maddox-Dolan, B. (2003, April). *Left out and left behind: NCLB and the American high school*. Washington, DC: Alliance for Excellent Education.

Johnson, J., Haslam, M., & White, R. (2006). *Improving student literacy in the Phoenix Union High School District, 2005–06*. Washington, DC: Policy Studies Associates.

Kemple, J., Corrin, W., Nelson, E., Salinger, T., Herrmann, S., Drummond, K., et al. (2008, January). *The enhanced reading opportunities study: Early impact and implementation findings* (NCEE 2008-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Kemple, J.J., Herlihy, C.M., & Smith, T.J. (2005, May). *Making progress toward graduation: Evidence from the Talent Development High School model*. New York: MDRC.

Kulik, J.A. (2003, May). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say. Final Report* (SRI Project No. P10446.001). Arlington, VA: SRI International.

Lee, J., Grigg, W., & Donahue, P. (2007). *The nation's report card: Reading 2007* (NCES 2007-496). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Liston, W.R. (1991). *The effects of computer-assisted instruction on remedial reading students' achievement in grade 10 identified South Carolina high schools as measured by BSAP state testing in school years 1988–89 and 1989–90*. Unpublished doctoral dissertation, University of South Carolina, Columbia.

Losh, M.A. (1991). *The effect of the strategies intervention model on the academic achievement of junior high learning-disabled students*. Unpublished doctoral dissertation, University of Nebraska–Lincoln.

MacIver, D.J., Balfanz, R., Ruby, A., Byrnes, V., Lorentz, S., & Jones, L. (2004). Developing adolescent literacy in high poverty middle schools: The impact of Talent Development's reforms across multi-

ple years and sites. In P.R. Pintrich & M.L. Maehr (Eds.), *Motivating students, improving schools, Vol. 13: The legacy of Carol Midgley, Advances in motivation and achievement* (pp. 185–207). Amsterdam: Elsevier.

Metrics Associates. (1981). *Evaluation of the Computer Assisted Instruction Title I Project, 1980–81. Research report*. Chelmsford, MA: Merrimack Education Center.

Mims, C., Lowther, D., Strahl, J.D., & Nunnery, J. (2006). *Little Rock School District READ 180 evaluation: Technical report*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Mothus, T.G. (1997). The effects of strategy instruction on the reading comprehension achievement of junior secondary school students. *Masters Abstracts International*, *42* (01), 44.

Murphy, R.F., Penuel, W.R., Means, B., Korbak, C., Whaley, A., & Allen, J.E. (2002, April). *E-DESK: A review of recent evidence on the effectiveness of discrete educational software* (SRI Project No. 11063). Menlo Park, CA: SRI International.

Nave, J. (2007). *An assessment of READ 180 regarding its association with the academic achievement of at-risk students in Sevier County schools*. Unpublished doctoral dissertation, East Tennessee State University, Johnson City, TN.

Papalewis, R. (2004). Struggling middle school readers: Successful, accelerating intervention. *Reading Improvement*, *41*(1), 24–37.

Rohrbeck, C.A., Ginsburg-Block, M.D., Fantuzzo, J.W., & Miller, T.R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*(2), 240–257.

Ross, S.M., & Nunnery, J.A. (2005, January). *The effect of School Renaissance on student achievement in two Mississippi school districts*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S., Nunnery, J., Avis, A., & Borek, T. (2005, July). *The effects of School Renaissance on student achievement in two Mississippi school districts: A longitudinal quasi-experimental study*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Rothstein, H.R., Sutton, A.J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, West Sussex, England: John Wiley.

Roy, J.W. (1993). *An investigation of the efficacy of computer-assisted mathematics, reading, and language arts instruction*. Unpublished doctoral dissertation, Baylor University, Waco, TX.

Schumaker, J.B., Denton, P.H., & Deshler, D.D. (1984). *The paraphrasing strategy*. Lawrence, KS: University of Kansas, Center for Research on Learning.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*(2), 309–316.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shneyderman, A. (2006). *Some results of the Voyager Passport Reading Intervention System in several school districts*. Miami, FL: Miami-Dade County Public Schools Office of Evaluation and Research.

Slavin, R.E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, *15*(9), 5–11.

Slavin, R.E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Boston: Allyn & Bacon.

Slavin, R.E. (2008). What works? Issues in synthesizing education program evaluations. *Educational Researcher*, *37*(1), 5–14.

Slavin, R.E. (in press). Cooperative learning. In G. McCulloch & D. Crook (Eds.), *The Routledge International Encyclopedia of Education*. Abington, UK: Routledge.

Slavin, R.E., Chamberlain, A., Daniels, C., & Madden, N.A. (2008, March). *The Reading Edge: A randomized evaluation of a middle school cooperative reading program*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Slavin, R.E., Daniels, C., & Madden, N.A. (2005). "Success for All" middle schools add content to middle grades reform. *Middle School Journal*, *36*(5), 4–8.

Slavin, R.E., & Lake, C. (in press). Effective programs in elementary math: A best evidence synthesis. *Review of Educational Research*.

Slavin, R.E., Lake, C., & Groff, C. (2007). *Effective programs in middle and high school math: A best evidence synthesis*. Manuscript submitted for publication.

Sterne, J.A.C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, *53*(11), 1119–1129.

Stevens, R.J., & Durkin, S. (1992, September). *Using student team reading and student team writing in middle schools: Two evaluations* (Report No. 36). Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Stevens, R.J., Madden, N.A., Slavin, R.E., & Farnish, A.M. (1987). Cooperative integrated reading and composition: Two field experiments. *Reading Research Quarterly*, *22*(4), 433–454.

Taylor, S., & Tweedie, R. (1998). *A non-parametric "trim and fill" method of assessing publication bias in meta-analysis*. Denver, CO: University of Colorado Health Sciences Center.

