Effective Reading Programs for the Elementary Grades: A Best-Evidence Synthesis

Article *in* Review of Educational Research · December 2009

DOI: 10.3102/0034654309341374

5 authors, including:

Robert E. Slavin
Johns Hopkins University

Cynthia Lake
Johns Hopkins University

Alan Cheung
The Chinese University of Hong Kong

Susan R Davis
Monash University (Australia)

Review of Educational Research
December 2009, Vol. 79, No. 4, pp. 1391–1466
DOI: 10.3102/0034654309341374

Effective Reading Programs for the Elementary Grades: A Best-Evidence Synthesis

Robert E. Slavin
Johns Hopkins University
University of York

Cynthia Lake
Johns Hopkins University

Bette Chambers
University of York
Johns Hopkins University

Alan Cheung
Johns Hopkins University

Susan Davis
Success for All Foundation

This article systematically reviews research on the achievement outcomes of four types of approaches to improving the reading success of children in the elementary grades: reading curricula, instructional technology, instructional process programs, and combinations of curricula and instructional process. Study inclusion criteria included use of randomized or matched control groups, a study duration of at least 12 weeks, valid achievement measures independent of the experimental treatments, and a final assessment at the end of Grade 1 or later. A total of 63 beginning reading (starting in Grades K or 1) and 79 upper elementary (Grades 2 through 5) reading studies met these criteria. The review concludes that instructional process programs designed to change daily teaching practices have substantially greater research support than programs that focus on curriculum or technology alone.

Keywords: elementary schools, meta-analysis, reading

From the first day of kindergarten to the last day of elementary school, children substantially define themselves as readers, and this has enormous influence on their development as learners and as members of society. Those who succeed in becoming fluent, strategic, and joyful readers are not guaranteed success in school or in life, but they are well on their way. However, those who do not succeed in reading, or who become reluctant readers, face long odds in achieving success in school and life. Every educator, parent, and policy maker knows the critical importance of reading in the elementary grades. Furthermore, the gap in reading performance between different ethnic groups, and between middle-class and disadvantaged children, is perhaps the most important policy issue in education in the United States. Because of the obvious importance of success in reading, schools invest enormous sums in initial teaching of reading and in remedial services for struggling readers.

Given the great importance of success in reading for millions of children and for our society as a whole, one would imagine that there would be a great deal of research on how teachers can most effectively teach children to read. There is in fact a great deal of basic research on reading, and we know a lot about how children learn to read and what goes wrong when they fail to learn (see, for example, National Early Literacy Panel, 2008; National Reading Panel [NRP], 2000; C. Snow, Burns, & Griffin, 1998). Yet there is much less research evaluating the practical programs actually available to schools and teachers to ensure reading success, and the research that does exist has not been comprehensively reviewed.

It is useful, for example, to know that effective beginning reading programs emphasize phonemic awareness, phonics, fluency, vocabulary, and comprehension, as concluded by the NRP (2000). Reviews by Adams (1990) and by C. Snow et al. (1998), as well as the NRP, have supported the importance of teaching with a strong emphasis on phonics and phonemic awareness. Yet school leaders and teachers do not choose between "phonics" and "no phonics." Instead, they choose among particular textbooks, software, and professional development approaches. Any particular program may incorporate the five NRP elements to a greater or lesser degree, but each also incorporates other features (such as classroom organization, motivation, grouping, assessment, and professional development) that also determine the outcomes of the program.

The importance of focusing attention on all aspects of reading approaches, not just on phonics or other NRP elements, was illustrated by the experience of the federal Reading First program. Based in large part on the findings of the NRP (2000) and earlier research syntheses, the Reading First program favored phonics and phonemic awareness, and national studies of Reading First by Gamse, Tepper-Jacob, Horst, Boulay, and Unlu (2008) and Moss et al. (2008) found that teachers in Reading First schools were in fact doing more phonics teaching than were those in similar, non–Reading First schools. Yet outcomes were disappointing, with small effects seen on first-grade decoding measures and no impact on comprehension measures in Grades 1 through 3. Similarly, a large study of intensive professional development focusing on phonics found no effects on the reading skills of second graders (Garet et al., 2008). The findings of these large-scale experiments imply that whereas the importance of phonics and phonemic awareness in reading instruction are well established, the addition of phonics to traditional basal instruction is not sufficient to bring about widespread improvement in children's reading. Other factors, especially relating to teaching methods, are also consequential.

The What Works Clearinghouse (WWC, 2009), in its beginning reading topic report, reviewed research on reading programs evaluated in Grades K through 3. However, the WWC only reports program ratings and does not include discussion of the findings or draw generalizations about the effects of types of programs. Furthermore, WWC inclusion standards applied in its beginning reading topic report include very brief studies (as few as 5 hours of instruction), and studies that used measures of skills taught in experimental but not control groups. It does not weight by sample size, and many of its conclusions are based on atypical effect sizes from studies with sample sizes as small as 46 (see Slavin, 2008).

The present article reviews research on the achievement outcomes of practical initial (nonremedial) reading programs for all elementary children, Grades K through 5, applying consistent methodological standards to the research. It is intended to provide fair summaries of the achievement effects of the full range of reading approaches available to educators and policy makers and to summarize for researchers the current state of the art in this area. The scope of the review includes all types of programs that teachers, principals, or superintendents might consider to improve the success of their children in reading: curricula, instructional technology (IT), instructional process programs, and combinations of curricula and instructional process. The review uses a form of best-evidence synthesis (Slavin, 1986), adapted for use in reviewing "what works" literatures in which there are generally few studies evaluating each of many programs. It is part of a series, all of which used the same methods with minor adaptations. Separate syntheses review research on remedial, preventive, and special education programs in elementary reading (Slavin, Lake, Davis, & Madden, 2009), middle and high school reading programs (Slavin, Cheung, Groff, & Lake, 2008), and reading programs for English language learners (Cheung & Slavin, 2005).

Focus of the Current Review

The present review uses procedures similar to those used in the secondary reading review to examine research on initial (nonremedial) programs for elementary reading. The purpose of the review is to place all types of initial reading programs intended to enhance reading achievement on a common scale, to provide educators and policy makers with meaningful, unbiased information that they can use to select programs most likely to make a difference with their students. The review emphasizes practical programs that are or could be used at scale. It therefore emphasizes large studies done over significant time periods that used standard measures, to maximize the usefulness of the review to educators. The review also seeks to identify common characteristics of programs likely to make a difference in reading achievement. This synthesis was intended to include all kinds of approaches to reading instruction, and groups them in four categories: reading curricula, IT, instructional process programs, and combinations of reading curricula and instructional process. *Reading curricula* primarily encompass core reading textbooks and curricula, such as *Reading Street* and *Open Court Reading*. *IT* refers to programs that use technology to enhance reading achievement. This includes traditional supplementary computer-assisted instruction (CAI) programs, in which students are sent to computer labs for additional practice. Other IT programs include *Reading Reels*, which provides embedded multimedia in daily lessons; and *Writing to Read*, which combines technology and nontechnology small group activities. *Instructional process programs* rely primarily on professional development to give teachers effective strategies for teaching reading. These include programs focusing on cooperative learning, such as *Peer-Assisted Learning Strategies* (*PALS*) and *Cooperative Integrated Reading and Composition* (*CIRC*), and programs focusing on phonics and phonological awareness. *Curriculum and instructional process programs*, specifically *Success for All* and *Direct Instruction*, provide specific phonetic curricula as well as extensive professional development focused on instructional strategies. Comprehensive school reform (CSR) programs were included only if they included specific reading programs; for a broader review of outcomes of elementary CSR models, see Comprehensive School Reform Quality Center (CSRQ; 2006) and Borman, Hewes, Overman, and Brown (2003).

Methodological Issues Characteristic of Elementary Reading Research

Although this review of research on reading programs shares methodological issues common to all systematic reviews, there are also some key issues unique to this subject and grade level. The thorniest of these relates to measurement. In the early stages of reading, researchers often use measures such as phonemic awareness that are not "reading" in any sense, though they are precursors. However, measures of reading comprehension and reading vocabulary tend to have floor effects at the kindergarten and first-grade levels. The present review included measures such as letter–word identification and word attack but did not accept measures such as auditory phonemic awareness. Measures of oral vocabulary, spelling, and language arts were excluded at all grade levels.

Another problem of early reading measurement is that in kindergarten, it is possible for a study to find positive effects of programs that introduce skills not ordinarily taught in kindergarten on measures of those skills. For example, until the late 1990s, it was not common in U.S. kindergartens for children to be taught phonics or phonemic awareness. Programs that moved these then-first-grade skills into kindergarten might appear very effective in comparison to control classes receiving little or no instruction on those skills but would, in fact, simply be teaching skills the control children would probably have mastered somewhat later.

Because of the difficulty of defining and measuring early literacy skills, multiyear evaluations of programs that may begin in kindergarten, but follow children at least through the end of first or second grade, are of particular value. By the end of second grade, it is certain that control students as well as experimental students have been seriously taught to read, and it becomes possible to use measures of reading comprehension and reading vocabulary that more fully represent the goals of reading instruction, not just precursors. Multiyear studies solve the problem of early presentation of skills ordinarily taught later. If kindergartners are taught certain first-grade reading skills, end-of-first-grade or -second-grade measures should be able to determine if this early teaching was truly beneficial. Due to the unique nature of research on kindergarten-only programs, studies whose final posttesting took place before spring of first grade are reviewed in a separate section of this article.

Review Methods

As noted earlier, the review methods used here are adaptations of a technique called *best-evidence synthesis* (Slavin, 1986, 2008). Best-evidence syntheses seek to apply consistent, well-justified standards to identify unbiased, meaningful information from experimental studies; discussing each study in some detail; and pooling effect sizes across studies in substantively justified categories. The method is very similar to meta-analysis (Cooper, 1998; Lipsey & Wilson, 2001), adding an emphasis on narrative description of each study's contribution. It is similar to the methods used by the WWC (2009), with a few important exceptions noted in the following sections. See Slavin (2008) for an extended discussion and rationale for the procedures used in all of these reviews.

Literature Search Procedures

A broad literature search was carried out in an attempt to locate every study that could possibly meet the inclusion requirements. Electronic searches were made of educational databases (JSTOR, ERIC, EBSCO, Psych INFO, Dissertation Abstracts) using various combinations of keywords (for example, *elementary students*, *reading*, *achievement*) and the years 1970 to 2009. Results were then narrowed by subject area (for example, *reading intervention*, *educational software*, *academic achievement*, *instructional strategies*). In addition to looking for studies by key terms and subject area, we conducted searches by program name. Web-based repositories and education publishers' Web sites were also examined. We attempted to contact producers and developers of reading programs to check whether they knew of studies that we had missed. Citations were obtained from other reviews of reading programs including the WWC (2009) beginning reading topic report; Adams (1990); NRP (2000); C. Snow et al. (1998); Torgerson, Brooks, and Hall (2006); Rose (2006); and August and Shanahan (2006) or potentially related topics such as IT (E. Chambers, 2003; Kulik, 2003; Murphy et al., 2002). We also conducted searches of recent tables of contents of key journals. We searched the following tables of contents from 2000 to 2009: *American Educational Research Journal*, *Reading Research Quarterly*, *Journal of Educational Research*, *Journal of Educational Psychology*, *Reading and Writing Quarterly*, *British Educational Research Journal*, and *Learning and Instruction*. Citations of studies appearing in the studies found in the first wave were also followed up.

Effect Sizes

In general, effect sizes were computed as the difference between experimental and control individual student posttests after adjustment for pretests and other covariates, divided by the unadjusted posttest control group standard deviation. If the control group *SD* was not available, a pooled *SD* 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, as when the only standard deviation presented was already adjusted for covariates or when only gain score *SD*s were available. If pretest and posttest means and *SD*s were presented but adjusted means were not, effect sizes for pretests were subtracted from effect sizes for posttests. In multiyear studies, effect sizes may be reported for each year but only the final year of treatment is presented in the tables. However, if there are multiple cohorts (e.g., K-1, K-2, K-3), each with adequate pretests, all cohorts are included in the tables.

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 reason for using weighted means is to maximize the importance of large studies, as the previous reviews and many others have found that small studies tend to overstate effect sizes (see Rothstein, Sutton, & Borenstein, 2005; Slavin & Smith, in press).

Effect sizes were broken down for measures of decoding (e.g., word attack, letter–word identification, and fluency), vocabulary, and comprehension/total reading. In general, comprehension, which is the ultimate goal of reading instruction, is the most important outcome measure. Very few studies reported separate vocabulary scores, so the tables only show separate outcomes for decoding and comprehension (although vocabulary measures are included in totals).

Criteria for Inclusion

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

1. The studies evaluated initial (i.e., nonremedial) classroom programs for elementary reading. Studies of variables, such as use of ability grouping, block scheduling, or single-sex classrooms, were not reviewed. Studies of tutoring and remedial programs for struggling readers are reviewed in a separate article (Slavin, Lake, Davis, & Madden, 2009).
2. The studies involved interventions that began when children were in elementary school, Grades K through 5. As noted earlier, studies that began and ended in kindergarten are reviewed separately. Programs beginning in Grades K or 1 were categorized as beginning reading, whereas those beginning in Grades 2 through 5 were categorized as upper elementary.
3. The studies compared children taught 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 had to be available in English.
5. Random assignment or matching with appropriate adjustments for any pretest differences (e.g., analyses of covariance) had to be used. Studies without control groups, such as pre–post comparisons and comparisons to "expected" scores, were excluded.
6. Pretest data had to be provided, unless studies used random assignment of at least 30 units (individuals, classes, or schools) and there were no indications of initial inequality. Studies with pretest differences of more than 50% of a standard deviation were excluded because, even with analyses of covariance, large pretest differences cannot be adequately controlled for as underlying distributions may be fundamentally different (Shadish, Cook, & Campbell, 2002).
7. The dependent measures included quantitative measures of reading performance, such as standardized reading measures. Experimenter-made measures were accepted if they were comprehensive measures of reading, which would be fair to the control groups, but measures of reading objectives inherent to the experimental program (but unlikely to be emphasized in control groups) were excluded. Studies using measures inherent to treatments, usually made by the experimenter or program developer, have been found to be associated with much larger effect sizes than are measures that are independent of treatments (Slavin & Madden, in press), and for this reason, effect sizes from treatment-inherent measures were excluded. The exclusion of measures inherent to the experimental treatment is a key difference between the procedures used in the present review and those used by the WWC (2009). Measures of reading individually administered by the children's own teachers were also excluded, on the basis that such assessments are susceptible to bias. As noted above, measures of prereading skills such as phonological awareness, as well as related skills such as oral vocabulary, language arts, and spelling, were not included in this review.

8. A minimum study duration of 12 weeks was required. This requirement is intended to focus the review on practical programs intended for use for the whole year, rather than brief investigations. Study duration is measured from the beginning of the treatments to posttest, so, for example, an intensive 8-week intervention in the fall of first grade would be considered a year-long study if the posttest were given in May. The 12-week criterion has been consistently used in all of the systematic reviews done previously by the current authors. This is another difference between the current review and the WWC (2009) beginning reading topic report, which included very brief studies.
9. Studies had to have at least 15 students and 2 teachers in each treatment group.

Limitations

It is important to note several limitations of the current review. First, the review focuses on experimental studies using quantitative measures of reading. There is much to be learned from qualitative and correlational research that can add depth and insight to understanding the effects of reading programs, but this research is not reviewed here. Second, the review focuses on replicable programs used in realistic school settings expected to have an impact 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, but it does not attend to shorter, more theoretically driven studies that may also provide useful information, especially to researchers. Finally, the review focuses on traditional measures of reading performance, primarily individually administered or group-administered standardized tests. These are useful in assessing the practical outcomes of various programs and are fair to control as well as experimental teachers, who are equally likely to be trying to help their students do well on these assessments. The review does not report on experimenter-made measures of content taught in the experimental group but not in the control group, even though results on such measures may also be of importance to some researchers or educators.

Categories of Research Design

Four categories of research designs were identified. *Randomized experiments* (R) 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* (RQE) (Slavin, 2008). *Matched* (M) studies were ones in which experimental and control groups were matched on key variables at pretest, before posttests were known, while *matched post hoc* (MPH) studies were ones in which groups were matched retrospectively, after posttests were known. Studies using fully randomized designs (R) are preferable to randomized quasi-experiments, but all randomized experiments are less subject to bias than matched studies. Among matched designs, prospective designs (M) were preferred to post hoc matched designs (MPH). In the text and in tables, studies of each type of program are listed in this order (R, RQE, M, MPH). Within these categories, studies with larger sample sizes are listed first. Therefore, studies discussed earlier in each section should be given greater weight than those listed later, all other things being equal.

