



## Review

## Effective programs for struggling readers: A best-evidence synthesis

Robert E. Slavin<sup>a,b,\*</sup>, Cynthia Lake<sup>a</sup>, Susan Davis<sup>c</sup>, Nancy A. Madden<sup>a,b</sup><sup>a</sup> Johns Hopkins University, 200 W. Towsontown Blvd., Baltimore, MD 21204, USA<sup>b</sup> University of York, Institute for Effective Education, Berrick Saul Building, Heslington, York YO10 5DD, England, United Kingdom<sup>c</sup> Success for All Foundation, 200 W. Towsontown Blvd., Baltimore, MD 21204, USA

## ARTICLE INFO

## Article history:

Received 4 June 2010

Received in revised form 7 July 2010

Accepted 8 July 2010

## Keywords:

Phrases: Struggling readers  
Elementary reading programs  
Student achievement  
Literacy  
Reading research

## ABSTRACT

This article reviews research on the achievement outcomes of alternative approaches for struggling readers ages 5–10 (US grades K-5): One-to-one tutoring, small-group tutorials, classroom instructional process approaches, and computer-assisted instruction. Study inclusion criteria included use of randomized or well-matched control groups, study duration of at least 12 weeks, and use of valid measures independent of treatments. A total of 97 studies met these criteria. The review concludes that one-to-one tutoring is very effective in improving reading performance. Tutoring models that focus on phonics obtain much better outcomes than others. Teachers are more effective than paraprofessionals and volunteers as tutors. Small-group, phonetic tutorials can be effective, but are not as effective as one-to-one phonetically focused tutoring. Classroom instructional process programs, especially cooperative learning, can have very positive effects for struggling readers. Computer-assisted instruction had few effects on reading. Taken together, the findings support a strong focus on improving classroom instruction and then providing one-to-one, phonetic tutoring to students who continue to experience difficulties.

© 2010 Elsevier Ltd. All rights reserved.

## Contents

1. Introduction .....	2
2. Focus of the current review .....	3
3. Methodological issues characteristic of research on struggling readers .....	4
4. Review methods .....	4
4.1. Literature search procedures .....	4
4.2. Effect sizes .....	5
4.3. Criteria for inclusion .....	5
4.4. Categories of research design .....	6
5. Findings .....	6
5.1. One-to-one tutoring by teachers .....	6
5.2. <i>Reading Recovery</i> .....	6
5.3. Other one-to-one tutoring by teachers .....	9
5.4. Conclusions: one-to-one tutoring by teachers .....	9
6. Tutoring by paraprofessionals and volunteers .....	9
7. Small-group tutorials .....	12
8. Classroom instructional process programs .....	12

\* Corresponding author at: Johns Hopkins University, 200 W. Towsontown Blvd., Baltimore, MD 21204, USA. Tel.: +1 410 616 2310/+44 1904 328 155; fax: +1 410 324 4440/+44 1904 328 156.

E-mail addresses: [rslavin@jhu.edu](mailto:rslavin@jhu.edu) (R.E. Slavin), [clake5@jhu.edu](mailto:clake5@jhu.edu) (C. Lake), [sdavis@successforall.org](mailto:sdavis@successforall.org) (S. Davis), [nmadden@jhu.edu](mailto:nmadden@jhu.edu) (N.A. Madden).

9.	Classroom instructional process with tutoring ( <i>Success for All</i> ).....	12
10.	Instructional technology .....	19
11.	Findings on key issues.....	19
11.1.	The importance of phonics .....	19
11.2.	Long-term effects of first-grade tutoring.....	19
11.3.	Teachers vs. paraprofessionals and volunteers as tutors.....	21
11.4.	One-to-one vs. small-group tutorials.....	21
11.5.	Classroom instructional process approaches vs. tutoring.....	21
11.6.	Implications for response to intervention.....	22
12.	Discussion .....	22
	For more information .....	23
	Disclosure statement.....	23
	Acknowledgements .....	23
	References .....	23