Texas Center for Educational Research. (2007, May). *Evaluation of the Texas Technology Immersion Pilot: Findings from the second year*. Austin, TX: Author.

Webb, N.M., & Palincsar, A.S. (1996). Group processes in the classroom. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 841–873). New York: Simon & Schuster Macmillan.

What Works Clearinghouse. (2007). *Beginning reading*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved March 17, 2008, from ies.ed.gov/ncee/wwc/reports/beginning_reading/topic

White, R.N., Haslam, M.B., & Hewes, G.M. (2006, July). *Improving student literacy in the Phoenix Union High School District 2003–04 and 2004–05. Final Report*. Washington, DC: Policy Studies Associates.

Woods, D.E. (2007). *An investigation of the effects of a middle school reading intervention on school dropout rates*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Robert E. Slavin is Director of the Center for Research and Reform in Education, Johns Hopkins University, Baltimore, Maryland, USA and Director of the Institute for Effective Education, University of York, England, UK; e-mail rslavin@successforall.org or rs553@york.ac.uk.

Alan Cheung is an Associate Professor at The Hong Kong Institute of Education, New Territories, Hong Kong; e-mail ckcheung@ied.edu.hk.

Cynthia Groff is currently pursuing a Ph.D. at the University of Pennsylvania, Philadelphia, USA; e-mail cgroff@dolphin.upenn.edu.

Cynthia Lake is an Instructor at Johns Hopkins University, Baltimore, Maryland, USA; e-mail clake5@jhu.edu.

Appendix
Studies Not Included in the Review

Program	Study	Reason for exclusion
	Reading curricula	
Corrective Reading	Airhart, K. (2005). *The effectiveness of direct instruction in reading compared to a state-mandated language arts curriculum for ninth and tenth graders with specific learning disabilities*. Unpublished doctoral dissertation, Tennessee State University, Nashville, TN.	Pretest differences > 0.5 *SD* on TORC-3
	Grossen, B., Hagen-Burke, S., & Burke, M.D. (2002). *An experimental study of the effects of considerate curricula in language arts on reading comprehension and writing* (Research Rep. No. 13). Lawrence, KS: University of Kansas, Institute for Academic Access.	Duration < 12 weeks
	Harris, R.E., Marchand-Martella, N., & Martella R.C. (2000). Effects of a peer-delivered Corrective Reading program. *Journal of Behavioral Education*, *10*(1), 21–36.	No control group
	Kalisek, A.M. (2004). *The effects of a middle school Corrective Reading intervention on high school passage rate*. Unpublished doctoral dissertation, University of La Verne, La Verne, CA.	Inadequate outcome measure
	Kasendorf, S.J., & McQuaid, P. (1987). Corrective Reading evaluation study. *ADI News*, *7*(1), 9.	No control group
	Lingo, A.S., Slaton, D.B., & Jolivette, K. (2006). Effects of corrective reading on the reading abilities and classroom behaviors of middle school students with reading deficits and challenging behavior. *Behavioral Disorders*, *31*(3), 265–283.	Multiple probe design; seven participants
	Shippen, M.E., Houchins, D.E., Steventon, C., & Sartor, D. (2005). A comparison of two direct instruction reading programs for urban middle school students. *Remedial and Special Education*, *26*(3), 175–182.	Duration < 12 weeks
	Sommers, J. (1995). Seven-year overview of Direct Instruction programs used in basic skills classes at Big Piney Middle School. *Effective School Practices*, *14*(4), 29–32.	No control group
	Strong, A.C., Wehby, J.H., Falk, K.B., & Lane, K.L. (2004). The impact of a structured reading curriculum and repeated reading on the performance of junior high students with emotional and behavioral disorders. *School Psychology Review*, *33*(4), 561–581.	Multiple baseline design; six participants
	Thorne, M.T. (1978). "Payment for reading": The use of the "Corrective Reading Scheme" with junior maladjusted boys. *Remedial Education*, *13*(2), 87–90.	No control group
Fluent Reader	Raile, C., & Seekal, P. (2004). *Curriculum-based measurements show improved fluency after only 12 weeks (Scientific Research: Quasi-Experimental series)*. Madison, WI: Renaissance Learning, Inc.	Inadequate control group; lower-ability group received treatment
LANGUAGE!	Greene, J.F. (1996). LANGUAGE! Effects of an individualized structured language curriculum for middle and high school students. *Annals of Dyslexia*, *46*, 97–121.	Inadequate control group
	Lawrence, A.J. (2003). *The effectiveness of the "Language!" program in improving the word recognition skills of middle school students with learning disabilities*. Unpublished master's thesis, California State University, Fullerton.	Duration < 12 weeks
Read XL	Holly, T.M. (2004). *Analyzing the effectiveness of reading intervention strategies on reading achievement in an urban West Tennessee school district*. Unpublished doctoral dissertation, Union University, Jackson, TN.	Inadequate outcome measure

(continued)