For Additional Information

The following sections present summaries of findings and tables showing characteristics and findings of individual studies. Descriptions of individual studies have been withheld to meet the page limits of this journal but can be seen in an online version at www.bestevidence.org. The Web site presents reviews separately for beginning and upper elementary reading. The Web versions also include appendices listing all relevant studies excluded from the review and the reasons for exclusion, as well as overall ratings of the strength of the evidence supporting use of individual programs.

Beginning Reading

From the first day of kindergarten to the last day of first grade, most children go through an extraordinary transformation as readers. If all goes well, children at the end of first grade know the sounds of all the letters and can form them into words, know the most common sight words, and can read and comprehend simple texts. The K-1 period is distinct from other stages of reading development because during this stage, children are learning all the basic skills of turning print into meaning. From second grade on, children build fluency, comprehension, and vocabulary for reading ever more complex text in many genres, but the K-1 period is qualitatively different in its focus on basic skills. The following sections summarize research on programs for beginning reading.

Research on Beginning Reading Curricula

The reading curricula category consists of textbooks for initial (nonremedial) reading instruction. Some professional development is typically provided with these textbooks, but far less than would be typical of instructional process approaches.

Table 1 summarizes descriptions and outcomes of all studies of curriculum programs for beginning reading.

Beginning reading curricula have been evaluated in seven studies, five of which used randomized quasi-experiments.

These studies evaluated three core basal reading programs, *Open Court Reading*, *Reading Street*, and *Scholastic Phonics Readers* with *Literacy Place*; plus three supplemental programs, the *Open Court Phonics Kit*, *Phonics in Context*, and *Elements of Reading: Phonics and Phonemic Awareness*. The sample size-weighted mean effect size (ES) across all seven was +0.12, with the four studies of core basal programs reporting a weighted mean ES of +0.11 and the three studies of supplementary programs with a weighted mean of +0.12. Effect sizes averaged +0.23 for decoding measures but only +0.09 for comprehension/total reading measures.

Research on Instructional Technology for Beginning Reading

The effectiveness of IT 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 IT in elementary or secondary reading, although E. Chambers (2003) came to a somewhat more positive conclusion.

TABLE 1
Beginning Reading Curricula

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Decoding	Compre-hension	Overall effect size
Core basal programs										
Open Court Reading										
Borman et al. (2007)	Randomized quasi-experiment (L)	1 year	16 classes (9 E, 7 C), 307 students (165 E, 139 C)	1	Schools in Idaho, Texas, Florida, and Indiana. 61% FL, 57% minority	Terra Nova		—	+0.06	+0.17
						Reading Comprehension	+0.06			
						Reading Vocabulary	+0.22			
						Reading Composite	+0.17			
Reading Street										
Wilkerson et al. (2007)	Randomized quasi-experiment (L)	1 year	18 teachers, 387 students (220 E, 167 C)	1	Schools in 4 sites around the U.S. 26% FL, 86% W, 8% H, 3% AA	Gates MacGinitie		—	+0.15	+0.15
Wilkerson et al. (2006)	Randomized quasi-experiment (L)	1 year	16 teachers (8 E, 8 C)	1	5 schools in 2 urban, 1 rural site. 54% FL, 57% W, 25% AA, 11% H	Gates MacGinitie		—	–0.02	–0.02

(continued)

TABLE 1 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect eize
Scholastic Phonics Readers and Literacy Place										
Schultz (1996)	Randomized quasi-experiment (L)	1 year	4 districts, 8 classes, 301 students (162 E, 139 C)	1	Large urban school districts in CA	CTBS Reading Vocabulary Comprehension Word Analysis	 +0.07 +0.11 +0.21 +0.23	+0.23	+0.14	+0.16
Supplemental curricula										
Open Court Phonics Kit										
Barrett (1995)	Matched (S)	1 year	9 classes (5 E, 4 C), 161 students (78 E, 83 C)	1	Middle class district in Riverside, CA	TERA-2 SAT Total	+0.36 +0.62	+0.54	+0.47	+0.49
Phonics in Context										
Barrett (1995)	Matched (S)	1 year	11 classes (7 E, 4 C), 170 students (87 E, 83 C)	1	Middle class district in Riverside, CA	TERA-2 SAT Total	+0.21 +0.47	+0.43	+0.40	+0.34
Elements of Reading: Phonics and Phonemic Awareness										
Apthorp (2005a)	Randomized quasi-experiment (L)	1 year	6 schools, 16 teachers (8 E, 8 C), 257 students (126 E, 131 C)	1	4 high-poverty, 2 middle class schools. Overall, 57% FL, 56% AA, 41% W, 5% H	ERDA Gates MacGinitie	–0.09 –0.29	–0.09	–0.29	–0.19

Note. L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; CTBS = Comprehensive Test of Basic Skills; SAT = Scholastic Achievement Test; TERA = Test of Early Reading Ability; ERDA = Early Reading Diagnostic Assessement, FL = Free/reduced-price lunch; W = White; AA = African American; H = Hispanic.

Thirteen studies of IT for beginning reading met the standards for the present review. These were divided into three categories. *Supplemental technology programs*, such as *Waterford*, *WICAT*, and *Phonics-Based Reading*, are programs that provide additional instruction at students' assessed levels of need to supplement traditional classroom instruction. *Mixed-method models*, represented by *Writing to Read*, are methods that use CAI along with noncomputer activities as students' core reading approach. *Embedded multimedia*, represented by *Reading Reels*, provides video content embedded in teachers' whole-class lessons.

Descriptions and outcomes of all studies of IT in beginning reading that met the inclusion criteria appear in Table 2.

The weighted mean ES for all technology approaches in beginning reading was only +0.09 across 13 studies. Large, randomized studies by Dynarski et al. (2007) and Campuzano, Dynarski, Agodini, and Rall (2009) found no impact of five current supplemental CAI models. These findings greatly affected the weighted mean of nine studies of supplementary CAI, estimated at +0.08. The weighted mean ES for decoding measures, also substantially affected by the Dynarski/Campuzano findings, was only +0.05, although comprehension/total reading effects (not measured in Dynarski/Campuzano) averaged +0.20. Large effect sizes were reported in small, matched studies of *Waterford* and *WICAT*. *Reading Reels*, which uses multimedia embedded in teachers' class lessons, had modest positive effects in two large randomized experiments (weighted mean ES = +0.20). With these potentially promising exceptions, research on the use of technology in beginning reading instruction does not show positive achievement effects of the types of software that have been most commonly used.

Research on Instructional Process Programs for Beginning Reading

Instructional process programs are methods that focus on providing teachers with extensive professional development to implement specific instructional methods. These fell into three categories. *Cooperative learning* programs (Slavin, 1995, 2009) use methods in which students work in small groups to help one another master academic content. *Phonological awareness training* is an approach that gives teachers specific classroom strategies for building phonics and phonemic awareness skills. *Phonics-focused professional development models*, including *Reading and Integrated Literacy Strategies* (*RAILS*); *Sing, Spell, Read, and Write*; *Ladders to Literacy*; *Early Reading Research*; and *Orton Gillingham*, provide training to teachers to help them effectively incorporate phonics, phonemic awareness, and other elements in beginning reading lessons. Note that two comprehensive programs combining instructional process approaches with innovative curricula, *Success for All* and *Direct Instruction*, are reviewed in a separate section of this article.

Descriptions and outcomes of all studies of instructional process programs meeting the inclusion criteria appear in Table 3.

Effects for instructional process programs were very positive. Across 17 studies, 5 of which were randomized quasi-experiments, the weighted mean ES for instructional process approaches in beginning reading was +0.37. The mean was +0.47 for decoding measures and +0.30 for comprehension/total reading measures. In particular, positive effects were seen on cooperative learning programs such as *PALS* and *Classwide Peer Tutoring* (mean ES = +0.46), phonics-focused

TABLE 2
Instructional Technology in Beginning Reading

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Computer-assisted instruction										
Destination Reading										
Campuzano et al. (2009)	Randomized (L)	1 year	21 teachers (21 E, 14 C), 742 students (448 E, 294 C)	1	Schools across the U.S. 71% FL, 31% AA, 34% H, 34% W	SAT-10		—	+0.11	+0.11
Headsprout										
Campuzano et al. (2009)	Randomized (L)	1 year	63 teachers (32 E, 31 C), 1,079 students (574 E, 505 C)	1	Schools across the U.S. 35% FL, 81% W, 13% AA, 67% H	SAT-10		—	+0.01	+0.01
Plato Focus										
Campuzano et al. (2009)	Randomized (L)	1 year	29 teachers (15 E, 14 C), 618 students (327 E, 291 C)	1	Schools across the U.S. 48% FL, 67% W, 27% H, 5% AA	SAT-10		—	+0.03	+0.03

(continued)

TABLE 2 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Waterford Early Reading Program										
Campuzano et al. (2009)	Randomized (L)	1 year	46 teachers (28 E, 18 C), 1,155 students (689 E, 466 C)	1	Schools across the U.S. 47% FL, 37% AA, 16% H	SAT-10		—	+0.02	+0.02
Cassady & Smith (2005)	Matched (S)	1 year	6 classes (3 E, 3 C), 93 students (46 E, 47 C)	1	School in rural Midwest	Terra Nova Reading		—	+0.71	+0.71
Phonics Based Reading										
Macaruso et al. (2006)	Matched (S)	7 mo.	5 schools, 10 classes (5 E, 5 C), 179 students (92 E, 87 C)	1	Boston area, 50% FL	Gates MacGinitie		—	+0.20	+0.20
The Literacy Center (LeapFrog)										
RMC (2004)	Randomized quasi-experiment (S)	1 year	6 schools, 195 students (109 E, 86 C)	1	High-poverty schools in Las Vegas, 30% ELL	Gates MacGinitie DIBELS	–0.04 –0.01	–0.01	–0.04	–0.02

(continued)

TABLE 2 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre- hension	Overall effect size
WICAT										
Erdner et al. (1997)	Matched (S)	1 year	2 schools, 85 students	1	Schools in north central OK	CTBS		—	+1.05	+1.05
Reading Machine										
Abram (1984)	Randomized (S)	12 weeks	103 students	1	Not stated	ITBS		—	+0.19	+0.19
Mixed-method models										
Writing to Read										
Collis et al. (1990)	Matched (S)	1 year	97 students (53 E, 44 C)	1	Schools in British Columbia, Canada	SAT		—	+0.47	+0.27
						Total Reading	+0.47			
						Word Study	+0.07			
Beasley (1989)	Matched (S)	6 months	74 students (42 E, 32 C)	1	Middle-class students in Athens, AL; 82% W, 18% AA	SESAT-2		–0.13	–0.52	–0.27
						Sounds & Letters	–0.09			
						Word Reading	+0.15			
						Sentence Reading	–0.44			
						Reading Comprehension	–0.52			
						Total Reading	–0.44			

(continued)

TABLE 2 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Embedded multimedia										
Reading Reels										
B. Chambers et al. (2006)	Randomized (L)	1 year	10 schools, 394 students	1	High-poverty schools in Hartford, CT; 61% H, 35% AA	Woodcock		+0.20	+0.08	+0.17
						Word ID	+0.15			
						Word Attack	+0.32			
						Passage Comp.	+0.08			
						DIBELS	+0.12			
B. Chambers et al. (2008)	Randomized (S)	1 year	2 schools, 159 students (75 E, 84 C)	1	Hispanic students in high-poverty schools in Los Angeles and Las Vegas	Woodcock		+0.30	+0.17	+0.27
						Letter-Word	+0.33			
						Word Attack	+0.28			
						GORT				
						Fluency	+0.28			
						Comprehension	+0.17			

Note. L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; SAT-9 = Stanford Achievement Test 9th Edition; TOWRE = Test of Word Reading Efficiency; CTBS = Comprehensive Test of Basic Skills; ITBS = Iowa Test of Basic Skills; SAT = Scholastic Achievement Test; SESAT = Stanford Early School Achievement Test; GORT = Gray Oral Reading Test; FL = Free/reduced-price lunch; W = White; AA = African American; H = Hispanic; ELL = English language learner.

professional development programs such as *Sing, Spell, Read, and Write*; *Early Reading Research*; and *RAILS* (mean ES = +0.43), and teaching of phonological awareness to kindergartners (mean ES = +0.22 on tests at the end of first or second grade).

Research on Combined Curriculum and Instructional Process Approaches for Beginning Reading

Evaluations of programs that provide complete curricula as well as extensive professional development in classroom instructional processes are summarized in Table 4. These consist of two programs, *Success for All* and *Direct Instruction*.

Success for All is a comprehensive school reform program designed to ensure success in reading for children in high-poverty schools (Slavin et al., 2009). It provides schools with a K-5 reading curriculum that focuses on phonemic awareness, phonics, comprehension, and vocabulary development, beginning with phonetically controlled minibooks in Grades K and 1. Cooperative learning is extensively used at all grade levels. Struggling students, especially first graders, receive one-to-one tutoring. Extensive professional development and a full-time facilitator help teachers effectively apply all program elements. Across 23 studies involving more than 12,000 children, the weighted mean ES for *Success for All* was +0.29. On decoding measures the overall mean was +0.33, and the mean was +0.27 for comprehension/total reading.

Dating back to the 1960s, *Direct Instruction* is an approach to beginning reading instruction that emphasizes a step-by-step approach to phonics, decodable texts that make use of a unique initial teaching alphabet; and structured, scripted manuals for teachers. Across three evaluations of *Direct Instruction*, the weighted mean ES for beginning reading was +0.10. However, it is important to note that in other reviews that examined effects of *Direct Instruction* in all elementary grades (not just K and 1), this program has been rated as among the strongest in reading outcomes (e.g., Borman et al., 2003; CSRQ, 2006; Herman, 1999).

Kindergarten-Only Studies

As noted earlier, studies that take place only during kindergarten can pose serious methodological challenges. Because the goals of kindergarten instruction vary a great deal from place to place, and have changed dramatically over the past 30 years, it is always possible that any experimental-control difference on an end-of-kindergarten reading measure is simply due to the fact that the control group was not being taught to read. Even when reading is being taught, kindergarten classes can vary greatly in their emphasis on phonics, so measures of word attack and phonological awareness can be easily inflated by programs that focus on these skills earlier than the control treatment does. Still, it is useful to know about kindergarten-only studies, as they can provide initial indications of programs worth following through to first grade and beyond.

Thirteen studies met the standards of the review but took place only during the kindergarten year. These are summarized in Table 5.

The kindergarten-only studies generally support the conclusions of the studies that follow children through first grade and beyond. It is important to note that many of the programs cited in the main review, which tested children at the end of first grade, also reported very positive outcomes during kindergarten. These are also programs

TABLE 3
Instructional Process Programs in Beginning Reading

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Decoding	Compre-hension	Overall effect size
Cooperative learning programs										
Classwide Peer Tutoring (CWPT)										
Greenwood et al. (1989)	Randomized quasi-experiment (S)	4 years	6 schools (3 E, 3 C), 123 students	1–4 (same students)	High-poverty schools in Kansas City, KS	MAT		—	+0.57	+0.57
						Grade 4	+0.57			
						Grade 6 (2 year followup)	+0.55			
PALS										
Mathes & Babyak (2001)	Randomized quasi-experiment (S)	14 weeks	20 classes (10 E, 10 C), 110 students (61 E, 49 C)	1	Schools in Florida; 63% W, 36% AA	Woodcock		+0.72	+0.41	+0.61
						Word Identification	+0.51			
						Word Attack	+0.92			
						Passage Comprehension	+0.41			
Calhoon et al. (2006)	Randomized quasi-experiment (S)	20 weeks	3 schools, 6 classrooms, 78 students (41 E, 37 C)	1	Students taught in English in a majority-Hispanic school in NM; 75% FL, 32% W, 68% H	DIBELS		+0.29	—	+0.29
						Nonsense Word Fluency	+0.58			
						Oral Reading Fluency	+0.00			

(continued)

TABLE 3 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre- hension	Overall effect size
Calhoon et al. (2007)	Randomized Quasi-Experiment (S)	16 weeks	3 schools, 6 classrooms, 76 students (43 E, 33 C)	1	Students in border schools in 2-way bilingual program; 88% FL, 79% H, 21% W, 28% ELL	DIBELS		+0.33	—	+0.33
						Nonsense Word Fluency	+0.51			
						Letter Naming Fluency	+0.20			
						Oral Reading Fluency	+0.29			
Mathes et al. (2001)	Matched (S)	16 weeks	24 classes (12 E, 12 C), 140 students (84 E, 56 C)	1	Schools in the Southeast; 65% W, 32% AA	Woodcock		+0.49	+0.56	+0.50
						Word Identification	+0.39			
						Word Attack	+0.59			
						Passage Comprehension	+0.56			
						TERA-2	+0.48			
Mathes et al. (1998)	Matched (S)	16 weeks	20 classes (10 E, 10 C), 96 students (48 E, 48 C)	1	Schools in southeastern city	Woodcock		+0.38	+0.37	+0.37
						Word Identification	+0.21			
						Word Attack	+0.54			
						Passage Comprehension	+0.37			

(continued)