## 1. Introduction

Over the past 25 years, there have been extraordinary developments in research, policy, and practice relating to programs for elementary-aged children who are struggling to learn to read. While there has long been concern about reading disabilities, dyslexia, and underachievement, research and development since the 1980s has created a sense of optimism that most children who start off their time in school struggling to learn to read can be quickly brought into the mainstream in this crucial skill. The appearance of *Reading Recovery*, first in New Zealand and later in the US, the UK, and throughout the English-speaking world, gave particular hope that tutors with extensive training could prevent reading failure with a substantial proportion of the children who were failing in first grade and were therefore at risk of serious difficulties throughout their time in school. In the US in the 1990s, the Clinton administration's America Reads initiative encouraged the creation of programs for volunteer tutors to work with struggling children, and this led to widespread development and evaluation of replicable programs for this purpose. Reading First, the Bush Administration's initiative for children in grades K-3, focused on high-poverty, low-achieving schools, with a particular focus on small-group interventions for struggling readers. A new focus in special education on response to intervention, particularly in the US (see Allington & Walmsey, 2007; Fuchs & Fuchs, 2006; Gersten et al., 2009), in which at-risk children receive small-group interventions and then possibly one-to-one tutoring to attempt to solve their problems before they might be referred to special education, has also encouraged development and research on small-group tutorials. In the UK, the government introduced a program called Every Child a Reader to disseminate *Reading Recovery* throughout England (see Burroughs-Lange, 2007, 2008; Policy Exchange, 2009). In addition to benefitting children, each of these initiatives has stimulated research of all kinds. In 1993, Wasik and Slavin reviewed research on tutoring programs, and found only five studies. In 2000, Elbaum, Vaughn, Hughes, and Moody reviewed one-to-one tutoring programs, but the great majority of the available research focused on just one program, *Reading Recovery*. Today, there are many programs designed to help struggling readers, and there is much research on factors that affect the impact of tutoring and other interventions. There is also much research on the effects on struggling readers of classroom programs and comprehensive school reform models that impact entire schools.

The importance of getting children off to a good start in reading cannot be overstated. In the elementary grades, success in school is virtually synonymous with success in reading, and children without strong reading skills by middle school are headed for disaster. Children who fail to read in the early grades incur so many costs to the education system, in special education, remediation, grade repetition, delinquency, and ultimate dropout, that even very expensive interventions can be justified on cost-effectiveness grounds alone, while at the same time preventing damage to young peoples' lives. Further, reading failure is not distributed randomly, but is concentrated among schools serving many disadvantaged, minority, and limited English proficient children. It is in the early elementary grades where the gap in performance between children of different races first appears, and this gap is perhaps the most important policy issue in education in the US. On the US fourth grade National Assessment of Educational Progress (NAEP, 2007), 43% of White children achieved at the "proficient" level, but only 14% of African American, 17% of Hispanic, and 8% of American Indian children scored at this level. In the UK and other European countries, gaps between middle class and disadvantaged students are equally problematic, and are particularly pronounced for boys and for members of various ethnic subgroups (Brooks, 2007; Harrison, 2000; Policy Exchange, 2009). Effective reading programs are important for children of all backgrounds, but for disadvantaged and many minority children and for children with learning disabilities, who particularly depend on school to achieve success, effective reading programs are especially important.

From the perspective of theory and basic research, the problem of early reading failure is reasonably well understood. A meta-analysis of basic research on the topic carried out by the US National Reading Panel (2000) identified five factors that must be emphasized if children are to succeed in reading: phonemic awareness, phonics, comprehension, fluency, and vocabulary. Another US consensus report, *Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffin, 1998), came to similar conclusions. Ysseldyke and Taylor (2007) identified problems that lead to reading difficulties, including processing deficits, deficient neuropsychological functioning, lack of sufficient experience or support at home, lack of facility in English, and insufficient preschool experience (see also Kamil, Mosenthal, Pearson, & Barr, 2000; Harrison, 2000). However, while

much is known about the characteristics of struggling readers, there is much less known about practical, replicable programs capable of helping educators prevent or remediate early reading deficits. Research in recent years has begun to evaluate a broad range of potential solutions but this newer research has not been systematically reviewed.