Program	Study	Reason for exclusion
	Computer-assisted instruction	
Accelerated Reader	Gibson, M.T. (2002). *An investigation of the effectiveness of the Accelerated Reader program used with middle school at-risk students in a rural school system*. Unpublished doctoral dissertation, Mississippi State University, Mississippi State.	Inadequate control group
	Goodman, G. (1999). *The Reading Renaissance/Accelerated Reader program: Pinal County school-to-work evaluation report*. Phoenix, AZ: Creative Research Associates.	No control group
	Kohel, P.R. (2003). *Using Accelerated Reader: Its impact on the reading levels and Delaware state testing scores of 10th grade students in Delaware's Milford High School*. Unpublished doctoral dissertation, Wilmington College.	No adequate control group; STAR pretest differences > 0.5 *SD*
	Lewis, S.C.S. (2005). *Evaluating alternative methodologies to teaching reading to sixth-grade students and the association with student achievement*. Unpublished doctoral dissertation, East Tennessee State University, Johnson City.	No pretests for Terra Nova
	McDurmon, A. (2001). *The effects of guided and repeated reading on English language learners*. Unpublished master's thesis, Berry College, Mount Berry, GA.	Outcome measure (STAR) inherent to treatment
	Nunnery, J.A., Ross, S.M., & Goldfeder, E. (2003, June). *The effect of School Renaissance on TAAS scores in the McKinney ISD*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.	Ceiling effect on TAAS
	Nunnery, J.A., Ross, S.M., & McDonald, A. (2006). A randomized experimental evaluation of the impact of Accelerated Reader/Reading Renaissance implementation on reading achievement in grades 3 to 6. *Journal of Education for Students Placed at Risk*, *11*(1), 1–18.	STAR Reading assessment inherent to treatment
	Paul, T.D. (2003). *Guided independent reading: An examination of the Reading Practice Database and the scientific research supporting guided independent reading as implemented in Reading Renaissance*. Wisconsin Rapids, WI: Renaissance Learning.	No control group
	Peak, J., & Dewalt, M.W. (1994). Reading achievement: Effects of computerized reading management and enrichment. *ERS Spectrum*, *12*(1), 31–34.	Insufficient information
	Scott, L.S. (1999). *The Accelerated Reader program, reading achievement, and attitudes of students with learning disabilities*. Unpublished master's thesis, Georgia State University, Atlanta. (ERIC Document Reproduction Service No. ED434431).	Inadequate control group; large pretest differences between groups
	Sims, S.P. (2002). *The effects of the Accelerated Reader program and sustained silent reading on reading attitudes and reading achievement of eighth-grade students*. Unpublished doctoral dissertation, Georgia State University, Atlanta.	Inadequate outcome measure
	Smith, I. (2005). *Can Accelerated Reader and cooperative learning enhance the reading achievement of Level 1 high school students on the Florida Comprehensive Assessment Test?* Unpublished doctoral dissertation, Nova Southeastern University, Fort Lauderdale-Davie, FL.	No control group
	Topping K.J., & Sanders, W.L. (2000). Teacher effectiveness and computer assessment of reading: Relating value added and learning information system data. *School Effectiveness and School Improvement*, *11*(3), 305–337.	No control group
	Vollands, S.R., Topping, K.J., & Evans, R.M. (1999). Computerized self-assessment of reading comprehension with the Accelerated Reader: Action research. *Reading & Writing Quarterly*, *15*(3), 197–211.	Large pretest differences
	Walberg, H.J. (2001). *Final evaluation of the reading initiative: Report to the J.A. & Kathryn Albertson Foundation Board of Directors*. Boise, ID: J.A. & Kathryn Albertson Foundation. Retrieved March 17, 2008, from jkaf.org/system/files/readevw.pdf	Program evaluations; insufficient data presented

(continued)

Program	Study	Reason for exclusion
		Computer-assisted instruction
	Walker, G.A. (2005). *The impact of Accelerated Reader on the reading levels of eighth-grade students at Delaware's Milford Middle School*. Unpublished doctoral dissertation, Wilmington College.	No untreated control group
Jostens	CompassLearning. (2005). *CompassLearning School Effectiveness Report: Daniel Boone Area School District, Birdsboro, Pennsylvania*. San Diego, CA: CompassLearning.	No control group
Failure Free Reading	Algozzine, B., Lockavitch, J.F., & Audette, R. (1997). Effects of Failure Free Reading on students at-risk for serious school failure. *Australian Journal of Learning Disabilities*, *2*(3), 14–17.	No control group
	Gum, L.I. (2003). *Collateral effects of computer-assisted reading instruction on the classroom behaviors of learners with emotional and/or behavioral disorders*. Unpublished doctoral dissertation, Tennessee Technological University, Cookeville.	Multiple baseline design; eight participants
	Rankhorn, B., England, G., Collins, S.M., Lockavitch, J.F., & Algozzine, B. (1998). Effects of the *failure free* reading program on students with severe reading disabilities. *Journal of Learning Disabilities*, *31*(3), 307–312.	No control group
	Slate, J., Algozzine, B., & Lockavitch, J.F. (1998). Effects of intensive remedial reading instruction. *Journal of At-Risk Issues*, *5*(1), 30–35.	No control group
Fast ForWord	Scientific Learning Corporation. (2006). Improved reading skills by students in Pocatello/Chubbuck School District #25 who used Fast ForWord® products. *MAPS for Learning: Educator Reports*, *10*(25), 1–5.	Pretest differences > 0.5 *SD*
	Scientific Learning Corporation. (2006). Improved reading skills by students in Washington local schools who used Fast ForWord® products. *MAPS for Learning: Educator Reports*, *11*(32), 1–6.	Duration < 12 weeks
	Scientific Learning Corporation. (2007). Improved reading skills by students in the South Euclid-Lyndhurst School District who used Fast ForWord® products. *MAPS for Learning: Educator Reports*, *11*(28), 1–5.	Inadequate control group
	Scientific Learning Corporation. (2007). Improved reading skills by students in Warren County schools who used Fast ForWord® products. *MAPS for Learning: Educator Reports*, *11*(29), 1–4.	No adequate comparison group
	Sharp, M.V.T. (2007). *An evaluation of the Fast ForWord program in the Christina School District*. Unpublished doctoral dissertation, University of Delaware.	No control group
Merit	Jones, J.D., Staats, W.D., Bowling, N., Bickel, R.D., Cunningham, M.L., & Cadle, C. (2004/2005). An evaluation of the Merit Reading Software Program in the Calhoun County (WV) Middle/High School. *Journal of Research on Technology in Education*, *37*(2), 177–195.	Duration < 12 weeks
MultiFunk	Fasting, R.B., & Lyster, S.-A.H. (2005). The effects of computer technology in assisting the development of literacy in young struggling readers and spellers. *European Journal of Special Needs Education*, *20*(1), 21–40.	Duration < 12 weeks
Peabody Literacy Lab	Hasselbring, T.S., & Goin, L.I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading & Writing Quarterly*, *20*(2), 123–144.	Inadequate control group; pretest differences > 0.5 *SD*
PLATO	Barnett, T.L. (1986). A comparative analysis of the PLATO computer-assisted instructional delivery system and the traditional individualized instructional program in two juvenile correctional facilities owned by the Commonwealth of Pennsylvania. *Dissertation Abstracts International*, *46* (09), 2668A. (UMI No. 8525658)	Duration < 12 weeks
	Brush, T. (2002, May). *PLATO evaluation series: Terry High School, Lamar Consolidated ISD, Rosenberg, TX*. Bloomington, MN: PLATO Learning. (ERIC Document Reproduction Service No. ED 469375)	No control group
	Elliott, E.L.L. (1985). The effects of computer-assisted instruction upon the basic skill proficiencies of secondary vocational education students. *Dissertation Abstracts International*, *46*(11), 3329A. (UMI No. 8600439)	Large pretest differences (> 0.5 *SD*) in reading and math