TABLE 3 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre- hension	Overall effect size
Phonological awareness training programs										
Lie (1991)	Randomized quasi-experiment (S)	2 years	10 schools, 208 students (Sequential analysis: 52 students; Positional analysis: 60 students; Control: 96 students)	1–2	Schools in Halden, Norway	Norwegian Reading Test				
						End of Grade 1	+0.34	—	+0.30	+0.30
						End of Grade 2	+0.30			
Lundberg et al. (1988)	Matched (L)	3 years	390 students (235 E, 155 C)	K-2	Schools in rural Denmark	Norwegian Reading Test				
						End of Grade 1	+0.40	—	+D540.48	+0.48
						End of Grade 2	+0.48			
Schneider et al. (1997) (Study 1)	Matched (L)	3 years	23 classes(11 E, 12 C), 371 students (205 E, 166 C)	K-2	Schools in rural Germany	German Reading Test			–0.19	–0.19
						End of Grade 1	+0.29			
						End of Grade 2	–0.19			
Schneider et al. (1997) (Study 2)	Matched (L)	3 years	18 classes (11 E, 7 C), 346 students (191 E, 155 C)	K-2	Schools in rural Germany	German Reading Test		—	+0.33	+0.33
						End of Grade 1	+0.53			
						End of Grade 2	+0.33			

(continued)

TABLE 3 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Blachman et al. (1999)	Matched (S)	1 1/2 years, 11 weeks in K-1, 1 year in 1st grade	4 schools (2 E, 2 C), 128 students (66 E, 62 C); 1-year follow-up: 106 students (58 E, 48 C)	K-1	High-poverty schools in Syracuse, NY	Woodcock Word ID	+0.28	+0.33	—	+0.33
						Decoding of Real Words	+0.64			
						Decoding of Non-Words	+0.74			
						1 year follow-up				
						Woodcock Word ID	+0.31			
						Decoding of Real Words	+0.34			
						Decoding of Non-Words	+0.36			
Phonics-focused professional development models										
Sing, Spell, Read, Write										
L. Jones (1995)	Matched (S)	7 months	4 classes, 97 students (50 E, 47 C)	1	School in Appalachian Mississippi; 55% FL, 78% W, 22% AA	Gates MacGinitie Reading Comprehension		—	+0.21	+0.21

(continued)

TABLE 3 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Decoding	Compre-hension	Overall effect size
Early Reading Research (ERR)										
Shapiro & Solity (2008)	Matched (S)	2 years	12 schools (6 E, 6 C), 434 students (235 E, 199 C)	K-1	Schools in England	British Achievement Scales				+0.54
						Word Reading	+0.62			
						NFER				
						Word Reading	+0.52			
						Accuracy	+0.59			
						Comprehension	+0.41			
Reading and Integrated Literacy Strategies (RAILS)										
Stevens et al. (2008)	Matched (S)	2 years	3 schools (2 E, 1 C), 237 students (112 E, 125 C)	K-1, 1–2	Schools in small city in PA; 71% FL, 94% W	MAT		—	+0.41	+0.41
						K-1	+0.39			
						1–2	+0.43			
Ladders to Literacy										
O'Connor (1999)	Matched (S)	1 year	4 schools (2 E, 2 C), 105 students (64 E, 41 C)	K-1	Large urban district, 46% AA, 51% W	Woodcock Letter-Word ID	+0.92	+0.20	—	+0.20
						Woodcock Letter-Word ID (1-year follow-up)	+0.02			
						Woodcock Word Attack (1-year follow-up)	+0.38			

(continued)

TABLE 3 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Decoding	Compreh-ension	Overall effect size
Orton-Gillingham										
Joshi et al. (2002)	Matched (S)	1 year	4 schools, 56 students (24 E, 32 C)	1	High-poverty schools in the Southwest; 81% FL, 53% minority	Woodcock Word Attack	+0.28	+0.28	+0.58	+0.43
						GMRT	+0.58			
Other professional development models										
Four Blocks										
Scarcelli & Morgan (1999)	Matched (S)	1 year	55 students (25 E, 30 C) in 4 classes (2 C, 2 E)	1	Title I school in Virginia Beach, VA	GMRT		—	+0.56	+0.56

Note. L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; MAT = Metropolitan Achievement Test; TERA = Test of Early Reading Ability; TOWRE = Test of Word Reading Efficiency; DORT = Durrell Oral Reading Test; GMRT = Gates-MacGinitie Reading Test; FL = Free/reduced-price lunch; W = White; AA = African American; H = Hispanic.

with a strong emphasis on phonics and/or cooperative learning, including *Success for All* (e.g., Jones, Gottfredson, & Gottfredson, 1997), and the phonological awareness training programs (e.g., Lundberg, Frost, & Petersen, 1988).

Overall Patterns of Outcomes: Beginning Reading

Across all categories, there were 63 qualifying studies of beginning reading programs that posttested children at the end of first grade or later. Nineteen of the studies used random assignment (8 were fully randomized and 11 were randomized quasi-experiments). The sample size-weighted mean ES was +0.22. These studies, involving more than 22,000 children, were identified from among more than 2,000 studies initially reviewed, and represent those that used rigorous experimental procedures.

Overall effects were somewhat stronger for decoding measures (such as *Woodcock Word Attack* and *Letter-Word Identification*) than for measures of comprehension and total reading. Across all studies, the weighted mean effect size was +0.27 for decoding measures and +0.20 for comprehension/total reading. Comprehension measures were more likely to show positive effects in multiyear studies that followed children into second grade or beyond.

There are several important patterns in the findings on beginning reading programs that are worthy of note. First, this article finds that successful programs almost always provide teachers with extensive professional development and follow-up focused on specific teaching methods. In particular, most of the beginning reading programs with strong evidence of effectiveness have cooperative learning at their core: *Success for All*, *PALS*, *Reading Reels*, and *Classwide Peer Tutoring* all emphasize children working with other children on structured activities. These are all forms 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, the elements that previous reviewers (e.g., Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003; Slavin, 1995, 2009; Webb, 2008) have identified as essential to the effectiveness of cooperative learning.

Second, all of the beginning reading programs found to be effective or promising in qualifying experiments have a strong focus on teaching phonics and phonemic awareness. This is particularly true of *Success for All*, *PALS*, *Reading Reels*, phonological awareness training, *Open Court Phonics Kits*, *Scholastic Phonics Readers* with *Literacy Place*, *Early Reading Research*, *RAILS*, *Direct Instruction*, and *Phonics-Based Reading*. It is important to note that studies of all of these programs found positive effects on comprehension and/or total reading measures, not just decoding measures that would appear more slanted toward phonetic approaches. However, an emphasis on phonics did not guarantee positive effects. Phonetic curricular approaches and supplemental CAI models, in particular, had minimal impacts on student outcomes. A large-scale evaluation of phonics-focused professional development by Garet et al. (2008) similarly found minimal effects for second graders. It clearly matters a great deal how reading is taught, and an emphasis on phonics may be necessary but it is not sufficient to ensure meaningful reading gains.

One key implication of the Gamse et al. (2008) evaluation of Reading First is that it is not enough to encourage teachers to emphasize phonics, phonemic awareness, and other elements. The Moss et al. (2008) report that analyzed differences

TABLE 4
Curriculum Plus Instructional Process Programs in Beginning Reading

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Success for All										
Borman et al. (2007)	Randomized (L)	3 years	35 schools, 2,108 students (1,085 E, 1, 023 C)	K-2	Title I schools throughout the U.S.; 72% FL, 57% AA, 31% W, 10% H	Woodcock		+0.28	+0.21	+0.25
						Word Identification	+0.22			
						Word Attack	+0.33			
						Passage Comprehension	+0.21			
Correnti (2009)	Matched (L)	4 years	115 schools (30 E, 85 C), 3,783 students (831 E, 2,932 C)	K-3	High-poverty schools in 17 states; 69% FL, 52% AA, 22% W, 19% H, 6% Asian	Terra Nova				+0.43
Madden et al. (1993); Slavin et al. (1993)	Matched (L)	5 years	10 schools (5 E, 5 C), 1,925 students (890 E, 1,035 C), 5 cohorts (1st grade in experiment 1 year, 2nd grade 2 years, etc.)	1–5	African American students in high-poverty schools in Baltimore, MD	Average of Woodcock, DORT, and CTBS		+0.55	+0.39	+0.46
						1st grade	+0.55			
						2nd grade	+0.32			
						3rd grade	+0.49			
						CTBS				
						4th grade	+0.45			
						5th grade	+0.48			

(continued)

TABLE 4 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Nunnery et al. (1996)	Matched (L)	2 years	64 schools (46 E, 18 C), 1,555 students	1–2	High-poverty schools in Houston, TX; 79% FL, 52% H, 48% AA	Average of Woodcock and DORT		+0.09	+0.02	+0.05
						First cohort (Gr. 2)	–0.08			
						Second cohort (Gr. 1)	+0.09			
						Spanish (Gr. 1)	+0.21			
Livingston & Flaherty (1997)	Matched (L)	2 years	6 schools (3 E, 3 C), 3 cohorts: English speakers (272 E, 184 C), Spanish bilingual (87 E, 93 C), Other ESL (80 E, 112 C)	1, 2	High-poverty multilingual schools in Modesto and Riverside, CA	Average of Woodcock and DORT across cohorts		+0.49	+0.49	+0.49
						English-Dominant	+0.28			
						Spanish Bilingual	+0.77			
						ESL	+0.43			
Ross et al. (1996)	Matched (L)	1 year	4 schools (2 E, 2 C), 540 students (169 E, 371 C)	1	Mostly Hispanic schools in Amphitheater District near Tucson, AZ	Average of Woodcock and DORT		+0.62	+0.33	+0.47

(continued)

TABLE 4 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Jones et al. (1997)	Matched (L)	3 years	2 schools (1 E, 1 C), 498 students (339 E, 159 C); Cohort 1:172 students (113 E, 59 C); Cohort 2: 157 students (109 E, 48 C); Cohort 3:169 students (117 E, 52 C)	3 Cohorts: Cohort 1: K-3; Cohort 2: K-2; Cohort 3: K-1	High-poverty AA schools in Charleston, SC	Woodcock		+0.23	+0.02	+0.27
						Kindergarten Woodcock and DORT	+0.98			
						1st grade SAT or BSAP	+0.20			
						1st grade SAT	–0.03			
						2nd grade SAT	+0.10			
						3rd grade	–0.06			
B. Chambers et al. (2005)	Matched (L)	1 year	8 schools (4 E, 4 C), 455 students (311 E, 144 C)	K-1	Mostly Hispanic communities in the U.S.	Woodcock Reading Mastery Test		+0.20	+0.21	+0.20
Ross, Smith, & Casey (1992)	Matched (L)	3 years	2 schools (1 E, 1 C), 370 students (223 E, 147 C), 3 cohorts	1–3	Rural schools in Caldwell, ID	Average of Woodcock and DORT		–0.10	–0.11	–0.10
Ross & Casey (1998b)	Matched (L)	2 years	8 schools (3 E, 5 C), 356 students (151 E, 205 C)	K-1	High-poverty schools in Ft. Wayne, IN; 75% FL, 45% minority	Woodcock		+0.33	+0.17	+0.25
						Word Identification	+0.22			
						Word Attack	+0.45			
						Passage Comprehension	+0.14			
						Durrell Oral	+0.21			

(continued)

TABLE 4 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre- hension	Overall effect size
Muñoz & Dossett (2004)	Matched (L)	3 years	6 schools (3 E, 3 C), 349 students (217 E, 132 C)	K-3	High-poverty schools in Louisville, KY	CTBS		—	+0.15	+0.15
Dianda & Flaherty (1995)	Matched (L)	2 years	6 schools (3 E, 3 C), 319 students (131 E, 188 C)	1	Mostly Hispanic students in schools in California; 72% FL, 42% H, 34% W, 32% ELL	Woodcock		+0.41	+0.45	+0.42
						Letter-Word Identification	+0.46			
						Word Attack	+0.36			
						Passage Comprehension	+0.45			
						Woodcock (all three measures)				
						English speakers	+0.55			
						Spanish bilingual	+0.84			
						Spanish dominant	+0.82			
						Non-English speakers	+0.11			
Ross & Casey (1998a)	Matched (L)	1 year	4 schools (2 E, 2 C), 316 students (156 E, 160 C)	1	Suburban schools in Portland, OR	Average of Woodcock and DORT		0.00	–0.02	–0.01

(continued)

TABLE 4 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Ross et al. (1997a)	Matched (L)	2 years	Cohort 1: 135 students (94 E, 41 C); Cohort 2: 146 students (106 E, 40 C)	K-1, 1–2	High-poverty schools in Clarke Co., GA	Average of Woodcock and DORT		+0.22	+0.08	+0.15
						1st grade	+0.27			
						2nd grade	+0.03			
Ross et al. (1995)	Matched (L)	3 years	2 schools, 3 cohorts, 251 students; Cohort 1: 59 E, 47 C; Cohort 2: 54 E, 20 C; Cohort 3: 45 E, 32 C	K-4	Title I schools in Ft. Wayne, IN	Average of Woodcock and DORT		+0.09	–0.09	0.00
						2nd grade	+0.10			
						3rd grade	–0.10			
						4th grade	0.00			
Casey et al. (1994)	Matched (S)	1 year	3 schools (2 E, 1 C), 189 students (116 E, 73 C)	1	High-poverty African American schools in Memphis, TN	Woodcock		+0.78	+0.53	+0.65
						Word Identification	+0.52			
						Word Attack	+1.03			
						Passage Comprehension	+0.63			
Ross, Smith, & Bond (1994)	Matched (S)	2 years	Cohort 1: 4 schools, 133 students(65 E, 68 C); Cohort 2: 2 schools, 46 students (20 E, 26 C)	K-1, 1–2	African American students in high-poverty schools in Montgomery, AL	Average of Woodcock and DORT		+0.76	+0.47	+0.62
						K-1 Cohort	+0.39			
						1–2 Cohort	+1.15			

(continued)

TABLE 4 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Smith et al. (1994)	Matched (S)	4 years	2 schools, 142 students (74 E, 68 C), 4 cohorts	1–4	High poverty AA school in Memphis	Average of Woodcock and DORT/Gray		+0.55	+0.65	+0.60
						1st grade	+1.15			
						2nd grade	+0.08			
						3rd grade	+0.56			
						4th grade	+0.04			
Wasik & Slavin (1993)	Matched (S)	3 years	2 schools (1 E, 1 C), 3 cohorts	1–3	High-poverty schools in Charleston, SC; 40% FL; 60% AA	Average of Woodcock and DORT		+0.39	+0.39	+0.39
						1st grade	+0.20			
						2nd grade	+0.67			
						3rd grade	+0.30			
Slavin & Madden (1991)	Matched (S)	2 years	2 schools (1 E, 1 C), 108 students (58 E, 50 C)	1–2	Small rural town in Maryland; 40% FL, 50% AA, 50% W	Average of Woodcock and DORT	+0.02	+0.02	+0.02	+0.02
						CTBS	+0.02			
Wang & Ross (1999a)	Matched (S)	1 year	4 schools (2 E, 2 C), 97 students (50 E, 47 C)	1	High-poverty schools in Little Rock, AK	Average of Woodcock and DORT		+0.20	+0.39	+0.30

(continued)

TABLE 4 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre- hension	Overall effect size
Wang & Ross (1999b)	Matched (S)	1 year	2 schools (1 E, 1 C), 82 students (43 E, 39 C)	1	High-poverty mostly Hispanic schools in Alhambra District near Phoenix, AZ	Average of Woodcock and DORT		+0.15	+0.16	+0.15
Slavin & Madden (1998)	Matched (S)	3 years	50 students (21 E, 29 C)	1–3	Spanish-dominant LEP students in Philadelphia, PA, who had transitioned to English classes	Woodcock		+0.36	–0.07	+0.22
						Word Attack	+0.65			
						Word Identification	+0.06			
						Passage Comprehension	–0.07			
Direct Instruction										
Kennedy (1978)	Matched (L)	4 years	2,216 children (1,161 E, 1,055 C)	K-3	High poverty schools in NY, RI, IL, & MS	MAT Reading Comprehension		—	+0.07	+0.07

(continued)

TABLE 4 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Decoding	Compre-hension	Overall effect size
Mac Iver et al. (2003)	Matched (L)	4 years	12 schools (6 E, 6 C), 275 students (171 E, 104 C)	K-3	High-poverty schools in Baltimore, majority African American	CTBS Reading Comprehension Vocabulary	 +0.13 0.00	—	+0.13	+0.07
Grant (1973)	Matched Post Hoc (S)	2 years	2 schools, 78 students (39 E, 39 C)	K-1	High-poverty African American students in WI	Wisconsin Reading Skill Development Long Vowels Base Words Dale Johnson Word Recognition	 +0.64 +1.33 +0.54	+0.84	—	+0.84

Note. L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; DORT = Durrell Oral Reading Test; CTBS = Comprehensive Test of Basic Skills; SAT = Scholastic Achievement Test; BSAP = Basic Skills Assessment Program; MAT = Metropolitan Achievement Test; FL = Free/reduced-price lunch; W = White; AA = African American; H = Hispanic; ELL = English language learner.