Because of the importance of ensuring success in reading for all children, the policy focus on the problem, and the costs involved, it is especially important to understand which types of programs are most likely to have a strong and lasting impact on the reading success of struggling children. The purpose of the present article is to review research on all types of approaches that have been evaluated as solutions for struggling readers. The article primarily addresses the US and UK policy contexts, but the review included all research available in English, and the implications of the review apply to all countries in the developed world.

As noted earlier, research on programs for struggling readers has been reviewed previously with conflicting conclusions. Wasik and Slavin (1993) reviewed research on one-to-one tutoring, concluding that tutoring had substantial positive impacts and that certified teachers obtained better results than paraprofessionals. Wasik (1997) then found positive outcomes for a variety of volunteer tutoring programs, such as those used in the federal *America Reads* initiative. Elbaum et al. (2000), in a review of one-to-one tutoring, focused primarily on *Reading Recovery*, finding positive effects but concluding (based on two small dissertations) that small-group tutorials might be as effective as one-to-one tutoring. Shanahan and Barr (1995) reviewed research on *Reading Recovery* and while recognizing its effectiveness, questioned its cost-effectiveness. In a recent What Works Clearinghouse practice guide, Gersten et al. (2009), in providing a research case for response to intervention, concluded that there is strong evidence to support use of small-group tutorials with struggling readers, based on 11 studies that fully or partially met the standards of the What Works Clearinghouse (WWC). The What Works Clearinghouse (2009) Beginning Reading Topic Report gave top ratings to *Reading Recovery*, though not to any of the programs cited by Gersten et al. (2009) as evidence for the effectiveness of small-group instruction. Brooks (2007) reviewed research on early interventions for reading difficulties in the UK policy context.

Few researchers or educators would doubt that one-to-one tutoring is effective for struggling readers. The questions today are beyond this and are critical for providing useful guidance to educators. After 25 years of research and reform, we want to know about long-term impacts of early intervention, about different effects of tutoring by certified teachers as opposed to paraprofessionals and volunteers, and whether small-group interventions can be as effective as one-to-one tutoring. We want to know the importance of a focus on phonics and phonological awareness in reading interventions. We want to know how to build on success in the early grades to maintain gains into secondary schools, and we want to know how improved classroom programs and technology might contribute to the success of struggling readers. Now that there is a rich diversity of approaches that have been rigorously evaluated, it is no longer enough to review tutoring or small-group interventions in isolation. Educators and policy makers deserve comparative, fair, and readily understood information about the likely impacts of alternative approaches or combinations of approaches. We also want to know more about for whom various models are likely to work and under what conditions. These are the questions that drive the current review.

The present article reviews research on the achievement outcomes of practical approaches for struggling readers, applying consistent methodological standards to the research. It is intended to provide fair summaries of the achievement effects of the full range of remedial and preventive 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 struggling readers: Tutoring, small-group tutorials, teacher professional development for effective core reading instruction, and instructional technology. 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 (see Slavin, 2008). It is part of a series, all of which used the same methods with minor adaptations. Separate research syntheses review research on beginning and upper-elementary reading programs (Slavin, Lake, Chambers, Cheung, & Davis, 2009), middle and high school reading programs (Slavin, Cheung, Groff, & Lake, 2008), and reading programs for English language learners (Cheung & Slavin, 2005).

The synthesis of research on beginning reading programs (Slavin, Lake, Chambers, Cheung, & Davis, 2009) provides the closest background for the present review. The beginning reading review identified 63 studies that met the inclusion standards. These were divided into four categories: reading curricula (core and supplementary textbooks), instructional technology, instructional process programs (such as cooperative learning), and combinations of curricula and instructional process. Effect sizes for curricula ( $ES = +0.12$ ) and for instructional technology ( $ES = +0.11$ ) were low. Larger effect sizes ( $ES = +0.31$ ) were found for instructional process programs, especially cooperative learning programs in which students help one another master reading comprehension skills in small teams or pairs. Effect sizes for *Success for All*, which combines curriculum and instructional process approaches, averaged  $+0.29$ .