(continued)

Program	Study	Reason for exclusion
	Computer-assisted instruction	
Quicktionary Reading Pen II	Higgins, E.L., & Raskind, M.H. (2005). The compensatory effectiveness of the Quicktionary Reading Pen II on the reading comprehension of students with learning disabilities. *Journal of Special Education Technology*, *20*(1), 31–40.	No pretest; duration < 12 weeks
READ 180	Admon, N. (2003). *READ 180 Stages A and B: Iredell-Statesville schools, North Carolina*. New York: Scholastic.	No control group
	Admon, N. (2005). *READ 180 Stage B: St. Paul School District, Minnesota*. New York: Scholastic.	No control group
	Brown, S.H. (2006). *The effectiveness of READ 180 intervention for struggling readers in grades 6–8*. Unpublished doctoral dissertation, Union University, Jackson, TN.	No control group
	Campbell, Y.C. (2006). *Effects of an integrated learning system on the reading achievement of middle school students*. Unpublished doctoral dissertation, University of Miami, Coral Gables, FL.	Inadequate control group; pretest differences > 0.5 *SD*
	Daviess County Public Schools, Assessment, Research and Curriculum Department. (2005). *READ 180 implementation year study*. Owensboro, KY: Author.	No control group
	Denman, J.S. (2004). *Integrating technology into the reading curriculum: Acquisition, implementation, and evaluation of a reading program with a technology component (READ 180) for struggling readers*. Newark, DE: University of Delaware.	No control group
	Dunn, C.A. (2002). *An investigation of the effects of computer-assisted reading instruction versus traditional reading instruction on selected high school freshmen*. Unpublished doctoral dissertation, Loyola University of Chicago.	Inadequate control group; pretest differences > 0.5 *SD*
	Ferguson, J.M. (2005). *The implementation of technology in reading classrooms and the impact of technology integration and student perceptions on reading achievement*. Unpublished doctoral dissertation, Texas A&M University, Commerce, TX.	No control group
	Gentry, L. (2006). *An evaluation of READ 180 in an urban secondary school*. Unpublished doctoral dissertation, American University, Washington, DC.	Inadequate control group; large pretest differences
	Goin, L., Hasselbring, T., & McAfee, I. (2004). *Executive summary, DoDEA/Scholastic READ 180 project: An evaluation of the READ 180 intervention program for struggling readers*. New York: Scholastic.	No control group
	Hasselbring, T.S., Goin, L., Taylor, R., Bottge, B., & Daley, P. (1997). The computer doesn't embarrass me. *Educational Leadership*, *55*(3), 30–33.	Descriptive article
	Hewes, G.M., Palmer, N., Haslam, M.B., & Mielke, M.B. (2006). *Five years of READ 180 in Des Moines: Improving literacy among middle school and high school special education students*. New York: Scholastic.	Inadequate control group
	Holyoke School District. (2005). *READ 180 Stage B: Holyoke School District, Massachusetts*. New York: Scholastic.	No control group
	Kratofil, M.D. (2006). *A comparison of the effect of Scholastic READ 180 and traditional reading interventions on the reading achievement of middle school low-level readers*. Unpublished master's thesis, Central Missouri State University, Warrensburg.	Inadequate control group; pretest differences > 0.5 *SD*
	Newman, D., Leuer, M., & Jaciw, A. (2006). *Effectiveness of Scholastic's READ 180 as a remedial reading program for ninth graders: Report of an implementation in Anaheim, CA*. Palo Alto, CA: Empirical Education.	No control group
	Palmer, N. (2003). *READ 180 middle-school study: Des Moines, Iowa, 2000–2002. Research report.* New York: Scholastic.	No control group
	Papalewis, R., & Scholastic Research and Evaluation Department. (2003, December). *Final Report: A study of READ 180 in middle schools in Clark County School District, Las Vegas, Nevada*. New York: Scholastic.	No control group

(continued)