TABLE 5
Kindergarten-Only Studies

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Reading curricula								
Superkids								
Borman & Dowling (2007)	Matched (L)	1 year	43 classes (23 E, 20 C), 750 students (400 E, 350 C)	K	Schools throughout the U.S., 52% minority	SAT-10		+0.20
						Sounds and Letters	+0.25	
						Word Reading	+0.14	
						Sentence Reading	+0.22	
D'Agostino (2009)	Matched (L)	1 year	43 classes (21 E, 22 C), 750 students (302 E, 368 C)	K	Schools throughout the U.S., 47% FL, 38% minority	ITBS		+0.23
						Word Analysis	+0.41	
						Reading Words	+0.23	
						Reading Comprehension	+0.24	
						Vocabulary	+0.02	
Voyager Universal Literacy								
Frechtling et al. (2006)	Matched (L)	1 year	8 schools (4 E, 4 C), 398 students (202 E, 196 C)	K	African American students in 8 urban schools	Woodcock		+0.67
						Word ID	+0.21	
						Word Attack	+1.11	
Hecht (2003)	Matched (S)	5 months	3 schools (1 E, 2 C), 213 students (101 E, 112 C)	K	High-poverty schools in Orlando	Woodcock		–0.02
						Word ID	–0.10	
						Word Analysis	+0.10	
						DIBELS		
						Nonsense Word	–0.07	

(continued)

TABLE 5 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Instructional technology								
Waterford Early Reading Program								
Paterson et al. (2003)	Matched (L)	1 year	16 classes (8 E, 8 C)	K	High-poverty community in western New York	Clay Word Recognition Test		0.00
Tracey & Young (2006)	Matched (L)	1 year	15 classes (8 E, 7 C), 265 children (151 E, 114 C)	K	High-minority northeastern community	TERA-2		+0.47
The Literacy Center (LeapFrog)								
RMC (2003)	Randomized quasi-experiment (S)	1 year	6 schools, 258 students (126 E, 132 C)	K	High-poverty schools in Las Vegas, 30% ELL	Gates MacGinitie DIBELS	+0.17 +0.12	+0.14
Destination Reading								
Barnett (2006)	Matched (L)	1 year	15 classes (8 E, 7 C)	K	High-poverty high-minority community in FL	DIBELS Clay Word Recognition Test Dolch	–0.56 –0.47 –0.56	–0.53
Writing to Read								
Stevenson et al. (1988)	Matched (S)	1 year	241 students (86 E, 155 C)	K	African American students in Washington, DC	MAT Reading		+0.35

(continued)

TABLE 5 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Granick & Reid (1987)	Matched (S)	1 year	2 schools, 73 students (37 E, 36 C)	K	High-poverty African American schools in Baltimore	MAT		+0.02
Instructional process programs								
Ladders to Literacy								
Fuchs et al. (2001)	Randomized (L)	20 weeks, with a one-year follow-up	8 schools (4 E, 4 C), 404 students, 3 groups: Ladders only: 11 teachers, 136 students; Ladders + PALS: 11 teachers, 133 students; Control: 11 teachers, 135 students	K	Title I and non–Title I kindergartens in Nashville, TN	*Ladders to Literacy Group*		+0.21
						End of kindergarten		
						Woodcock		
						Word Attack	+0.17	
						Word ID	–0.25	
						Followup to Fall of first grade		
						Word Attack	+0.38	
						Word ID	+0.05	
						Ladders + PALS Group		
						End of kindergarten		
						Word Attack	+0.36	
						Word ID	+0.25	
						Followup to Fall of first grade		
						Word Attack	+0.41	
						Word ID	+0.43	

(continued)

TABLE 5 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
O'Connor (1999)	Matched (L)	1 year	17 classes (9 E, 8 C), 318 students (192 E, 89 C)	K	Rural midwestern district, 100% White	Woodcock Johnson Letter Word ID		+0.43
						Typical children	+0.33	
						At-risk children	+0.68	
Little Books								
Phillips et al. (1990)	Randomized Quasi-Experiment (L)	1 year	18 classes, 309 students	K	Urban and rural schools in Newfoundland, Canada	MET		+0.22
						School + home	+0.33	
						School only	+0.19	
						Home only	+0.14	

Note. L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; ITBS: Iowa Test of Basic Skills; SAT-10: Stanford Achievement Test; TERA = Test of Early Reading Ability; MAT = Metropolitan Achievement Test; FL = Free/reduced-price lunch; W = White; AA = African American; H = Hispanic; ELL = English language learner.

between Reading First and similar Title I schools that did not receive Reading First funding found that Reading First teachers were in fact spending more time teaching reading and, specifically, more time on phonics, phonemic awareness, fluency, vocabulary, and comprehension. The Reading First teachers were significantly more likely to use basal textbooks that were revisions of traditional basals designed primarily to increase the focus on phonics and phonemic awareness. In order of popularity in Reading First schools, these were *Harcourt Trophies* (22.5% of Reading First, 15.0% of non–Reading First), *Open Court Reading* (15.4% vs. 9.8%), *Scott Foresman Reading* (13.0% vs. 12.2%), and Houghton Mifflin's *Nation's Choice* (10.7% vs. 2.5%). Yet none of these had ever been evaluated at the beginning of Reading First, and only *Open Court Reading* has been adequately evaluated since then, in a study that found modest impacts (ES = +0.17; Borman et al., 2007). If adopting books with more phonics and spending a few more minutes each day on the five elements recommended by the NRP (2000) were sufficient to improve beginning reading performance, the Gamse et al. (2008) national evaluation would have found significant positive effects. The research summarized in the present review points in a different direction. It supports the use of well-developed programs that integrate curriculum, pedagogy, and extensive professional development.

Upper Elementary Reading Programs

From second to fifth grade, children go through a critical transformation as readers. Most beginning second graders are able to decode, to recognize key sight words, to comprehend simple texts, and to read with some degree of fluency. The tasks that lay ahead of them, however, are qualitatively different from those they have navigated so far. They must consolidate and extend their basic skills, to be sure; and they must become fluent, confident readers. But most important, children in the upper elementary grades must become strategic comprehenders of increasingly sophisticated text. They must build a vocabulary of words and concepts as well as a vocabulary of cognitive and metacognitive approaches to texts. Whereas decoding skills may develop in a fairly step-by-step progression, the skills mastered in the upper elementary grades emerge as children read in many genres and learn how to make sense of what they read, a less straightforward process. Early decoding success is a key predictor of success in the upper elementary grades and beyond (e.g., Juel, 1988), yet there are many children who are successful decoders but poor comprehenders. This period is also distinct from the middle grades, when reading instruction is not typically taught as a separate subject but is subsumed in English or language arts.

Because of the different objectives and requirements of the upper elementary grades, programs that are effective in building beginning reading skills are not necessarily optimal in the upper elementary grades, and vice versa. For this reason, in reviewing research on effective reading programs, it is important to review programs at each of these levels separately. This section focuses on studies of nonremedial classroom reading approaches that begin in Grades 2 through 5.

Current Issues in Upper Elementary Reading

In recent years, reading in the upper elementary grades has taken on particular centrality because of the growing importance of test-based accountability. In the United States, state accountability systems have long emphasized performance in Grades 3 through 5 as the indicator of elementary school success; and in 2001, No Child Left Behind heightened this emphasis, requiring testing of reading and math in every grade from 3 to 8 and adding sanctions for schools not making adequate yearly progress. In England, Key Stage 2 assessments in reading and math in Year 6 (age 11) are the main indicators of primary school success.

Despite the obvious importance of upper elementary reading for policy and practice, there has never been a review of research on effective programs at this grade level. The federal WWC (2009) has created a topic report on beginning reading programs, and this synthesis included studies with students up to third grade. However, the WWC excluded studies that included grades above 3 if they did not analyze data separately for grades above and below third grade, and this excluded many upper elementary studies that included Grades 2 through 4, 3 through 5, and so on. At this writing, the WWC has not announced a plan to do an upper elementary reading review. Deshler, Palincsar, Biancarosa, and Nair (2007) published a major "research-based guide to instructional programs and practices" for struggling adolescent readers. It contains brief discussions of the research evidence supporting each of 48 widely used programs, as well as lists of articles for each, and many of the articles reported studies of Grades 3 through 6. Yet Deshler et al. did not attempt to synthesize or compare the evidence bases for the programs at any grade level.

The review of research on upper elementary reading programs summarized in this section uses methods identical to those used in the beginning reading review, except that programs had to have begun in Grades 2 through 5. This synthesis groups upper elementary reading programs in three categories, defined previously for beginning reading programs: reading curricula, IT, and instructional process programs. *Reading curricula* primarily encompass core reading textbooks and curricula, such as Scott Foresman's *Reading Street*, as well as supplementary texts such as Scholastic's *Fluency Formula*. *IT* refers to programs that use technology to enhance reading achievement, especially CAI. *Instructional process programs* are the most diverse. All programs in this category rely primarily on professional development to give teachers effective strategies for teaching reading. These include programs focusing on cooperative learning, classroom motivation and management, and metacognitive strategies. Examples include *Cooperative Integrated Reading and Composition*, *PALS*, *Exemplary Center for Reading Instruction* (*ECRI*), and *Consistency Management-Cooperative Discipline* (*CMCD*).

Research on Upper Elementary Reading Curricula

The reading curricula category includes seven qualifying studies of core basal textbooks and eight studies of supplementary texts used as initial instruction with all students. Characteristics and findings of individual studies appear in Table 6.

Both core and supplemental reading curricula for the upper elementary grades have been studied in high-quality evaluations. Among 15 studies, there were 5 randomized experiments as well as 4 randomized quasi-experiments, involving more than 10,000 students. These studies found few effects on student reading achievement.

TABLE 6
Upper Elementary Reading Curricula

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size
Core basal programs								
Open Court								
Borman et al. (2007)	Randomized (L)	1 year	5 schools 33 teachers (18 E, 15 C), 613 students (342 E, 271 C)	2–5	High-poverty schools in ID, FL, NC, TX; 77% FL, 73% minority, 11% ESL	Terra Nova		+0.15
						Comprehension	+0.15	
						Composite	+0.15	
						Vocabulary	+0.13	
Skindrud & Gersten (2006)	Matched (L)	2 years	Grade 2 cohort: 434 students (292 E, 142 C); Grade 3 cohort: 642 students (350 E, 292 C)	2–3, 3–4	High-poverty schools in Sacramento	SAT-9		+0.20
						Grade 2–3 cohort	+0.30	
						Grade 3–4 cohort	+0.10	
Reading Street								
Wilkerson et al. (2006)	Randomized (L)	1 year	5 schools, 32 teachers	2–3	3 middle-class schools; 2 Title I, high-poverty schools; 54% FL, 57% W, 25% AA, 11% H	Gates MacGinitie		–0.06
						2nd grade	–0.10	
						3rd grade	–0.01	
Wilkerson et al. (2007)	Randomized (L)	1 year	40 teachers, 793 students (409 E, 384 C)	2–3	4 schools nationwide; 86% W, 8% H, 3% AA, 26% FL	Gates MacGinitie		–0.04
						2nd grade	–0.14	
						3rd grade	+0.06	

(continued)

TABLE 6 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Houghton Mifflin Reading								
Swartz & Johnson (2003)	Matched (L)	Cohort 1: 2 years; Cohort 2: 1 year	10 schools (5 E, 5 C); 2 Cohorts: Cohort 1: 586 students (220 E, 326 C); Cohort 2: 465 students (91 E, 374 C)	Cohort 1: Grades 2–3; Cohort 2: Grade 3	Mostly AA schools in Chicago; 94% FL, 76% AA, 16% W, 9% H	ITBS		+0.11
						Cohort 1		
						Reading	–0.08	
						Vocabulary	+0.38	
						Total	+0.15	
						Cohort 2		
						Reading	–0.04	
						Vocabulary	+0.17	
						Total	+0.07	
Harcourt Reading Program								
Conner, Greene, & Munroe (2004)	Matched (L)	1 year	63 schools (18 E, 45 C), 12,832 students (3,928 E, 8,904 C)	3–5	High-poverty schools in Philadelphia	Terra Nova		+0.10

(continued)

TABLE 6 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Whole language basals								
Rigby								
Wilkerson (2004)	Matched (L)	32 weeks	13 classes in 4 schools (2 E, 2 C), 472 students (245 E, 227 C)	2 and 4	High-poverty schools; 80% FL, 57% AA, 29% H, 5% W	Gates-MacGinitie		–0.26
						Second Graders		
						Word Decoding	+0.22	
						Word Knowledge	–0.07	
						Comprehension	–0.23	
						Total	–0.03	
						Fourth Graders		
						Vocabulary	–0.61	
						Comprehension	–0.33	
						Total	–0.48	
Supplementary curricula								
Schoolwide Enrichment Reading Model								
Reis et al. (2008)	Randomized (L)	14 weeks	31 teachers (17 E, 14 C), 544 students (306 E, 238 C)	3–5	2 middle-class schools in New England towns; 36% FL, 64% W, 28% H, 3% AA, 3% Asian, 18% LEP	Oral Reading Fluency	+0.08	+0.12
						ITBS	+0.15	

(continued)

TABLE 6 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Elements of Reading: Comprehension								
Resendez et al. (2006)	Randomized (L)		18 teachers (10 E, 8 C), 413 students (229 E, 184 C)	3	Schools in AZ, KY, VA, and OR; 69% FL, 36% W, 28% H, 20% AA, 6% Native American	Gates-MacGinitie		+0.09
						Vocabulary	+0.21	
						Comprehension	+0.11	
						Total	+0.17	
						ERDA		
						Target Words in Context	+0.05	
						Narrative Passage Fluency	+0.03	
						Informational Passage Fluency	0.00	
						Reading Comprehension	+0.12	
Elements of Reading: Vocabulary								
Apthorp (2005b)	Randomized quasi-experiment (L)	1 year	7 schools, 268 students (147 E, 121 C)	3	High-poverty schools in AL and NY. 83% FL, 49% AA, 46% W, 10% LEP	Gates-McGinitie		+0.10
						Vocabulary	+0.21	
						Comprehension	+0.10	
						ERDA Sight Vocabulary	0.00	

(continued)

TABLE 6 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Elements of Reading: Fluency								
Apthorp (2005c)	Randomized quasi-experiment (S)	1 year	10 classes, 184 students (97 E, 87 C)	2	Majority White high-poverty Title I schools; 74% FL, 82% W, 12% AA, 4% H, 8% LEP	ERDA		+0.10
						Word Identification	0.00	
						Narrative Passage Fluency	+0.15	
						Informational Passage Fluency	+0.18	
						Gates McGinitie Comprehension	+0.05	
Fluency Formula								
Sivin-Kachala & Bialo (2005)	Randomized quasi-experiment (S)	1 year	12 classes, 128 students (66 E, 62 C)	2	Suburban districts in Long Island, NY; 20% FL, 7% LEP	Woodcock Passage Comprehension		–0.24
Jacob's Ladder								
Stambaugh (2007)	Matched (S)	12 weeks	2 schools	3–5	Rural high-poverty schools in OH. 27% FL	ITBS		+0.02

(continued)

TABLE 6 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Contextually-Based Vocabulary Instruction								
Nelson & Stage (2007)	Randomized quasi-experiment (L)	3 months	16 classes (8 E, 8 C), 308 students (159 E, 149 C)	3, 5	Schools in midwestern district. 32% FL, 70% W, 24% H, 24% LEP	Gates-MacGinitie		+0.15
						Comprehension	+0.27	
						Vocabulary	+0.03	
QuickReads								
Huxley (2006)	Matched (S)	12 weeks	4 classes (2 E, 2 C), 61 students (35 E, 26 C)	3	High-poverty suburban school; 69% FL, 63% AA, 35% W	Gates-MacGinitie		+0.24
						Comprehension	+0.32	
						Rate	+0.30	
						Accuracy	+0.42	
						TOWRE		
						Sight Word	+0.13	
						Decoding	+0.12	

Note. L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; SAT-9 = Stanford Achievement Test 9th Edition; ITBS = Iowa Test of Basic Skills; ERDA = Early Reading Diagnostic Assessement, FL = Free/reduced-price lunch; W = White; AA = African American; H = Hispanic; LEP = Limited English Proficient.

The weighted mean ES for core reading curricula was only +0.06, and for supplementary curricula it was +0.08, with an overall weighted mean of +0.06. The mean for the randomized studies and randomized quasi-experiments was +0.04. The only curriculum with promising effects was *Open Court* (average ES = +0.18), but in both of the studies of this program, teachers received far more professional development than that usually provided, and in both studies *Open Court* was used for 2 and a half hours per day whereas control students had 90 minutes of reading.