## 2. Focus of the current review

The present review uses procedures similar to those used in the beginning reading review to examine research on programs for struggling readers. The purpose of the review is to place all types of interventions intended to enhance the achievement of students experiencing difficulties in learning to read on a common scale, to provide educators and policy makers with meaningful, unbiased information that they can use to select approaches 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 using 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 interventions for struggling readers' instruction, and groups them in six categories as follows:

1. *One-to-one tutoring by teachers*. This includes *Reading Recovery* and other tutoring models.
2. *One-to-one tutoring by paraprofessionals and volunteers*.
3. *Small-group tutorials*.
4. *Classroom instructional process approaches*. This category includes cooperative learning (e.g., *Cooperative Integrated Reading and Composition (CIRC)* and *Peer-Assisted Learning Strategies (PALS)*), *Direct Instruction*, and other approaches that focus on training teachers in effective classroom methods rather than pullout services for at-risk students. Although these methods target all students, the studies report outcomes separately for students in the lowest 25% of their cohort at pretest.
5. *Classroom instructional process with tutoring*. This category consists of studies of *Success for All*, which trains teachers in reading strategies, cooperative learning, and other methods throughout the elementary grades, and provides 1-1 tutoring by teachers to low achieving children. Outcomes are reported for students in the lowest 25% of their cohort at pretest.
6. *Instructional technology*. These studies report effects of computer-assisted instruction and other uses of technology for children in the lowest 25% of their cohort at pretest.

### 3. Methodological issues characteristic of research on struggling readers

While a review of research on reading programs for struggling readers shares methodological issues common to all systematic reviews, there are also some key issues unique to this topic. One 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. The present review emphasizes measures of reading comprehension and general reading, but also accepts measures of key components of reading such as letter–word identification and word attack as measures of reading. It did not, however, accept measures such as auditory phonemic awareness. Measures of oral vocabulary, spelling, and language arts were also excluded.

Another problem of early reading measurement is that in kindergarten studies, it is possible 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 US 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 them, but would in fact simply be teaching skills the children would probably have mastered somewhat later. For this reason, kindergarten interventions are included in this review only if they followed students to the end of first grade or later, by which time it is certain that control students as well as experimental students would have been formally taught to read.

### 4. Review methods

The review methods used here are similar to those used by Slavin, Lake, et al. (2009), who adapted 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 *What Works Clearinghouse* (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 this series of best-evidence reviews.

#### 4.1. 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 different combinations of key words (for example, “elementary or primary students,” “reading,” “achievement”) and the years 1970–2009. 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' websites 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 *What Works Clearinghouse* (2009) beginning reading topic report, Gersten et al. (2009), Ritter, Barnett, Denny, and Albin (2009), Elbaum et al. (2000), Wasik and Slavin (1993), Wasik (1997), National Reading Panel (2000), Snow et al. (1998), Torgerson, Brooks, and Hall (2006), and Brooks (2007). We also conducted searches of recent tables of contents of key journals. We searched the following tables of contents from 2000 to 2008: *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. Articles from additional journals, as well as technical reports, conference papers, dissertations, and other sources, were also obtained. Citations of studies appearing in the articles found in the first wave were also followed up.

#### 4.2. Effect sizes

In general, effect sizes were computed as the difference between experimental and control individual student posttest means 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 were usually reported for each year but only the final year of treatment is presented in the tables. However, if there were multiple cohorts (e.g., K-1, K-2, K-3), each with adequate pretests, all cohorts were 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, 2008; Slavin & Smith, 2009).

Statistical significance is reported in this review as it was in the original articles, but is not reported if the article did not report significance for a given comparison. For example, if a study reported outcomes separately for high, average, and low achievers and reported treatment by achievement level interactions, but did not report a separate analysis for low achievers, we would report the effect size for low achievers but not statistical significance.

#### 4.3. Criteria for inclusion

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

1. The studies evaluated specific, potentially replicable programs for children ages 5–10 (US grades K-5) who were having difficulties learning to read. These are defined as children with reading disabilities, children in the lowest 33% (or lower) of their classes, or any children receiving tutoring or other intensive services to prevent or remediate serious reading problems. Studies of variables (e.g., mainstreaming or after school attendance) or of government funding streams (e.g., Title I, Reading First, Supplemental Educational Services) are not included.
2. The included studies compared children taught using a given reading program to those in a control group taught using an alternative program or standard methods. Studies that compared two experimental methods without a control group representing ordinary practice are not included in the main tables but are discussed as appropriate to answer relevant questions.
3. Studies could have taken place in any country, but the report had to be available in English. In practice, most qualifying studies took place in the US, and others took place in the UK and Australia.
4. 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.
5. Pretest data had to be provided. 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).
6. 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 What Works Clearinghouse. As noted earlier, measures of pre-reading skills such as auditory phonological awareness, as well as reading-related outcomes such as oral vocabulary, language arts, and spelling, were not included in this review.
7. 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 What Works Clearinghouse (2009) beginning reading topic report, which included very brief studies.
8. Studies had to have at least 15 students and two teachers or tutors in each treatment group.