Program	Study	Reason for exclusion
	Computer-assisted instruction	
	Pearson, L.M., & White, R.N. (2004, June). *Study of the impact of READ 180 on student performance in Fairfax County Public Schools*. New York: Scholastic.	No control group
	Scholastic Research and Evaluation Department. (2004, June). *Final report: A study of READ 180 at Shiprock High School in Central Consolidated School District on the Navajo Indian Reservation, New Mexico*. New York: Scholastic.	No control group
	Scholastic Research and Evaluation Department. (2005). *Special education students: Selbyville Middle and Sussex Central Middle Schools, Indian River School District (Delaware)*. New York: Scholastic.	No control group
	Thomas, D.M. (2005). *Examining the academic and motivational outcomes of students participating in the READ 180 program*. Unpublished doctoral dissertation, University of Kentucky, Lexington.	Pretest equivalence not established
	Thomas, J. (2003). *Reading program evaluation: READ 180, grades 4–8, November, 2003*. Kirkwood, MO: Kirkwood School District.	No control group
	White, R.N., Williams, I.J., & Haslem, M.B. (2005, June). *Performance of District 23 students participating in Scholastic READ 180*. Washington, DC: Policy Studies Associates.	Pretest differences > 0.5 *SD*
	Witkowski, P.M. (2004). *A comparison study of two intervention programs for reading-delayed high school students*. Unpublished doctoral dissertation, University of Missouri–Saint Louis.	Inadequate control group
	Zvoch, K., & Letourneau, L. (2006). *Closing the achievement gap: An examination of the status and growth of ninth grade READ 180 students*. Las Vegas, NV: Clark County School District.	No control group
Reading Partner	Salomon, G., Globerson, T., & Guterman, E. (1989). The computer as a zone of proximal development: Internalizing reading-related metacognitions from a reading partner. *Journal of Educational Psychology*, *81*(4), 620–627.	Duration < 12 weeks
Reading Plus	Marrs, H., & Patrick, C. (2002). A return to eye-movement training? An evaluation of the Reading Plus program. *Reading Psychology*, *23*(4), 297–322.	Inadequate control group
Reading Renaissance	Renaissance Learning. (2002). *Results from a three-year statewide implementation of Reading Renaissance in Idaho*. Madison, WI: Author.	Inadequate control group
Roland Reading Method	Hardiman, M.M. (2004). *Teaching adolescents with reading deficits: The effects of a phonics-based approach*. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore, MD.	Inadequate control group; large differences on free lunch % and some pretests
Student Assistant for Learning from Text (SALT)	MacArthur, C.A., & Haynes, J.B. (1995). Student Assistant for Learning from Text (SALT): A hypermedia reading aid. *Journal of Learning Disabilities*, *28*(3), 150–159.	Duration < 12 weeks
SuccessMaker & talking books	Underwood, J.D.M. (2000). A comparison of two types of computer support for reading development. *Journal of Research in Reading*, *23*(2), 136–148.	Insufficient information on pretest scores
Other computer-assisted instruction programs	Arroyo, C. (1992). *What is the effect of extensive use of computers on the reading achievement scores of seventh grade students?* (ERIC document Reproduction Service No. ED353544)	Insufficient information
	Cicchetti, G., Sandagata, A., Suntag, M., & Tarnuzzer, J. (2003). *The effects of web-based instruction in digital classrooms on math and reading performance on the CT Academic Performance test (CAPT) and related outcomes for a 10th grade cohort of CT urban vocational-technical school students*. Providence, RI: Brown University.	No control group
	Gentry, M.M., Chinn, K.M., & Moulton, R.D. (2004/2005). Effectiveness of multimedia reading materials when used with children who are deaf. *American Annals of the Deaf*, *149*(5), 394–403.	Inadequate control group
	Kim, A.-H. (2002). *Effects of computer-assisted collaborative strategic reading on reading comprehension for high-school students with learning disabilities*. Unpublished doctoral dissertation, University of Texas at Austin.	Inadequate control group; large differences on % free lunch

Program	Study	Reason for exclusion
	Computer-assisted instruction	
	Kim, A.-H., Vaughn, S., Klingner, J.K., Woodruff, A.L., Reutebuch, C.K., & Kouzekanani, K. (2006). Improving the reading comprehension of middle school students with disabilities through computer-assisted collaborative strategic reading. *Remedial and Special Education*, *27*(4), 235–249.	Average duration < 12 weeks
	Koza, J.L. (1989). *Comparison of the achievement of mathematics and reading levels and attitude toward learning of high-risk secondary students through the use of computer-aided instruction*. Unpublished doctoral dissertation, University of Minnesota.	Duration < 12 weeks; inadequate control group
	Kramarski, B., & Feldman, Y. (2000). Internet in the classroom: Effects on reading comprehension, motivation and metacognitive awareness. *Education Media International*, *37*(3), 149–155.	Duration < 12 weeks
	Lynch, L., Fawcett, A.J., & Nicolson, R.I. (2000). Computer-assisted reading intervention in a secondary school: An evaluation study. *British Journal of Educational Technology*, *31*(4), 333–348.	Duration < 12 weeks
	Reinking, D. (1988). Computer-mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. *Reading Research Quarterly*, *23*(4), 484–498.	Duration < 12 weeks
	Traynor, P.L. (2003). Effects of computer-assisted instruction on different learners. *Journal of Instructional Psychology*, *30*(2), 137–143.	No control group
	Instructional-process programs	
AMP Reading System	Mid-continent Research for Education and Learning. (n.d.). *Final Evaluation Report, AGS Globe's AMP Reading System Efficacy Study*. Denver, CO: Author.	Inadequate control group; pretest differences > 0.5 *SD*
BIG Accommodation Model	Grossen, B.J. (2002). The BIG Accomodation Model: The Direct Instruction model for secondary schools. *Journal of Education for Students Placed at Risk*, *7*(2), 241–263.	Inadequate control groups
Career Academies	Elliot, M.N., Hanser, L.M., & Gilroy, C.L. (2002). Career academies: Additional evidence of positive student outcomes. *Journal of Education for Students Placed at Risk*, *7*(1), 71–90.	Inadequate outcome measure
Classwide Peer Tutoring	Neddenriep, C.E. (2003). *Classwide peer tutoring: Three experiments investigating the generalized effects of increased oral reading fluency to silent reading comprehension*. Unpublished doctoral dissertation, University of Tennessee, Knoxville.	Duration < 12 weeks
	Stevens, M.L. (1998). *Effects of classwide peer tutoring on the classroom behavior and academic performance of students with ADHD*. Unpublished doctoral dissertation, Alfred University, Alfred, NY.	Duration < 12 weeks
	Veerkamp, M.B. (2001). *The effects of Classwide Peer Tutoring on the reading achievement of urban middle school students*. Unpublished doctoral dissertation, University of Kansas, Lawrence.	Inadequate outcome measure
Corrective Reading	Grossen, B. (2004). Success of a direct instruction model at a secondary level school with high-risk students. *Reading & Writing Quarterly*, *20*(2), 161–178.	No control group
Content Reading in Secondary Schools (CRISS)	Allen, R. (2000, Summer). Before it's too late: Giving reading a last chance. *ASCD Curriculum Update*, pp. 1–3, 6–8.	Inadequate outcome measure; uncertain validity and reliability
	Havens, L. (1993). *Project CRISS: Reading, writing, and studying strategies for literature and content*. Kalispell, MT: Project CRISS.	Inadequate information on outcome measure validity
	Pearson, J.W., & Santa, C.M. (1995). Students as researchers of their own learning. *Journal of Reading*, *38*(6), 462–469.	Inadequate outcome measure; uncertain validity and reliability
	Santa, C.M. (2004, January). *Project CRISS: Evidence of effectiveness*. Kalispell, MT: Project CRISS.	Inadequate information on outcome measure validity
Direct Instruction Corrective Reading Program	Maggs, A., & Murdoch, R. (1979). Teaching low performers in upper primary and lower secondary to read by direct instruction methods. *Reading Education*, *4*(1), 35–39.	No control group