Research on Instructional Technology Programs for Upper Elementary Grades

Thirty-one studies of IT for Grades 2 through 6 met the standards for this review. These were divided into three categories. *Supplemental CAI* programs, such as *Jostens/Compass Learning*, *Academy of Reading*, *LeapTrack*, *My Reading Coach*, and *CCC/Successmaker* provided additional instruction at students' assessed levels of need to supplement traditional classroom instruction. *Computer-managed learning systems* included only *Accelerated Reader*. This program uses computers to assess students' reading levels, assign reading materials at students' levels, score tests on those readings, and chart students' progress, but students do not work directly on the computer. *Innovative technology applications* included *Fast ForWord* and *Lightspan*.

Descriptions and outcomes of all studies of IT in upper elementary reading that met the inclusion criteria appear in Table 7.

Among the 31 qualifying upper elementary studies that evaluated various forms of IT, 8 used random assignment to treatments. The studies involved a total of more than 10,000 students. Overall, the sample size-weighted mean ES was very small (+0.06). The randomized evaluations ($n = 8$) had a weighted mean ES of +0.05. These findings support Kulik's (2003) conclusion that effects of CAI in reading are minimal.

None of the three categories of IT programs had convincing positive effects. Across 25 studies of supplemental programs (such as *Jostens* and *CCC*), the weighted mean ES was +0.05. Two studies of *Accelerated Reader* had a mean ES of + 0.06. Effect sizes were higher but samples were small in two studies of *Fast ForWord*, which had a mean ES of +0.21, and a small study of *Lightspan* had an ES of +0.42.

It is important to note that there is no trend toward more positive effects of IT in more recent studies. Among 11 studies reported since 2000, the weighted mean ES was only +0.06; and the large, randomized study by Dynarski et al. (2007; Campuzano et al., 2009) found no significant effects of use of a variety of modern software on the reading achievement of fourth graders (ES = +0.02). Most of the IT studies involved use of computers as supplements to regular classroom instruction, usually for about 30 minutes, one to three times a week. It may be that more intensive uses of IT would produce more robust effects, and the study of *My Reading Coach*, which provided computerized instruction 45 minutes every day and showed positive effects (ES = +0.24) in a large randomized evaluation, is a hint in this direction. Another promising use of technology is in integrated computer and noncomputer instruction, as done in *Read 180*, successfully evaluated in the middle grades (Slavin et al., 2008). However, the evidence summarized here clearly indicates that the types of supplementary CAI programs that have dominated the use of technology in education for 30 years are not producing significant effects in upper elementary reading. Many studies of IT are of high quality, and many of them involve large samples. It is difficult to imagine that such a large number of studies would fail to detect a meaningful impact if it existed.

TABLE 7
Upper Elementary Technology Programs

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Supplemental computer-assisted instruction (CAI) programs								
Academy of Reading								
Campuzano et al. (2009)	Randomized (L)	1 year	41 teachers (22 E, 19 C), 899 students (495 E, 404 C)	4	Schools across the U.S.; 65% FL, 54% AA, 29% H, 17% W	SAT-10		–0.01
LeapTrack								
Campuzano, et al. (2009)	Randomized (L)	1 year	55 teachers (29 E, 26 C), 1,274 students (665 E, 609 C)	4	Schools across the U.S.; 61% FL, 57% AA, 33% W, 10% H	SAT-10		+0.09
Jostens (Earlier form of Compass Learning)								
Alifrangis (1991)	Randomized (S)	1 year	12 classes (6 E, 6 C)	4–6	School at an army base near Washington, DC; 37% minority.	CTBS Reading		+0.15
						4th grade	+0.30	
						5th grade	+0.20	
						6th grade	–0.04	

(continued)

TABLE 7 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size
Becker (1994)	Randomized (S)	1 year	1 school, 187 students	2–5	Inner-city Baltimore, high-poverty	CAT		+0.09
Standish (1995)	Matched (S)	1 year	2 schools, 139 students (56 E, 83 C)	2	Students in suburban DE	MAT 6 Reading Comprehension		+0.05
Estep (1997)	Matched post hoc (L)	4 years	106 schools (53 E, 53 C)	3	Elementary schools in IN	ISTEP		+0.03
						Reading Vocabulary	+0.03	
						Reading Total	+0.03	
Clariana (1994)	Matched post hoc (S)	1 year	85 students (47 E, 38 C)	3	School in a predominantly White, rural area.	CTBS		+0.20
Compass Learning								
Kadel Research Consulting (2006)	Matched post hoc (S)	2 years	138 students (69 E, 224 C)	4–5	Garfield Heights, OH; 50% FL, 63% W, 24% H, 13% AA	OAT		+0.29
						1 year	–0.10	
						2 years	+0.29	
CCC Successmaker								
J. Campbell (2000)	Matched (L)	1 year	13 schools (7 E, 6 C), 701 students (310 E, 391 C)	4–5	Middle-class students in Etowah, AL	SAT		–0.02
						Comprehension	–0.09	
						Vocabulary	+0.04	

(continued)

TABLE 7 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Ragosta (1983)	Matched (L)	3 years	6 schools (4 E, 2 C), eight 1-year cohorts, three 2-year cohorts, one 3-year cohort	4–6	High-poverty schools in Los Angeles	CTBS		+0.17
						One year		
						Comprehension	+0.23	
						Vocabulary	+0.25	
						Two years		
						Comprehension	–0.01	
						Vocabulary	+0.17	
						Three years		
						Comprehension	–0.24	
						Vocabulary	+0.58	
Saracho (1982)	Matched (L)	1 year	256 students (128 E, 128 C)	3–6	Spanish-speaking migrant students	CTBS Reading		–0.09
						3rd	–0.04	
						4th	–0.25	
						5th	+0.16	
						6th	–0.17	
Classworks Gold								
Whitaker (2005)	Matched post hoc (S)	1 year	2 schools, 218 students	4, 5	Schools in rural Tennessee, 62% low SES	TCAP		–0.14
						4	–0.10	
						5	–0.19	

(continued)

TABLE 7 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size
My Reading Coach								
Vaughan et al. (2006)	Randomized (L)	1 year	4 schools, 284 students (127 E, 157 C)	2–4	Predominately minority students from 4 schools in 3 states; 27% ELLs, 36% AA, 36% H, 22% W	GRADE		+0.24
						Vocabulary	+0.24	
						Comprehension	+0.22	
WICAT								
Miller (1997)	Matched post hoc (L)	3 years	30 schools (10 E, 20 C)	3–5	New York City Public Schools, almost all AA or Hispanic, 1/6 ESL	DRP		+0.02
Clayton (1992)	Matched post hoc (L)	1 year	5 schools (1 E, 4 C), 426 students (181 E, 245 C)	2–5	Schools in northwest SC; 46% FL, 59% W, 39% AA	CTBS		–0.01
Mys & Petrie (1988)	Matched post hoc (L)	3 years	4 schools (1 E, 3 C), 257 students (81 E, 176 C)	2–4	Schools in Dearborn, MI	ITBS		–0.15
Open Book to Literacy								
Williams (2005)	Matched (S)	1 year	2 schools (1 E, 1 C), 127 students (66 E, 61 C)	4	High-poverty schools in Memphis; 51% W, 24% H, 21% AA	TORC		+0.28

(continued)

TABLE 7 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Other supplemental CAI								
Becker (1994)	Randomized (S)	1 year	9 classes, 199 students	2–5	Schools in inner-city Baltimore. 50% FL, 99% AA	CAT		+0.06
Easterling (1982) (MicroSystem 80)	Randomized (S)	4 months	2 schools, 42 students (21 E, 21 C)	5	Schools in suburban school district	CAT Reading Comprehension		+0.05
Schmidt (1991) (Wasatch ILS)	Matched (L)	1 year	4 schools (2 E, 2 C), 1,224 students (646 E, 578 C)	2–6	Schools in Southern CA. 25% FL	CTBS		+0.04
Cooperman (1985)	Matched (L)	1 year	3 schools (1 E, 2 C), 470 students (204 E, 266 C)	2–4	Students from 3 low to middle-class schools; 86% W, 13% AA	CAT		–0.06
Bryg (1984)	Matched (S)	15 weeks	9 teachers (5 E, 4 C), 152 students (83 E, 69 C)	4	Schools in Omaha, NE	CAT Reading Comprehension		+0.20
Roth & Beck (1987)	Matched (S)	1 year	6 classes (3 E, 3 C), 108 students (59 E, 49 C)	4	High-poverty low-achieving urban schools; 100% AA	Woodcock Word Attack	+0.60	+0.38
						CAT Vocabulary	+0.53	
						CAT Reading Comprehension	0.00	
Coomes (1985)	Matched (S)	1 year	4 schools, 102 students (51 E, 51 C)	4	Middle class schools in TX; 90% W	CTBS		+0.02

(continued)

TABLE 7 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Hoffman (1984)	Matched (S)	1 year	3 schools, 96 students (51 E, 45 C)	3	Schools in suburban Midwest; 11% minority	Gates MacGinitie		–0.07
						Comprehension	–0.04	
						Vocabulary	–0.10	
Levy (1985)	Matched post hoc (L)	1 year	4 schools, 581 students (293 E, 288 C)	5	Suburban NY school district	SAT		+0.19
Computer-managed learning systems								
Accelerated Reader								
Knox (1996)	Randomized (S)	3 months	77 students (40 E, 37 C)	3–4	Low-SES students in a southeastern state; 72% FL, 79% W, 13% AA, 8% H	DRS Reading Vocab	+0.25	–0.03
						DRS Reading Comp	–0.13	
						SAT Reading Vocab	–0.07	
						SAT Reading Comp	–0.17	
Yee (2007)	Matched (L)	1 year	3 schools (1 E, 2 C), 2,072 students (612 E, 1,460 C)	2–5	Majority-Hispanic schools in Los Angeles Co.; 92% FL, 79% H, 17% AA, 61% ELL	CST		+0.06

(continued)

TABLE 7 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Innovative technology applications								
Fast ForWord								
Marion (2004)	Matched (L)	1 year	349 students (215 E, 134 C)	5–6	Schools in Appalachian TN; 52% FL, 100% W	Terra Nova		+0.25
Scientific Learning (2006)	Matched (S)	15 weeks	142 students (55 E, 87 C)	5–6	Middle-class schools in northwest OH	Gates MacGinitie		+0.11
						Comprehension	+0.12	
						Vocabulary	+0.11	
Lightspan								
Birch (2002)	Matched post hoc (S)	2 years	101 students (50 E, 51 C)	2, 3	Schools in the Caesar Rodney School District in DE	SAT		+0.42
						Vocabulary	+0.59	
						Comprehension	+0.25	

Note. L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; CTBS = Comprehensive Test of Basic Skills; CAT = California Achievement Test; CST = California Standards Test; MAT = Metropolitan Achievement Test; ITBS = Iowa Test of Basic Skills; ISTEP = Indiana Statewide Testing for Educational Progress; OAT = Ohio Achievement Test; CCC = Computer Curriculum Corporation; TCAP = Tennessee Comprehensive Assessment Program; GRADE = Group Reading Assessment and Diagnostic Examination; DRP = Degrees of Reading Power; WRAT = Wide Range Achievement Test; SAT = Scholastic Achievement Test; DRS = Diagnostic Reading Scales; FL = Free/reduced price lunch; W = White, AA = African American, H = Hispanic, ELL = English language learners; LEP = Limited English Proficient.

Research on Upper Elementary Instructional Process Programs

Instructional process programs are methods that focus on providing teachers with extensive professional development to implement specific instructional methods. In upper elementary reading, instructional process programs are quite diverse. Thirty-three studies, 6 of which used random assignment, evaluated a broad range of approaches. *Cooperative learning programs* (Slavin, 1995, 2009; Webb, 2008) use methods in which students work in small groups to help one another master academic content.

Strategy instruction programs teach students cognitive and metacognitive skills such as summarization, graphic organizers, and prediction to help them comprehend text. Strategy instruction is often combined with other methods, especially cooperative learning and peer tutoring. *Structured phonetic intervention programs* are approaches emphasizing phonics, systematic instruction, and frequent assessment of student progress. *Phonics-focused professional development programs* are ones that teach teachers the NRP elements, especially phonics and phonemic awareness, mostly in workshops. *Integrated language arts programs* are less structured and less phonetic and focus on integrating reading and writing, literature study, and pleasure in reading. *Cross-age tutoring programs* involve older children working with younger ones, and *same-age tutoring* involves having children take turns tutoring one another. *Classroom management and motivation programs* focus on building a positive learning environment.

Descriptions and outcomes of all studies of upper elementary instructional process programs meeting the inclusion criteria appear in Table 8.

Both the methods and the findings of instructional process programs for upper elementary reading were quite diverse. Across 33 experimental-control comparisons, involving more than 17,000 students, the weighted mean ES was +0.21. These include 4 randomized and 2 randomized quasi-experiment studies.

Ten of the studies evaluated two forms of cooperative learning. These had a weighted mean ES of +0.21. All but one of the cooperative learning studies evaluated *CIRC*, which involves students in well-structured cooperative groups within which they help each other master and apply metacognitive learning strategies. *CIRC* was the basis for middle school reading programs called *Student Team Reading* and *The Reading Edge*, which had a weighted mean ES of +0.29 in four secondary studies. The consistent positive effects of this family of cooperative learning approaches support the idea that programs focusing on professional development in structured activities that engage children in discussions about reading, giving them opportunities to help each other learn and use metacognitive skills, may have particular promise for enhancing reading achievement from the second grade onward. Positive effects were also found for cross-age tutoring programs (ES = +0.26 in four studies) and for same-age tutoring (ES = +0.26 in two studies), reinforcing the conclusion that structuring interaction among students on reading strategies is an effective approach. Another promising category was programs emphasizing metacognitive strategy instruction, such as *Reciprocal Teaching* and *Thinking Maps*, which had a weighted mean ES of +0.32 in five studies. In these programs, students were taught skills such as prediction, summarization, and self-evaluation.

TABLE 8
Upper Elementary Instructional Process Programs

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Cooperative learning								
Cooperative Integrated Reading and Composition (CIRC)								
Stevens and Slavin (1995b)	Matched (L)	2 years	7 schools (3 E, 4 C), 63 classes (31 E, 32 C), 1,299 students (635 E, 664 C)	2–6	Mostly White students in working-class suburb of Baltimore; 9% FL, 95% W	CAT		+0.23
						Vocabulary	+0.20	
						Comprehension	+0.26	
Stevens & Slavin (1995a)	Matched (L)	2 years	5 schools (2 E, 3 C), 45 classes (21 E, 24 C), 873 students (411 E, 462 C)	2–6	Suburban district in Maryland; 10% FL, 93% W	CAT		+0.25
						Comprehension	+0.28	
						Vocabulary	+0.21	
Jenkins et al. (1994)	Matched (L)	1 year	2 schools, 860 students (332 E, 528 C)	1–6	Mount Vernon, WA; 36% FL	MAT		+0.18
						Comprehension	+0.09	
						Vocabulary	+0.31	
						Total	+0.18	
Stevens et al. (1987; Study 1)	Matched (L)	12 weeks	10 schools (6 E, 4 C), 21 classes (11 E, 10 C)	3–4	Middle-class suburb of Baltimore; 4% FL, 84% W, 16% AA	CAT		+0.18
						Comprehension	+0.19	
						Vocabulary	+0.17	

(continued)

TABLE 8 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Stevens et al. (1987; Study 2)	Matched (L)	6 months	9 schools (4 E, 5 C), 22 classes (9 E, 13 C), 450 students	3–4	Middle-class suburb of Baltimore; 18% FL, 78% W, 22% AA	CAT		+0.45
						Comprehension	+0.35	
						Vocabulary	+0.11	
						Total	+0.23	
						Durrell	+0.54	
Bramlett (1994)	Matched (L)	1 year	8 schools (9 C, 9 E), 18 classes, 392 students (198 E, 194 C)	3	Rural southern Ohio	CAT		+0.08
						Comprehension	+0.10	
						Total Reading	+0.07	
						Word Analysis	+0.10	
						Vocabulary	+0.03	
Rapp (1991)	Matched (S)	1 year	2 schools (1 E, 1 C), 88 students (43 E, 45 C)	3	Working-class schools in Lewistown, ID	ITBS		+0.14
						Comprehension	+0.09	
						Vocabulary	+0.18	
Calderon et al. (1998)	Matched (S)	2 years	7 schools (3 E, 4 C) Year 1: 84 students (51 E, 33 C); Year 2: 59 students (26 E, 33 C)	2 and 3	Spanish-dominant students transitioning to English in high-poverty schools near the Mexican border in Texas; 79% H	STAAS 2nd graders	+0.30	+0.87
						NAPT 3rd graders		
						1 year	+0.62	
						2 years	+0.87	

(continued)