#### 4.4. Categories of research design

Three 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. Studies using fully randomized designs (R) are preferable to randomized quasi-experiments (RQE), but all randomized experiments are less subject to bias than matched studies. In the text and in tables, studies of each type of program are listed in this order (R, RQE, M). 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.

### 5. Findings

#### 5.1. One-to-one tutoring by teachers

One-to-one instruction from certified teachers and reading specialists is the gold standard among interventions for struggling readers. It is the most expensive solution, but the expense is more than justified if it can make a substantial difference for children at a critical point in their reading development and therefore reduce later needs for special education, remediation, or grade retention. Studies of one-to-one tutoring by teachers are summarized in Table 1, and then described in the following sections.

#### 5.2. Reading Recovery

*Reading Recovery* is by far the most widely researched and widely used tutoring program in the world. Originally developed in New Zealand by Clay (1985), *Reading Recovery* provides extensive training, observation, and feedback to certified teachers, who provide daily 30-min lessons to the lowest 20–30% of first graders in a school until they are reading at the level of average first graders in their school. In general, teachers work with about four children during half of each school day and teach a regular first-grade class the other half of the day. A *Reading Recovery* session involves: (a) re-reading of a familiar book, (b) independent reading of a text at the child's level, (c) teaching of letter knowledge, (d) composing and writing a sentence, (e) re-constructing a cut-up sentence, and (f) introducing a new book. The books are leveled readers with predictable text. Over the years, *Reading Recovery* has added more of an emphasis on phonics and decoding skills. Teacher training for *Reading Recovery* involves about 75 contact hours and includes live observations through a one-way glass screen and feedback from expert teacher leaders. The training takes place over an entire school year concurrent with practice with children.

From early on, *Reading Recovery* leaders have placed a strong emphasis on program evaluation, and there are many studies that have evaluated program outcomes. However, there are characteristic elements of many *Reading Recovery* studies that are important to understand. First, most *Reading Recovery* studies use as posttests measures from Clay's (1985) Diagnostic Observation Survey. Given particular emphasis is a measure called Text Reading Level, in which children are asked to read aloud from leveled readers, while testers (usually other *Reading Recovery* teachers) record accuracy using a running record. Unfortunately, this and other Diagnostic Observation Survey measures are closely aligned to skills taught in *Reading Recovery* and are considered inherent to the treatment; empirically, effect sizes on these measures are typically much greater than those on treatment-independent measures. For example, the review by Elbaum et al. (2000) reported a mean effect size of +0.64 for Text Reading Level and +0.30 for other oral reading measures. In accord with this review's requirement of treatment-independent measures, Diagnostic Observation Survey measures in studies of *Reading Recovery* are not reported, but this issue is discussed further below.

A second typical procedure in *Reading Recovery* evaluations is to divide children into "discontinued" and "not discontinued" categories. Discontinued children are those who met criteria for ending tutoring because they were reading at the level of their schoolmates. "Not discontinued" children are those who did not meet criteria, and are in general referred for special education or other services. A third category is children who received some tutoring but not the 60 sessions held to be a full treatment. Some studies of *Reading Recovery* only include the discontinued (i.e., successful) children. In this review, studies are included only if all children assigned to *Reading Recovery*, discontinued or not, are included in the analyses.

Finally, some *Reading Recovery* studies have the tutors themselves administer the tests used as the outcome measures in the evaluation. An example is a widely cited study by Schwartz (2005). This introduces a strong possibility of bias, and such studies were excluded.