(continued)

Program	Study	Reason for exclusion
	Instructional-process programs	
Exemplary Center for Reading Instruction (ECRI)	Reid, E.R. (1996). *Exemplary center for reading instruction (ECRI) validation study*. Salt Lake City, UT: Reid Foundation.	One study with control group but pretest differences > 0.5 *SD*
Fluent Reader	Palumbo, T.J. (2004). *Effects of the Fluent Reader program on reading performance*. Unpublished master's thesis, University of Minnesota. Retrieved March 17, 2008, from www.tc.umn.edu/~samue001/papers.htm	Duration < 12 weeks
Great Leaps	Dudley, A.M. (2005). *Effects of two fluency methods on the reading performance of secondary students*. Unpublished doctoral dissertation, University of Arizona, Tucson.	Inadequate control group
	Mercer, C.D., Campbell, K.U., Miller, M.D., Mercer, K.D., & Lane, H.B. (2000). Effects of a reading fluency intervention for middle schoolers with specific learning disabilities. *Learning Disabilities Research & Practice*, *15*(4), 179–189.	Inadequate control group
	Pruitt, B.A. (2000). *The effects of "Great Leaps Reading" on the reading fluency of students served in special education*. Unpublished doctoral dissertation, University of Kentucky, Lexington.	No control group
Intensive Reading Strategies Instruction (IRSI)	Seybert, L.G. (1998). *The development and evaluation of a model of intensive reading strategies instruction for teachers in inclusive, secondary-level classrooms*. Unpublished doctoral dissertation, University of Kansas, Lawrence.	Inadequate control group; pretest differences > 0.5 *SD*
Multicultural Reading and Thinking (McRAT)	Hoskyn, J.J. (1994). *Multicultural reading and thinking: A three year report—1989–92*. (ERIC Document Reproduction Service No. ED380416)	No reading outcomes
	Hoskyn, J.J., et al. (1993, April). *Multicultural reading and thinking program (McRAT)*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.	Inadequate outcome measure; writing, not reading
Lindamood-Bell	Kennedy, K.M., & Backman, J. (1993). Effectiveness of the Lindamood Auditory Discrimination in Depth program with students with learning disabilities. *Learning Disabilities Research & Practice*, *8*(4), 253–259.	Duration < 12 weeks
Pathway Project	Olson, C.B., & Land, R. (2007). A cognitive strategies approach to reading and writing instruction for English Language Learners in secondary school. *Research in the Teaching of English*, *41*(3), 269–303.	Inadequate control group
Phonological Analysis and Blending/ Direct Instruction (PHAB/DI), Western Institute for Science and Technology (WIST)	Lovett, M.W., Lacerenza, L., Borden, S.L., Frijters, J.C., Steinbach, K.A., & De Palma, M. (2000). Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology*, *92*(2), 263–283.	Inappropriate control group (not studying reading)
Phono-Graphix	Endress, S.A., Weston, H., Marchand-Martella, N.E., Martella, R.C., & Simmons, J. (2007). Examining the effects of Phono-Graphix on the remediation of reading skills of students with disabilities: A program evaluation. *Education and Treatment of Children*, *30*(2), 1–20.	Duration < 12 weeks
	McGuinness, C., McGuinness, D., & McGuinness, G. (1996). Phono-Graphix™: A new method for remediating reading difficulties. *Annals of Dyslexia*, *46*, 73–96.	No control group
Read Now	Algozzine, B. (2004). Effects of Read Now on adolescents at risk for school failure. *Journal of At-Risk Issues*, *10*(2), 1–8.	Duration < 12 weeks; STAR Reading assessment inherent to treatment
Read Right	Green, J. (1998). *Project Report: READ RIGHT Juvenile Detention Pilot Project, Mission Creek Youth Camp, Belfair, Washington*. Shelton, WA: Read Right Systems.	No control group
	Litzenberger, J. (2001). *Reading research results: WASL 2001, Using READ RIGHT as an intervention program for at-risk 10th graders. Final report prepared for Read Right Systems and Kent School District*. Shelton, WA: Read Right.	Insufficient information