TABLE 8 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Skeans (1991)	Matched post hoc (L)	19 months	630 students (348 E, 282 C)	3 and 5	Suburban district near Houston	MAT: 3rd grade		–0.03
						Vocabulary	+0.20	
						Comprehension	+0.08	
						MAT: 5th grade		
						Vocabulary	–0.15	
						Comprehension	–0.24	
Reader's Theater								
Carrick (2000)	Matched (S)	14 weeks	98 students (53 E, 45 C)	5	Urban New Jersey; 80% FL, 85% AA, 11% H	Compared to control		+0.29
						Terra Nova	+0.22	
						Oral Reading	+0.50	
						Compared to paired reading		
						Terra Nova	+0.12	
						Oral Reading	+0.30	
Same-age tutoring programs								
PALS								
Fuchs et al. (1999)	Randomized quasi-experiment (S)	21 weeks	45 students; 15 students each in PALS, PALS-HG (PALS + tutoring strategies), or control	2–3	Students in a southeastern city; 24% FL, 62% W, 38% AA	SDRT Reading Comprehension		+0.36
						PALS	+0.72	
						PALS HG	0.00	

(continued)

TABLE 8 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Same-Age Tutoring + Strategy Instruction								
Van Keer & Verhaeghe (2005)	Matched (L)	1 year	Second graders: 11 classes (5 E, 6 C), 215 students (91 E, 124 C); Fifth graders: 10 classes (4 E, 6 C), 208 students (101 E, 107 C)	2, 5	Middle class schools in Flanders, Belgium	Dutch Reading Comprehension Test		+0.29
						2nd graders	+0.17	
						5th graders	+0.40	
Van Keer & Verhaeghe (2008)	Matched (L)	1 year	Second graders: 12 classes (6 E, 6 C), 234 students (110 E, 124 C); Fifth graders: 15 classes, (9 E, 6 C), 293 students (186 E, 107 C)	2, 5	Middle class schools in Flanders, Belgium	Dutch Reading Comprehension Test		+0.24
						2nd graders	+0.26	
						5th graders	+0.21	
Cross-age tutoring programs								
Reading Together								
Policy Studies Associates (2007)	Randomized (S)	1 year	124 students (56 E, 68 C)	2	School in Irving, TX	Terra Nova		–0.01
Cross-Age Tutoring								
Hilger (2000)	Matched (S)	1 year	1 school 72 students (47 E, 35 C)	3	High-poverty school; 78% FL; 34% AA, 34% Asian, 26% W, 5% H	STAR	+0.16	+0.37
						Fluency	+0.58	

(continued)

TABLE 8 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size
Cross-Age Tutoring + Strategy Instruction								
Van Keer & Verhaeghe (2005)	Matched (L)	1 year	Second graders: 9 classes (3 E, 6 C), 190 students (66 E, 124 C); Fifth graders: 10 classes (4 E, 6 C), 276 students (169 E, 107 C)	2, 5	Middle-class schools in Flanders, Belgium	Dutch Reading Comprehension Test		+0.27
						2nd graders	+0.22	
						5th graders	+0.32	
Van Keer & Verhaeghe (2008)	Matched (L)	1 year	Second graders: 14 classes (8 E, 6 C), 286 students (162 E, 124 C); Fifth graders: 13 classes (7 E, 6 C), 263 students (156 E, 107 C)	2, 5	Middle class schools in Flanders, Belgium	Dutch Reading Comprehension Test		+0.35
						Second graders	+0.42	
						Fifth graders	+0.28	
Strategy instruction								
Belgian Strategy Model								
Van Keer & Verhaeghe (2005)	Matched (L)	1 year	Second graders: 14 classes (8 E, 6 C), 287 students (163 E, 124 C); Fifth graders: 14 classes (8 E, 6 C), 284 students (177 E, 107 C)	2, 5	Middle-class schools in Flanders, Belgium	Dutch Reading Comprehension Test		+0.30
						Second graders	+0.24	
						Fifth graders	+0.35	

(continued)

TABLE 8 (continued)

Study	Design (Large/Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size
Thinking Maps								
Leary (1999)	Matched (S)	1 year	2 schools (1 E, 1 C), 78 students (41 E, 37 C)	4	High-poverty schools in southeastern Virginia; 79% FL; 69% AA, 31% W	SAT-9		+0.31
Hickie (2006)	Matched post hoc (S)	2 years	2 schools (1 E, 1 C), 54 students (24 E, 30 C)	4–5	High-poverty White schools in northeastern Tennessee; 91% FL	TCAP		+0.70
Foundations and Frameworks								
Blackmon (2008)	Matched (S)	1 year	5 schools (3 E, 2 C), 103 students (52 E, 51 C)	4–5	Philadelphia Christian schools, mostly AA, H	Gates MacGinitie		–0.02
						Comprehension	–0.08	
						Vocabulary	+0.04	
Reciprocal Teaching								
Spörer, Brunstein, & Kieschke (2007)	Matched (S)	19 weeks	105 students	3–6	Middle-class schools in Germany	German standardized comprehension test		+0.57

(continued)

TABLE 8 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Fluency instruction								
FORI								
Kuhn et al. (2006)	Randomized quasi-experiment (S)	1 year	5 schools (3 E, 2 C), 227 students (143 E, 84 C)	2	High-poverty schools in New Jersey and Georgia; 58% FL, 51% AA, 23% W, 21% H, 5% Asian	TOWRE GORT—4 WIAT	+0.29 +0.10 +0.18	+0.19
Structured phonetic intervention programs								
Exemplary Center for Reading Instruction (ECRI)								
Reid (1996)	Matched post hoc (L)	1 year	5 schools (4 E, 1 C), 921 students (590 E, 331 C)	2–6	High-poverty White schools in eastern Tennessee; 99% W	SAT Comprehension Vocabulary	 +0.71 +0.59	+0.65
Cohen (1991)	Matched post hoc (L)	1 year	473 students (242 E, 231 C)	3	Urban school district; 45% AA, 34% W, 21% H	ITBS Comprehension Vocabulary	 +0.07 +0.21	+0.14
Phonics-based professional development								
Language Essentials for Teachers of Reading and Spelling (LETRS)								
Garet et al. (2008)	Randomized (L)	1 year	90 schools, 5,530 students (1,983 LETRS, 1,738 LETRS + Coaching, 1,809 C)	2	6 urban districts; 78% FL, 78% AA, 15% W, 5% H	Various state assessments LETRS LETRS + Coaching	 +0.08 +0.03	+0.06

(continued)

TABLE 8 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Integrated language arts programs								
Literature-Based Program								
Morrow (1992)	Randomized quasi-experiment (S)	1 year	9 classes, 166 students (56 LBP + parents, 46 LBP only, 64 C)	2	Students in two suburban schools in New Jersey; 24% FL, 43% AA, 37% White, 14% Asian	CAT		+0.21
						School + home	+0.21	
						School only	+0.20	
Success in Reading and Writing								
Lindsey (1988)	Matched (S)	1 year	2 schools (1 E, 1 C), 97 students (56 E, 41 C)	2–3	Elementary schools in the Pacific Northwest	CAT		–0.11
						Comprehension	–0.23	
						Vocabulary	+0.01	
Carbo Reading Styles								
Oglesby & Suter (1995)	Matched (S)	1 year	13 classes (6 E, 7 C), 198 students (105 E, 93 C)	3 and 6	Urban, mostly AA school in the mid-South; 80% AA, 20% W, 81% remedial	Gates MacGinitie		+0.27
Classroom management and motivation programs								
Consistency Management-Cooperative Discipline (CMCD)								
Freiberg et al. (1990)	Matched post hoc (L)	2 years	10 schools (5 E, 5 C), 699 students (364 E, 335 C)	2–5	High-poverty schools in Houston; 72% FL, 90% AA	MAT-6 (Grades 2–5)	+0.09	+0.12
						TEAMS (Grades 3 and 5)	+0.14	

(continued)

TABLE 8 (continued)

Study	Design (Large/ Small)	Duration	*N*	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
Opuni (2006)	Matched post hoc (L)	1 year	14 schools (7 E, 7 C), 456 students (228 E, 228 C)	3	High-poverty schools in Newark, NJ; 78% FL, 90% AA	SAT-9		+0.26
Student Success Skills								
C. Campbell and Brigman (2005)	Randomized (L)	6 months	20 schools, 480 students (240 E, 240 C)	5-6	Low-achieving students in Florida; 62% FL, 82% W, 9% AA, 5% H	FCAT		+0.23
Responsive Classroom								
Rimm-Kaufman et al. (2007)	Matched post hoc (L)	3 years	6 schools (3 E, 3 C), 3 groups: Grades 2–5, 381 students (211 E, 170 C); Grades 3–5, 502 students (282 E, 220 C); Grades 4–5, 506 students (266 E, 240 C)	2–5	Schools in a northeastern urban district; 35% FL, 57% W, 22% AA, 21% H	DRP		+0.15
						Grades 2–5	+0.21	
						Grades 3–5	+0.16	
						Grades 4–5	+0.07	

Note. L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; CAT = California Achievement Test; MAT = Metropolitan Achievement Test; ITBS = Iowa Tests of Basic Skills; STAAS = Texas Assessment of Academic Skills–Spanish; NAPT-Norm-Referenced Assessment Program for Texas; SDRT = Stanford Diagnostic Reading Test; SAT = Stanford Achievement Test; TCAP = Tennessee Comprehensive Assessment Program; PALS = Peer-Assisted Learning Strategies; PALS-HG = Peer-Assisted Learning Strategies with Help-Giving Training; TOWRE = Test of Word Reading Efficiency; GORT = Gray Oral Reading Test; GRADE = Group Reading Assessment and Diagnostic Examination; STAR = Standardized Test for Assessment of Reading; WIAT = Wechsler Individual Achievement Test; TEAMS = Texas State Assessment of Academic Skills; SAT = Scholastic Achievement Test; DRP = Degrees of Reading Power; FCAT = Florida's Comprehensive Assessment Test; FL = Free/Reduced lunch, W = White, AA = African American, H = Hispanic, CTBS = Comprehensive Test of Basic Skills.

It is important to note that additional instructional process programs also showed positive effects, but because the studies evaluating these approaches involved small groups of struggling readers rather than students in general, these findings are reviewed by Slavin et al. (2009). These include *DISTAR/Corrective Reading*, *PALS*, and *Empower Reading*.

Overall Patterns of Outcomes: Upper Elementary Reading

Across all categories, there were 79 qualifying studies of upper elementary school reading programs involving a total of more than 32,000 students, of which 23 used random assignment (16 were fully randomized and 7 were randomized quasi-experiments). The overall sample size-weighted mean ES was +0.13. The mean ESs of +0.06 for reading curricula and +0.06 for technology contrast with a mean of +0.21 for instructional process programs, such as cooperative learning and strategy instruction, reinforcing the findings of the beginning reading review.

Outcomes for High-Poverty Schools

An important question for policy and practice is whether effects of various programs are particularly strong or weak for students in high-poverty schools. To examine this question, schools in each study were defined as "high-poverty" if at least 50% of their students qualified for free or reduced-price lunches, or if other information in the study (such as a description of schools as serving high-poverty neighborhoods) indicated high-poverty status. Forty-one beginning reading and 31 of the upper elementary studies involved high-poverty schools, by this definition. At beginning and upper elementary grade levels, outcomes were very similar for high-poverty schools (mean ES = +0.15) and low-poverty schools (mean ES = +0.14). Among the studies of reading curricula, weighted mean ESs were +0.07 (n = 14) for high-poverty schools and +0.09 (n = 8) for low-poverty schools. For IT, the weighted mean effect sizes were +0.08 (n = 17) for high-poverty schools and +0.06 (n = 26) for low-poverty schools. Among studies of instructional process programs, including beginning reading programs that combine instructional process and curriculum, the weighted mean ESs were +0.27 (n = 45) for high-poverty schools and +0.20 (n = 31) for low-poverty schools.

As in the overall set of studies, the studies of high-poverty schools supported the observation that programs that provide extensive professional development to teachers in specific classroom strategies are most likely to make a difference in the achievement of students in high-poverty schools. From a policy perspective, what these findings imply is that proven models could be used effectively in any type of school, but to reduce gaps according to socioeconomic status, these programs should be particularly encouraged among high-poverty Title I schools.

Overall Discussion

The research reviewed in this article provides reason for optimism about the improvement of basic reading instruction in the elementary grades. Sixty-three studies of beginning reading programs and 79 studies of upper elementary reading programs met stringent methodological requirements, and these studies provide support for many replicable approaches. More research on a larger set of programs is needed, of course, but the research that already exists provides educators and policy makers with several robust approaches they could choose to improve their students' reading performance. Those programs have been shown to be effective in high-poverty as well as less disadvantaged schools, so if the effective programs were implemented with integrity by many schools serving disadvantaged students, this could significantly reduce achievement gaps between middle class and lower class children. The research also identified types of approaches that have not been successful in improving elementary reading performance.

There are several important patterns in the findings that are worthy of note. First, for both beginning reading and upper elementary reading, this article finds extensive evidence supporting forms 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. In beginning reading, examples of cooperative learning included *PALS*, and cooperative learning is a key component of *Success for All*. In upper elementary reading, the category is primarily represented by *CIRC*. Positive effects for studies of cross-age and same-age tutoring at all grade levels also reinforce the value of engaging students in structured peer-to-peer interactions. The finding of positive effects of cooperative learning programs is consistent with the findings of reviews of secondary reading programs (Slavin et al., 2008) and elementary and secondary math programs (Slavin & Lake, 2008; Slavin, Lake, & Groff, in press).

Also consistent with previous reviews is the finding that both alternative curricula and IT generally produced small effects on reading measures at all grade levels. In particular, the evidence did not support the idea that simply introducing materials or training with a strong emphasis on phonics will significantly improve reading outcomes. Effects of adopting phonetic textbooks were very small, and a large study of *Language Essentials for Teachers of Reading and Spelling* (*LETRS*), a professional development program focused on phonics, also found disappointing results (Garet et al., 2008). These findings suggest that whereas phonics appears necessary in reading instruction, adding a phonics focus is not enough to increase reading achievement.

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 together every day. These programs are characterized by extensive professional development in classroom strategies intended to maximize students' participation and engagement, give them effective metacognitive strategies for comprehending text, and strengthen their phonics skills. As in earlier reviews, such strategies had outcomes that were clearly and consistently more positive than those found for curricula or IT alone. These positive effects were found equally for high-poverty and low-poverty schools, and they were found on comprehension as well as decoding measures. More research and development of reading programs for elementary students is clearly needed, but this review identifies several promising approaches that could be used today to help students succeed in reading in the elementary grades.

Note

We thank Marilyn Adams, Steven Ross, Michael McKenna, Henry Becker, and Nancy Madden for comments on an earlier draft. The first author is one of

the developers of Success for All and CIRC, the third author is one of the developers of Reading Reels, and the last author works for the Success for All Foundation.

References

Abram, S. L. (1984). *The effect of computer assisted instruction on first grade phonics and mathematics achievement computation*. Unpublished doctoral dissertation, Northern Arizona University, Flagstaff.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Alifrangis, C. M. (1991). An integrated learning system in an elementary school: Implementation, attitudes, and results. *Journal of Computing in Childhood Education*, *2*(3), 51–66.

Apthorp, H. (2005a). *Elements of reading: Phonics and phonemic awareness*. Orlando, FL: Harcourt.

Apthorp, H. (November, 2005b). *A study of the effects of Harcourt Achieve's Elements of Reading: Fluency*. Denver, CO: Mid-Continent Research for Education and Learning.

Apthorp, H. (2005c, December). *A study of the effects of Harcourt Achieve's Elements of Reading: Vocabulary*. Denver, CO: Mid-Continent Research for Education and Learning.

August, D., & Shanahan, T. (2006). Synthesis: Instruction and professional development. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners* (pp. 351–364). Mahwah, NJ: Erlbaum.

Baer, J., Baldi, S., Ayotte, K., Green, P. J., & McGrath, D. (2007). *The reading literacy of U.S. fourth-grade students in an international context: Results from the 2001 and 2006 Progress in International Reading Literacy Study (PIRLS)*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Barnett, L. B. (2006). *The effect of computer-assisted instruction on the reading skills of emergent readers*. Unpublished doctoral dissertation, Florida Atlantic University, Boca Raton.

Barrett, T. J. (1995). *A comparison of two approaches to first grade phonics instruction in the Riverside Unified School District*. Paper presented at the annual meeting of the California Educational Research Association, Lake Tahoe.

Beasley, N. (1989). *The effects of the IBM Writing to Read program on the achievement of selected first grade students*. Dissertation Abstracts International, 51 (3), 739A. (UMI No 9122247).

Becker, H. J. (1994). Mindless or mindful use of integrated learning systems. *International Journal of Educational Research*, *21*(1), 65–79.

Birch, J. (2002). *The effects of the Delaware Challenge Grant Program on the standardized reading and mathematics test scores of second and third grade students in the Caesar Rodney School District*. Unpublished doctoral dissertation, Wilmington College, Wilmington, DE.