(continued)

Program	Study	Reason for exclusion
		Instructional-process programs
	Mercer, C.D., Campbell, K.U., Miller, M.D., Mercer, K.D., & Lane, H.B. (2000). Effects of a reading fluency intervention for middle schoolers with specific learning disabilities. *Learning Disabilities Research & Practice*, *15*(4), 179–189.	Inadequate control group; compares groups receiving the treatment for different durations
Reading Apprenticeship	Greenleaf, C.L., & Mueller, F.L. (with Cziko, C.). (1997, September). *Impact of the Pilot Academic Literacy Course on ninth grade students' reading development: Academic year 1996–1997. A report to the Stuart Foundations*. San Francisco, CA: WestEd.	No control group
Reciprocal Teaching	Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading classes. *American Educational Research Journal*, *35*(2), 309–332.	Duration < 12 weeks
	Alfassi, M. (2004). Reading to learn: Effects of combined strategy instruction on high school students. *The Journal of Educational Research*, *97*(4), 171–184.	Duration < 12 weeks
	Brady, P.L. (1990). *Improving the reading comprehension of middle school students through reciprocal teaching and semantic mapping strategies*. Unpublished doctoral dissertation, University of Oregon, Eugene.	Duration < 12 weeks
	Levin, M.C. (1989). *An experimental investigation of reciprocal teaching and informed strategies for learning taught to learning-disabled intermediate school learners*. Unpublished doctoral dissertation, Columbia University, New York.	Duration < 12 weeks
	Lovett, M.W., Borden, S.L., Warren-Chaplin, P.M., Lacerenza, L., DeLuca, T., & Giovinazzo, R. (1996). Text comprehension training for disabled readers: An evaluation of reciprocal teaching and text analysis training programs. *Brain & Language*, *54*(3), 447–480.	Duration < 12 weeks
	Lysynchuk, L., Pressley, M., & Vye, N.J. (1989, March). *Reciprocal instruction improves standardized reading comprehension performance in poor grade-school comprehenders*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.	Duration < 12 weeks
	Lysynchuk, L.M., Pressley, M., & Vye, N.J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *The Elementary School Journal*, *90*(5), 469–484.	Duration < 12 weeks
	Serran, G. (2002). *Improving reading comprehension: A comparative study of metacognitive strategies*. Unpublished master's thesis, Kean University, Union, NJ.	Duration < 12 weeks
	Westera, J., & Moore, D.W. (1995). Reciprocal teaching of reading comprehension in a New Zealand high school. *Psychology in the Schools*, *32*(3), 225–232.	Duration < 12 weeks
Talent Development High School	McPartland, J., Balfanz, R., Jordan, W., & Legters, N. (1998). Improving climate and achievement in a troubled urban high school through the Talent Development model. *Journal of Education for Students Placed at Risk*, *3*(4), 337–361.	Inadequate control group
REWARDS	Gowan, D.W. (2006). *REWARDS: Structural analysis as an effective means for decoding multisyllabic words*. Unpublished master's thesis, California State University, Fullerton.	Inadequate control group; pretest differences > 0.5 *SD*
Spalding Method	Mohler, G.M. (2002). *The effect of direct instruction in phonemic awareness, multisensory phonics, and fluency on the basic reading skills of low-ability seventh grade students*. Unpublished doctoral dissertation, University of Nebraska–Lincoln.	No control group
Thinking Works	Lynch, M.T. (2001).*The effects of strategy instruction on reading comprehension in junior high students*. Unpublished doctoral dissertation, University of Toledo, Toledo, OH.	Inadequate control group
Wilson Reading System	Moats, L.C. (1998). Reading, spelling and writing disabilities in the middle grades. In B.Y.L. Wong (Ed.), *Learning about learning disabilities* (2nd ed., pp. 367–389). San Diego, CA: Academic Press.	No control group

(continued)