Blachman, B. A., Tangel, D., Ball, E., Black, R., & McGraw, C. (1999). Developing phonological awareness and word recognition skills: A two-year intervention with low-income, inner-city children. *Reading and Writing: An Interdisciplinary Journal*, *11*, 239–273.

Blackmon, C. M. (2008). *The efficacy of Foundations & Frameworks on elementary students' reading achievement in urban Christian schools*. Unpublished doctoral dissertation, Liberty University, Lynchburg, VA.

Bond, C., Ross, S. M., Smith, L. J., & Nunnery, J. A. (1995). The effects of the Sing, Spell, Read, and Write program on reading achievement of beginning readers. *Reading Research and Instruction*, *35*, 122–141.

Borman, G., & Dowling, N. (2007). *Student and teacher outcomes of the Superkids quasi-experimental study*. Madison, WI: University of Wisconsin.

Borman, G. D., Dowling, N. M., & Schneck, C. (2007). *The national randomized field trial of Open Court Reading*. Madison: University of Wisconsin Press.

Borman, G. D., & Hewes, G. (2003). Long-term effects and cost effectiveness of Success for All. *Educational Evaluation and Policy Analysis*, *24*(2), 243–266.

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.

Borman, G. D., & Rachuba, L. T. (2001, August). *Evaluation of the Scientific Learning Corporation's Fast ForWord computer-based training program in the Baltimore City Public Schools*. Report prepared for the Abell Foundation.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, *44*(3), 701–731.

Bramlett, R. K. (1994). Implementing cooperative learning: A field study evaluating issues for school-based consultants. *Journal of School Psychology*, *32*(1), 67–84.

Brown, I. S., & Felton, R. H. (1990). Effects of instruction on beginning reading skills in children at risk for reading disability. *Reading and Writing*. *An Interdisciplinary Journal*, *2*, 223–241.

Bryg, V. (1984). *The effect of computer assisted instruction upon reading achievement with selected fourth grade children*. Unpublished doctoral dissertation, University of Nebraska, Lincoln.

Calderon, M., Hertz-Lazarowitz, R., & Slavin, R. E. (1998). Effects of bilingual cooperative integrated reading and composition on students making the transition from Spanish to English reading. *Elementary School Journal*, *99*(2), 153–165.

Calhoon, M., Al Otaiba, S., Cihak, D., King, A., & Avalos, A. (2007). The effects of a peer-mediated program on reading skill acquisition for two-way bilingual first-grade classrooms. *Learning Disability Quarterly*, *30*(3), 169–184.

Calhoon, M., Otaiba, S., Greenberg, D., King, A., & Avalos, A. (2006). Improving reading skills in predominately Hispanic Title I first grade classrooms: The promise of Peer-Assisted Learning Strategies. *Learning Disabilities Research and Practice*, *21*(4), 261–272.

Campbell, C., & Brigman, G. (2005). Closing the achievement gap: A structured approach to group counseling. *Journal for Specialists in Group Work*, *30*(1), 67–82.

Campbell, J. P. (2000). *A comparison of computerized and traditional instruction in the area of elementary reading*. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts*. Washington, DC: U.S. Department of Education.

Carbo, M. (1982). *Carbo Reading Styles Inventory*. New York: Learning Research Associates.

Carbo, M., Dunn, R., & Dunn, K. (1986). *Teaching students to read through their individual learning styles*. New York: Allyn & Bacon.

Carlson, C. D., & Francis, D. J. (2002). Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE). *Journal of Education for Students Placed at Risk*, *7*(2), 141–166.

Carrick, L. U. (2000). *The effects of Reader's Theater on fluency and comprehension of fifth grade students in regular classes*. Unpublished doctoral dissertation, Lehigh University, Bethlehem, PA.

Casey, J., Smith., L., & Ross, S. (1994). *Final report: 1993–94 Success for All program in Memphis*, *Tennessee: Formative evaluation of new SFA schools*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Cassady, J., & Smith, L. (2005). The impact of a structured integrated learning system on first grade students' reading gains. *Reading and Writing Quarterly*, *21*(4), 361–376.

Chall, J. S. (1983). Literacy: Trends and explanations. *Educational Researcher*, *12*(9), 3–8.

Chambers, B., Cheung, A., Madden, N., Slavin, R. E., & Gifford, R., (2006). Achievement effects of embedded multimedia in a Success for All reading program. *Journal of Educational Psychology*, *98*(1), 232–237.

Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P.C., Tucker, B. J. Cheung, A., et al. (2008). Technology infusion in Success for All: Reading outcomes for first graders. *Elementary School Journal*, *109*(1), 1–15.

Chambers, B., Slavin, R. E., Madden, N. A., Cheung, A., & Gifford, R. (2004). *Enhancing Success for All for Hispanic students: Effects on beginning reading achievement* (Technical Rep.). Baltimore: Johns Hopkins University, Center for Date-Driven Reform in Education.

Chambers, B., Slavin, R. E., Madden, N. A., Cheung, A., & Gifford, R. (2005). *Effects of Success for All with embedded video on the beginning reading achievement of Hispanic children* (Technical Rep.). Baltimore: Center for Research and Reform in Education, Johns Hopkins University.

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.

Clariana, R. B. (1994). *The effects of an integrated learning system on third graders' mathematics and reading achievement*. San Diego, CA: Jostens Learning Corporation. (ERIC Document Reproduction Service No. ED 409 181)

Clayton, I. L. (1992). *The relationship between computer-assisted instruction in reading and mathematics achievement and selected student variables*. Unpublished doctoral dissertation, University of Southern Mississippi, Hattiesburg.

Cohen, K. (1991). *A comparative study of reading instruction management for selected third-grade students in an urban school district*. Unpublished doctoral dissertation, University of North Texas, Denton.

Collis, B., Ollila, L, & Ollila, K. (1990). Writing to Read: An evaluation of a Canadian installation of a computer-supported initial language environment. *Journal of Educational Computing Research, 6*(4), 411-27.

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

Conner, J. M., Greene, B. G., & Munroe, K. (2004). *An experimental study of the instructional effectiveness of the Harcourt Reading Program in academically at-risk* *schools in the Philadelphia City School District*. Bloomington, Indiana: Educational Research, Institute of America.

Cook, T., Shadish, W. R., & Wong, V. C. (2008). *Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within study comparisons*. Paper presented at the annual meetings of the Society for Research on Effective Education, Crystal City, VA.

Coomes, P. (1985). *The effects of computer assisted instruction on the development of reading and language skills*. Unpublished doctoral dissertation, North Texas State University, Denton.

Cooper, H. (1998). *Synthesizing research* (3rd ed.). Thousand Oaks, CA: Sage.

Cooperman, K. S. (1985). *An experimental study to compare the effectiveness of a regular classroom reading program to a regular classroom reading program with a computer-assisted instruction program in reading comprehension skills in grades two through four*. Unpublished doctoral dissertation, American University, Washington, DC.

Correnti, R. (2009). *Examining CSR program effects on student achievement: Causal explanation through examination of implementation rates and student mobility*. Paper presented at the 2nd annual conference of the Society for Research on Educational Effectiveness, Washington, DC, March, 2009.

D'Agostino, J. (2009). *The effectiveness of the Superkids on student achievement and teacher outcomes*. Columbus, OH: Ohio State University.

Deshler, D., Palincsar, A., Biancarosa, G., & Nair, M. (2007). *Informed choices for struggling adolescent readers*. Newark, DE: International Reading Association.

Dianda, M., & Flaherty, J. (1995, April). *Effects of Success for All on the reading achievement of first graders in California bilingual programs*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: Institute of Education Sciences.

Easterling, B. (1982). *The effects of computer assisted instruction as a supplement to classroom instruction in reading comprehension and arithmetic*. Unpublished doctoral dissertation, North Texas State University, Denton.

Erdner, R., Guy, R., & Bush, A. (1997). The impact of a year of computer assisted instruction on the development of first grade reading skills. *Journal of Educational Computing Research*, *18*(4), 369–388.

Estep, S. (1997). *An investigation of the relationship between integrated learning systems and academic achievement*. Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, *90*, 37–55.

Frechtling, J., Zhang, X., & Silverstein, G. (2006). The Voyager Universal Literacy System: Results from a study of kindergarten students in inner-city schools. *Journal of Education for Students Placed at Risk*, *11*(1), 75–95.

Freiberg, H. J., Prokosch, N., Tresister, E. S., & Stein, T. (1990). Turning around five at-risk elementary schools. *Journal of School Effectiveness and School Improvement*, *1*(1), 5–25.

Fuchs, D., Fuchs, L., Mathes, G., & Simmons, D. (1997). Peer-Assisted Learning Strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*(1), 174–206.

Fuchs, D., Fuchs, S., Thompson, A., Al-Otaiba, S., Yen, L., Yang, N., et al. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*(2), 251.

Fuchs, L., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of Peer-Assisted Learning Strategies in reading with and without training in elaborated help giving. *Elementary School Journal*, *99*(3), 201–220.

Gamse, B. C., Tepper-Jacob, R., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First impact study: Final report*. Washington, DC: Institute for Education Sciences, U.S. Department of Education.

Garet, M., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W. et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. New York: MDRC.

Granick, L., & Reid, E. (1987). *Writing to Read program*, *FY 87*. Baltimore: Baltimore City Public Schools.

Grant, E. M. (1973). *A study of comparison of two reading programs (Ginn 360 and DISTAR) upon primary inner city students*. Unpublished doctoral dissertation, University of Washington, Seattle.

Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of Classwide Peer Tutoring. *Journal of Educational Psychology*, *81*(3), 371–383.

Greenwood, C. R., Terry, B., Utley, C. A., Montagna, D., & Walker, D. (1993). Achievement, placement, and services: Middle school benefits of Classwide Peer Tutoring used at the elementary level. *School Psychology Review*, *22*(3), 497–516.

Hecht, S. (2003). *A study between Voyager and control schools in Orange County, Florida 2002–2003*. Davie: Florida Atlantic University.

Hecht, S., & Close, L. (2002). Emergent literacy skills and training time uniquely predict variability in responses to phonemic awareness training in disadvantaged kindergartners. *Journal of Experimental Child Psychology*, *82*, 93–115.

Herman, R. (1999). *An educator's guide to schoolwide reform*. Arlington, VA: Educational Research Service.

Hickey, K. (2006). *An examination of student performance in reading/language and mathematics after two years of Thinking Maps® implementation in three Tennessee schools*. Unpublished doctoral dissertation, East Tennessee State University.

Hilger, L. (2000). *Cross-age tutoring in reading: Academic and attitudinal effects from high-school tutors and third-grade tutees*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis–St. Paul.

Hoffman, J. T. (1984). *Reading achievement and attitude toward reading of elementary students receiving supplementary computer assisted instruction compared with students receiving supplementary traditional instruction*. Unpublished doctoral dissertation, Ball State University, Muncie, IN.

Huxley, A. (2006). *A text-based intervention of reading fluency, comprehension, and content knowledge*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Jenkins, J., Jewell, M., Leicester, N., O'Connor, R., Jenkins, L., & Troutner, N. (1994). Accommodations for individual differences without classroom groups: An experiment in school restructuring. *Exceptional Children*, *60*(4), 344–358.

Jones, E. M., Gottfredson, G. D., & Gottfredson, D. C. (1997). Success for some: An evaluation of the Success for All program. *Evaluation Review*, *21*(6), 643–670.

Jones, L. R. G. (1995). *The effects of an eclectic approach versus a modified whole language approach on the reading and writing skills of first-grade students*. Unpublished doctoral dissertation, University of Mississippi, University.

Joshi, R. M., Dahlgren, M., & Boulware-Gooden, R. (2002). Teaching reading in an inner city school through a multisensory teaching approach. *Annals of Dyslexia*, *52*(1), 229.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437–447.

Kadel Research Consulting. (2006). *Garfield Heights City Schools Maple Leaf intermediate, enhancing education through technology, end-year evaluation*. Hyde Park, VT: Kadel Research Consulting.

Kennedy, M. M. (1978). Findings from the Follow Through Planned Variation study. *Educational Researcher*, *7*(6), 3–11.

Knox, M. (1996). An experimental study of the effects of the Accelerated Reader program and a teacher directed program on reading comprehension and vocabulary of fourth and fifth grade students. *Dissertation Abstracts International*, *57*(10), 4208A. (UMI No. 9710798)

Kuhn, M. R., Schwanenflugel, P. J., Morris, R. D., Morrow, L. M., Woo, D. G., Meisinger, E. B., et al. (2006). Teaching children to be fluent and automatic readers. *Journal of Literacy Research*, *38*(4), 357–387.

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

Leary, S. F. (1999). *The effect of Thinking Maps® instruction on the achievement of fourth-grade students*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg.

Levy, M. H. (1985). *An evaluation of computer assisted instruction upon the achievement of fifth grade students as measured by standardized tests*. Unpublished doctoral dissertation, University of Bridgeport, CT.

Lie, A. (1991). Effects of a training program for stimulating skills in word analysis in first-grade children. *Reading Research Quarterly*, *26*, 234–250.

Lindsey, M. M. (1998). *A comprehensive evaluation of an integrated reading and language arts curriculum, with attention to the experiences of low achieving children*. Unpublished doctoral dissertation, University of Oregon, Eugene.

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

Livingston, M., & Flaherty, J. (1997). *Effects of Success for All on reading achievement in California schools*. Los Alamitos, CA: WestEd.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, *23*, 263–284.

Macaruso, P., Hook, P. E., & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading*, *29*, 162–172.

Mac Iver, M., Kemper, E., & Stringfield, S. (2003). *The Baltimore Curriculum Project: Final report of the four-year evaluation study*. Baltimore: Johns Hopkins University, Center for Social Organization of Schools.

Madden, N. A., Slavin, R. E., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal*, *30*, 123–148.

Marion, G. G. (2004). *An examination of the relationship between students' use of the Fast ForWord Reading Program and their performance on standardized assessments in elementary schools*. Unpublished doctoral dissertation, East Tennessee State University, Johnson City.

Mathes, P., & Babyak, A. (2001). The effects of Peer-Assisted Literacy Strategies for first-grade readers with and without additional mini-skills lessons. *Learning Disabilities Research & Practice*, *16*(1), 28–44.

Mathes, P., Torgesen, J., & Allor, J. (2001). The effects of Peer-Assisted Literacy Strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal*, *38*(2), 371-410.

Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (l998). Peer-Assisted Learning Strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly*, *33*, 62–94.

Mathes, P. G., Torgesen, J. K., Clancy-Menchetti, J., Santi, K., Nicholas, K., Robinson, C., et al. (2003). A comparison of teacher-directed versus peer-assisted instruction to struggling first-grade readers. *Elementary School Journal*, *103*(5), 461–479.

Merzenich, M., Jenkins, W., Johnston, P., Schreiner, C., Miller, S., & Tallal, P. (1996). Temporal processing deficits of language learning impaired children ameliorated by training. *Science*, *271*, 77–80.

Miller, H. (1997). Quantitative analyses of student outcome measures. *International Journal of Educational Research*, *25*, 119–136.

Morrow, L. M. (1992). The impact of a literature-based program on literacy achievement, use of literature, & attitudes of children from minority backgrounds. *Reading Research Quarterly*, *27*, 250–275.

Moss, M., Fountain, A. R., Boulay, B., Horst, M., Rodger, C., & Brown-Lyons, M. (2008). *Reading First implementation evaluation: Final report*. Cambridge, MA: Abt Associates.

Muñoz, M. A., & Dossett, D. (2004). Educating students placed at risk: Evaluating the impact of Success for All in urban settings. *Journal of Education for Students Placed at Risk*, *9(3)*, 261–277.

Murphy, R., Penuel, W., Means, B., Korbak, C., Whaley, A., & Allen, J. (2002). *E-DESK: A review of recent evidence on discrete educational software*. Menlo Park, CA: SRI International.

Mys, D. P., & Petrie, J. (1988). *Evaluation of student reading and math WICAT computer managed instructional program Salina Elementary School November, 1985–June 1988* (Bulletin No. 1345). Bearhorn, MI: Office of Research and Evaluation, Public Schools.

National Center for Education Statistics. (2007). *The National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education.

National Early Literacy Panel. (2008). *Developing early literacy*. Washington, DC: National Institute for Literacy.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Rockville, MD: National Institute of Child Health and Human Development.

Nelson, J. R., & Stage S. A. (2007). Fostering the development of vocabulary knowledge and reading comprehension through contextually based supplemental multiple meaning vocabulary instruction. *Education and Treatment of Children*, *30*(1), 1-22.

Nunnery, J., Slavin, R. E., Ross, S. M., Smith, L. J., Hunter, P., & Stubbs, J. (1996, April). *An assessment of Success for All program component configuration effects on the reading achievement of at-risk first grade students*. Paper presented at the annual meeting of the American Educational Research Association, New York.