Program	Study	Reason for exclusion
	Instructional-process programs	
	Moccia, J.P. (2005). *The influence of multi-sensory, multi-component reading intervention strategies with middle school poor readers*. Unpublished doctoral dissertation, Seton Hall University, South Orange, NJ.	Inadequate control group
	Reuter, H.B. (2006). *Phonological awareness instruction for middle school students with disabilities: A scripted multisensory intervention*. Unpublished doctoral dissertation, University of Oregon, Eugene.	Inadequate control group; pretest differences > 0.5 *SD*
	Wilson, B.A., & O'Connor, J.R. (1995). Effectiveness of the Wilson Reading System used in public school training. In C. McIntyre & J.S. Pickering (Eds.), *Clinical studies of multisensory structured language education for students with dyslexia and related disorders* (pp. 247–253). Salem, OR: International Multisensory Structured Language Education Council.	No control group
Other instructional-process programs	Alfassi, M. (2000). Using Information and Communication Technology (ICT) to foster literacy and facilitate discourse within the classroom. *Education Media International*, *37*(3), 137–148.	No control group
	Anderson, V., & Roit, M. (1993). Planning and implementing collaborative strategy instruction for delayed readers in grades 6–10. *The Elementary School Journal*, *94*(2), 121–137.	Quantitative data not provided
	Bakken, J.P., Mastropieri, M.A., & Scruggs, T.E. (1997). Reading comprehension of expository science material and students with learning disabilities: A comparison of strategies. *The Journal of Special Education*, *31*(3), 300–324.	Duration < 12 weeks
	Bos, C.S., Anders, P.L., Filip, D., & Jaffe, L.E. (1989). The effects of an interactive instructional strategy for enhancing reading comprehension and content area learning for students with learning disabilities. *Journal of Learning Disabilities*, *22*(6), 384–390.	Duration < 12 weeks
	Campbell, B.W. (2002). *Genre studies: Temporary homogeneous grouping to improve reading or merely another form of tracking?* Unpublished doctoral dissertation, University of San Diego, San Diego, CA.	Pretest equivalence not established
	DiCecco, V.M., & Gleason, M.M. (2002). Using graphic organizers to attain relational knowledge from expository text. *Journal of Learning Disabilities*, *35*(4), 306–320.	Inadequate outcome measure
	Dickson, S.V., & Bursuck, W.D. (1999). Implementing a model for preventing reading failure: A report from the field. *Learning Disabilities Research & Practice*, *14*(4), 191–202	No control group
	Dole, J.A., Brown, K.J., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly*, *31*(1), 62–88.	Duration < 12 weeks
	Erickson, E.A. (2006). *A comparison of reading growth between Intensive Supplemental Reading and Second Chance Reading for the Des Moines Public Schools, 2002–2005*. Unpublished doctoral dissertation, Drake University, Des Moines, IA.	Inadequate control group; pretest differences > 0.5 *SD*
	Esser, M.M.S. (2001). *The effects of metacognitive strategy training and attribution retraining on reading comprehension in African-American students with learning disabilities*. Unpublished doctoral dissertation, University of Wisconsin–Milwaukee.	Duration < 12 weeks
	Jitendra, A.K., Hoppes, M.K., & Xin, Y.P. (2000). Enhancing main idea comprehension for students with learning problems: The role of a summarization strategy and self-monitoring instruction. *The Journal of Special Education*, *34*(3), 127–139.	Duration < 12 weeks
	Katims, D.S., & Harris, S. (1997). Improving the reading comprehension of middle school students in inclusive classrooms. *Journal of Adolescent & Adult Literacy*, *41*(2), 116–123.	Duration < 12 weeks
	Klingner, J.K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal*, *96*(3), 275–293.	Duration < 12 weeks

(continued)

Program	Study	Reason for exclusion
	Instructional-process programs	
	Ligas, M.R. (2002). Evaluation of Broward County Alliance of Quality Schools project. *Journal of Education for Students Placed at Risk*, *7*(2), 117–139.	No control group
	Malmgren, K.W., & Leaone, P.E. (2000). Effects of a short-term auxiliary reading program on the reading skills of incarcerated youth. *Education and Treatment of Children*, *23*(3), 239–247.	No control group
	Manset-Williamson, G., & Nelson, J.M. (2005). Balanced, strategic reading instruction for upper-elementary and middle school students with reading disabilities: A comparative study of two approaches. *Learning Disability Quarterly*, *28*(1), 59–74.	Duration < 12 weeks
	Mastropieri, M.A., Scruggs, T.E., Hamilton, S.L., Wolfe, S., Whedon, C., & Canevaro, A. (1996). Promoting thinking skills of students with learning disabilities: Effects on recall and comprehension of expository prose. *Exceptionality*, *6*(1), 1–11.	Duration < 12 weeks
	Mastropieri, M.A., Scruggs, T., Mohler, L., Beranek, M., Spencer, V., Boon, R.T., & Talbott, E. (2001). Can middle school students with serious reading difficulties help each other and learn anything? *Learning Disabilities Research & Practice*, *16*(1), 18–27.	Duration < 12 weeks
	Nesbitt, J.S., & Wang, N. (2007, April). *An evaluation of a high school reading remediation program*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.	Insufficient information on control group pretests
	Newborn, S.L. (1998). The effects of instructional settings on the efficacy of strategy instruction for students with learning disabilities. *Dissertation Abstracts International*, *59* (05), 1527A.	Duration < 12 weeks
	Penney, C.G. (2002). Teaching decoding skills to poor readers in high school. *Journal of Literacy Research*, *34*(1), 99–118.	Pretest differences > 0.5 *SD*
	Peverly, S.T. & Wood, R. (2001). The effects of adjunct questions and feedback on improving the reading comprehension skills of learning-disabled adolescents. *Contemporary Educational Psychology*, *26*(1), 25–43.	Duration < 12 weeks
	Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, B. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, *28*(2), 147–169.	Duration < 12 weeks; inadequate outcome measure
	Sandora, C.A. (1995). *A comparison of two discussion techniques: Great Books (post-reading) and Questioning the Author (on-line) on students' comprehension and interpretation of narrative texts*. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.	Inadequate outcome measure
	Sandora, C., Beck, I., & McKeown, M. (1999). A comparison of two discussion strategies on students' comprehension and interpretation of complex literature. *Journal of Reading Psychology*, *20*(3), 177–212.	Duration < 12 weeks
	Topping K.J., & Fisher, A.M. (2003). Computerized formative assessment of reading comprehension: Field trials in the UK. *Journal of Research in Reading*, *26*(3), 267–279.	No control group
	Ugel, N.S. (1999). The effects of a multicomponent reading intervention on the reading achievement of middle school students with learning disabilities. *Dissertation Abstracts International*, *61* (01), 118A.	Duration < 12 weeks
	Wilder, A.A., & Williams, J.P. (2001). Students with severe learning disabilities can learn higher order comprehension skills. *Journal of Educational Psychology*, *93*(2), 268–278.	Inadequate outcome measure
	Williams, J.P. (1998). Improving the comprehension of disabled readers. *Annals of Dyslexia*, *48*, 213–238.	Duration < 12 weeks; inadequate outcome measure

Note. TORC-3 = Test of Reading Comprehension, 3rd edition; TAAS = Texas Assessment of Academic Skills.