O'Connor, R. (1999). Teachers learning Ladders to Literacy. *Learning Disabilities Research & Practice*, *14*(4), 203–214.

O'Connor, R., Notari-Syverson, A., & Vadasy, P. (1996, March). *The effect of kindergarten phonological intervention on the first grade reading and writing of children with mild disabilities*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Oglesby, F., & Suter, W. N. (1995). Matching reading styles and reading instruction. *Research in the Schools* (Mid-South Educational Research Association), *2*(1), 11–15.

Opuni, K. A. (2006). *The effectiveness of the Consistency Management & Cooperative Discipline (CMCD) model as a student empowerment and achievement enhancer: The experiences of two K-12 inner-city school systems*. Paper presented at the 4th Annual Hawaii International Conference of Education, Honolulu, Hawaii.

Papalwis, R. (2004). *Struggling middle school readers: Successful, accelerating intervention*. Sacramento: California State University.

Paterson, W., Henry, J., O'Quin, K., Ceprano, M., & Blue, E. (2003). Investigating the effectiveness of an integrated learning system on early emergent readers. *Reading Research Quarterly*, *38*(2), 172–206.

Phillips, L., Norris, S., Mason, J., & Kerr, B. (1990). *Effect of early literacy intervention on kindergarten achievement* (Technical Rep. No. 520). Champaign: University of Illinois at Urbana-Champaign, Center for the Study of Reading.

Policy Studies Associates. (2007). *Evidence of long-term learning outcomes among Reading Together tutees*. Washington, DC: Author.

Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching*. New York: Guilford.

Ragosta, M. (1983). Computer-assisted instruction and compensatory education: A longitudinal analysis. *Machine-Mediated Learning*, *1*(1), 97–127.

Rapp, J. C. (1991). *The effect of cooperative learning on selected student variables (Cooperative Integrated Reading and Composition) on academic achievement in reading comprehension, vocabulary and spelling and on student self-esteem*. EdD dissertation, Washington State University. Retrieved September 5, 2007, from ProQuest Digital Dissertations database. (Publication No. AAT 9207225)

Reid, E. (1996). *Exemplary Center for Reading Instruction (ECRI) validation study*. Salt Lake City, UT: Exemplary Center for Reading. (ERIC No. ED 414560)

Reis, S. M., Eckert, R. D., McCoach, D. B., Jacobs, J. K., & Coyne, M. (2008). Using enrichment reading practices to increase reading fluency, comprehension, and attitudes. *Journal of Educational Research*, *101*(5), 299–314.

Renzulli, J. S., & Reis, S. M. (1998). *The schoolwide enrichment model: A how-to guide for educational excellence* (2nd ed.). Mansfield Center, CT: Creative Learning Press.

Resendez, M., Sridiharan, S., & Azin, M. (2006). *Harcourt Achieve's Elements of Reading: Comprehension randomized control trial*. Jackson, WY: PRES Associates.

Rimm-Kaufman, S., Fan, X., Chiu, Y., & You, W. (2007). The contribution of the Responsive Classroom Approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology*, *45*, 401–421.

RMC Research Corporation (2003). *The Literacy Center K-1 Las Vegas research project*. Las Vegas, NV: Author.

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*, *94*(2), 240–257.

Rose, J. (2006). *Independent review of the teaching of early reading*. London: Department for Education and Skills.
Ross, S., Smith., L., Bond, C., & Casey, J. (1994). *Final report: 1993-1994 Success for All program in Montgomery, Alabama*. Memphis, TN: Center for Research in Education Policy, University of Memphis.
Ross, S. M., & Casey, J. (1998a). *Longitudinal study of student literacy achievement in different Title I school-wide programs in Ft. Wayne community schools, Year 2: First grade results*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.
Ross, S. M., & Casey, J. (1998b). *Success for All evaluation*, *1997–98 Tigard-Tualatin School District*. Memphis, TN: University of Memphis, Center for Research on Educational Policy.
Ross, S. M., Nunnery, J. A., & Smith, L. J. (1996). *Evaluation of Title I reading programs: Amphitheater Public Schools Year 1: 1995–1996*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.
Ross, S. M., Smith, L. J., & Casey, J. (1992). *Final report: 1991–92 Success for All program in Caldwell*, *Idaho*. Memphis, TN: Memphis State University.
Ross, S. M., Smith, L. J., & Casey, J. (1995). *Final Report: 1994–95 Success for All program in Fort Wayne*, *Indiana*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.
Ross, S. M., Smith, L., & Casey, J. (1997a). *Final report: 1996–97 Success for All program in Clarke County*, *Georgia*. Memphis, TN: University of Memphis, Center for Research on Educational Policy.
Ross, S. M., Smith, L., & Casey, J. (1997b). Preventing early school failure: Impacts of Success for all nonstandardized test outcomes, minority group performance, and school effectiveness. *Journal of Education for Students Placed at Risk*, *2*(1), 29–53.
Roth, S., & Beck, I. (1987). Theoretical and instructional implications of the assessment of two microcomputer word recognition programs. *Reading Research Quarterly*, *22*(2), 197–218.
Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention assessment*, *and adjustments*. Chichester, UK: John Wiley.
Saracho, O. (1982). The effects of a computer-assisted computer program on basic skills achievement and attitudes toward instruction of Spanish-speaking migrant children. *American Educational Research Journal*, *19*(2), 201–219.
Scarcelli, S., & Morgan, R. (1999). The efficacy of using a direct reading instruction approach in literature based classrooms. *Reading Improvement*, *36*(4), 172–179.
Schmidt, S. (1991). *Technology for the 21st century: The effects of an integrated distributive computer network system on student achievement*. Unpublished doctoral dissertation, University of La Verne, LaVerne, CA.
Schneider, W., Kuspert, P., Roth, E. Visé, M., & Marx, H. (1997). Short- and long-term effects of training phonological awareness in kindergarten: Evidence from two German studies. *Journal of Experimental Child Psychology*, *66*, 311–340.
Schultz, L. (1996). *Effectiveness study of Scholastic Phonics Readers and a comprehensive reading program*. New York: Scholastic.
Scientific Learning Corporation. (2006). Improved reading skills by students in Boone County School District who used Fast ForWord products. *Maps for Learning: Educator Reports*, *10*(15), 1–7.

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

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

Shapiro, L. R. & Solity, J. (2008). Delivering phonological and phonics training within whole class teaching. *The British Journal of Educational Psychology*, *78*(4), 597-620.

Sivin-Kachala, J., & Bialo, E. (2005). *Fluency Formula second grade study*. Long Island, NY: Scholastic.

Skeans, S. (1991). *The effects of Cooperative Integrated Reading and Composition, fidelity of implementation, and teacher concerns on student achievement*. PhD dissertation, Texas A&M University, College Station. Retrieved September 5, 2007, from ProQuest Digital Dissertations database. (Publication No. AAT 9217026)

Skindrud, K., & Gersten, R. (2006). An evaluation of two contrasting approaches for improving reading achievement in a large urban district. *The Elementary School Journal*, 106, 389.

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

Slavin, R. E. (2009). Cooperative learning. In G. McCulloch & D. Crook (Eds.), *International encyclopedia of education*. Abington, UK: Routledge.

Slavin, R. E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best evidence synthesis. *Reading Research Quarterly*, *43*(3), 290–322.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best evidence synthesis. *Review of Educational Research*, *78*(3), 427–515.

Slavin, R. E., Lake, C., Davis, S. & Madden, N. A. (2009, April). *Effective programs for struggling readers: A best-evidence synthesis*. Paper presented at the annual meetings of the American Educational Research Association, San Diego, CA.

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

Slavin, R. E., & Madden, N. A. (1991). *Success for All at Buckingham Elementary: Second year evaluation*. Baltimore: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Slavin, R. E., & Madden, N. (1998). *Success for All/Exito Para Todos: Effects on the reading achievement of students acquiring English* (Report No. 19). Baltimore: Center for Research on the Education of Students Placed at Risk.

Slavin, R. E., & Madden, N. A. (in press). Measures inherent to treatments in systematic reviews in education. *Journal of Research on Educational Effectiveness*.

Slavin, R. E., Madden, N. A., Chambers, B., & Haxby, B. (2009). *Two million children: Success for All*. Thousand Oaks, CA: Corwin.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1993). *Success for All in the Baltimore City Public Schools: Year 6 report*. Baltimore: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Slavin, R. E., & Smith, D. (in press). Effects of sample size on effect size in systematic reviews in education. *Educational Evaluation and Policy Analysis*.

Smith, L., & Ross, S. (1992). *1991–1992 Ft Wayne, IN SFA results*. Memphis, TN: Memphis State University, Center for Research in Educational Policy.

Smith, L. J., Ross, S. M., & Casey, J. P. (1994). *Special education analyses for Success for All in four cities*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Snow, C. E., Burns, S. M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Snow, M. F. (1993). *The effects of computer-assisted instruction and focused tutorial services on the achievement of marginal learners*. EdD dissertation, University of Miami, FL. Retrieved September 5, 2007, from ProQuest Digital Dissertations database. (Publication No. AAT 9401831)

Social Programs That Work. (2008). *Success for All*. Retrieved December 12, 2008, from www.evidencebasedprograms.org

Sporer, N., Brunstein, J., & Kieschke, U. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and Instruction, 19*, 272-286.

Stambaugh, T. (2007). *Effects of the Jacob's Ladder reading comprehension program on reading comprehension and critical thinking skills of third, fourth, and fifth grade students in rural, Title I schools*. PhD dissertation, College of William and Mary, Williamsburg, VA. Retrieved November 21, 2007, from ProQuest Digital Dissertations database. (Publication No. AAT 3254406)

Standish, D. (1995). The effects on reading comprehension of Jostens' Integrated Language Arts for second-grade students along with Jostens' Basic Learning System for second-grade Chapter 1 students. *Dissertation Abstracts International*, *57*(3), 1079A. (UMI No. 9623238)

Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. (1976). *Education as experimentation: A planned variation model* (Vols. IIIA, IIIB). Cambridge, MA: Abt Associates. (ERIC No. ED 148489)

Stein, M., Berends, M., Fuchs, D., McMaster, K., Saenz, L., Yen, L., et al. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis*, *30*(4), 368–388.

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*, 433–454.

Stevens, R. J., & Slavin, R. E. (1995a). The cooperative elementary school: Effect on student achievement and social relations. *American Educational Research Journal*, *32*, 321–351.

Stevens, R. J., & Slavin, R. E. (1995b). Effects of a cooperative approach in reading and writing on academically handicapped and nonhandicapped students. *Elementary School Journal*, *95*(3), 241–262.

Stevens, R. J., Van Meter, P., Garner, J., Warcholak, N., Bochna, C., & Hall, T. (2008). The Reading and Integrated Literacy Strategies (RAILS): An integrated approach to early reading. *Journal of Education for Students Placed at Risk*, *13*(4), 357–380.

Stevenson, Z., Cathey-Pugh, J., & Kosmidis, M. (1988). Achievement in the Writing to Read program: A comparative evaluation study. Washington, DC: District of Columbia Public Schools, Division of Quality Assurance and Management Planning. (ERIC No. ED293147)

Swartz, J., & Johnston, K. (2003). *Efficacy study of Houghton Mifflin Reading: A legacy of literacy*. Cambridge, MA: Abt Associates.

Tallal, P., Miller, S., Bedi, G., Byma, G., Wang, X., Nagarajan, S., et al. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, *271*, 81–84.

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

Torgerson, C. (2006). The quality of systematic reviews of effectiveness in literacy learning in English: A tertiary review. *Journal of Research in Reading*, *29*(2), 1–29.

Torgerson, C. J., Brooks, G., & Hall, J. (2006). *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling* (DfES Research Rep. 711). Department for Education and Skills, London.

Tracey, D., & Young, J. (2006). *Technology and early literacy: The impact of an integrated learning system on high-risk kindergartners' achievement*. Union, NJ: Kean University.

Van Keer, H., & Verhaeghe, J. (2005). Comparing two teacher development programs for innovating reading comprehension instruction with regard to teachers' experiences and student outcomes. *Teaching and Teacher Education*, *21*, 543–562.

Van Keer, H., & Verhaeghe, J. (2008). *Strategic reading in peer tutoring dyads in second- and fifth-grade classrooms*. Unpublished manuscript, Ghent University, Ghent, Belgium.

Vaughan, J., Serido, J., & Wilhelm, M. (2006). *The effects of My Reading Coach on reading achievement of elementary education students*. Phoenix: Arizona Board of Regents.

Wang, L. W., & Ross, S. M. (1999a). *Evaluation of Success for All program*. *Little Rock School District*, *Year 2: 1998–99*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Wang, L. W., & Ross, S. M. (1999b). *Results for Success for All Program*, *Alhambra (AZ) School District*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Wasik, B., & Slavin, R. (1993). Success for All at Pepperhill Elementary School: 1993 evaluation. Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students. Johns Hopkins University.

Webb, N. M. (2008). Learning in small groups. In T. L. Good (Ed.), *21st century education: A reference handbook* (pp. 203–211). Thousand Oaks, CA: Sage.

What Works Clearinghouse. (2009). *Beginning reading topic report*. Washington, DC: U.S. Department of Education. Retrieved March 15, 2009, from http://ies.ed.gov/ncee/wwc

Whitaker, J. C. (2005). *Impact of an integrated learning system on reading and mathematics achievement*. Unpublished doctoral dissertation, Tennessee State University, Johnson City.

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

Wilkerson, S. B. (2004). *A study of the effectiveness of Harcourt Achieve's Rigby Literacy Program: Final evaluation report*. Aurora, CO: Mid-Continent Research for Education and Learning.

Wilkerson, S. B., Shannon, L. C., & Herman, T. L. (2006). *An efficacy study on Scott Foresman's Reading Street Program: Year one report*. Austin, TX: Magnolia Consulting.

Wilkerson, S. B., Shannon, L. C., & Herman, T. L. (2007). *An efficacy study on Scott Foresman's Reading Street Program: Year two report*. Austin, TX: Magnolia Consulting.

Williams, B. J. (2005). *A quasi-experimental study on the effects of the OpenBook to Literacy program on fourth-grade students*. Unpublished doctoral dissertation, Tennessee State University, Johnson City.

Yee, V. N. (2007). *An evaluation of the impact of a standards-based intervention on the academic achievement of English language learners*. Unpublished doctoral dissertation, University of Southern California, Los Angeles

Bios

ROBERT E. SLAVIN is director of the Center for Research and Reform in Education at Johns Hopkins University, director of the Institute for Effective Education at the University of York, and the cofounder and chairman of the Success for All Foundation. He received his BA in psychology from Reed College in 1972 and his PhD in social relations in 1975 from Johns Hopkins University. He has authored or coauthored more than 200 articles and 20 books, including *Educational Psychology: Theory Into Practice* (Allyn & Bacon, 1986, 1988, 1991, 1994, 1997, 2000, 2003, 2006, 2009); *Cooperative Learning: Theory, Research, and Practice* (Allyn & Bacon, 1990, 1995); *Show Me the Evidence: Proven and Promising Programs for America's Schools* (Corwin, 1998); *Effective Programs for Latino Students* (Erlbaum, 2000); and *Two Million Children: Success for All* (Corwin, 2009). He received the American Educational Research Association's Raymond B. Cattell Early Career Award for Programmatic Research in 1986, the Palmer O. Johnson award for the best article in an AERA journal in 1988 and 2008, the Charles A. Dana award in 1994, the James Bryant Conant Award from the Education Commission of the States in 1998, the Outstanding Leadership in Education Award from the Horace Mann League in 1999, the Distinguished Services Award from the Council of Chief State School Officers in 2000, and the AERA Review of Research Award in 2009.

CYNTHIA LAKE is a research scientist for the Center for Research and Reform in Education at Johns Hopkins University. She received her BA in development studies from Brown University in 1996 and her MA in sociology in 2004. Her research interests include comprehensive school reform, school improvement, and educational research methods and statistics. She is currently a data analyst of meta-analyses of education for the *Best Evidence Encyclopedia*.

BETTE CHAMBERS is currently a professor in the Institute for Effective Education at the University of York in England and at the Center for Research and Reform in Education at Johns Hopkins University, where she conducts research in early childhood education and early literacy. She also directs the development and dissemination of the early childhood education and technology-embedded programs at the Success for All Foundation in Baltimore. She received her BA in early childhood education from Concordia University in 1982 and her PhD in educational psychology in 1990 from McGill University. She has authored and coauthored numerous articles, books, and practical guides on cooperative learning, technology infusion in literacy, and early childhood education.

ALAN CHEUNG is an associate professor at Johns Hopkins University. He received his PhD in education in 2002 from Brigham Young University. He has authored and coauthored more than 30 professional publications, including book chapters, articles, technical reports, and creative works. He specializes in research reviews, childhood education, and curriculum reform.

SUSAN DAVIS is a research assistant at the Success for All Foundation.