The outcomes for *Reading Recovery* were positive, but less so than might have been expected. Across all 8 studies, the weighted mean effect size was +0.23. This mean did not include Text Reading Level or other measures from the Clay Diagnostic Survey, which were considered inherent to the treatment and were therefore excluded, as discussed previously. Text Reading Level, often administered by *Reading Recovery* teachers and aligned with the focus of *Reading Recovery* teaching methods, produces much higher effect sizes than other reading measures. Across six qualifying studies that used it, the mean effect size was +1.32 for Text Reading Level, but about half of this, or +0.68, for other measures. Even including Text Reading Level,

**Table 1**

(a) One-to-one and (b) other one-to-one tutoring by teachers.

Study	Design large/small	Duration	N	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size
<b>(a)</b>								
<i>Reading Recovery</i> Pinnell et al. (1994)	Randomized (L)	1 year	33 schools 193 students (31E, 162C)	1	Urban, suburban, and rural schools in Ohio; 74%W, 26% AA, 65% FL	Gates MacGinitie		+0.19
Pinnell et al. (1988)	Randomized (S)	1 year	12 schools 187 students (126E, 26C)	1	Low-achieving students in Columbus, Ohio	CTBS	Comprehension Vocabulary	+0.55 +0.48
Curry, Griffith, and Williams (1995)	Matched (L)	1 year	553 students (268E, 285C)	1	High-poverty students in Austin, TX. 47%AA, 47%H, 6%W	ITBS		–0.16
Burroughs-Lange (2008), Hurry and Holliman (2009)	Matched (S)	1 year	42 schools (21E, 21C) 234 students (87E, 147C)	Ages 5–7	Low-achieving students in London, England with high proportions of ELLs and children receiving FL	BAS Word Reading WRAPS		+0.87 +0.65
Hurry and Sylva (2007)	Matched (S)	1 year	198 students (89E, 109C)	Ages 6–7	Primary schools in England 42% FL, 16% ELLs	BAS Word Reading Neale Prose Reading		+0.84 +0.85
Center et al. (1995)	Matched (S)	1 year	15 schools (10E, 5C) 56 students (22E, 34C)	Year 1	Urban schools in New South Wales, Australia	Neale Analysis of Reading Ability Passage Reading Cloze Test of Syntactic Awareness Word Attack Skills Test		+1.15 +1.00 +0.46 +0.82
Escamilla (1997)	Matched (S)	1 year	6 schools 46 students (23E, 23C)	1	Spanish-dominant students in Southern Arizona	Aprenda		+0.30
Huggins (1999)	Matched (S)	1 year	30 schools 122 students (70E, 52C)	1	High-poverty students in Detroit, MI	CAT	Comprehension Vocabulary	+0.03 –0.15
<b>(b)</b>								
<i>Auditory Discrimination in Depth</i> Torgesen et al. (1997)	Randomized (S)	2 1/2 years	65 students (33E, 32C)	K-2	50%W, 49%AA	Woodcock Johnson	Word ID Word Attack Passage Comprehension Word Efficiency Non-Word Efficiency	+0.65 +1.02 +0.39 +1.28 +1.17
<i>Early Steps/Howard Street Tutoring</i> Morris et al. (2000)	Matched (S)	1 year	11 schools (6E, 5C), 86 students (43E, 43C)	1	High-poverty AA schools in Tennessee	Woodcock Johnson adaptations	Word Attack Passage Comprehension	+0.92 +0.80
Brown et al. (2005) (Teacher tutors only)	Matched (S)	1 year	7 schools, 59 students (17E, 42C)	2–6	Urban schools in the intermountain west 46% FL, 42% minority, 25% ELLs	Woodcock Johnson	Word Attack Passage Comprehension Word recognition Passage reading	+1.00 +1.31 +0.81 +1.00

Table 1 (Continued)

Study	Design large/small	Duration	N	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size	
Santa and Hóien (1999)	Matched (S)	1 year	4 schools (2E, 2C) 49 students (23E, 26C)	1	Lower middle class schools in Montana	Woodcock Johnson	Word ID Word Attack Passage Comprehension	+0.70 +1.28 +1.14	+1.04
<i>Reading Rescue</i> Ehri et al. (2007) (Teacher tutors only)	Matched (S)	6 months	102 students (32E, 70C)	1	Spanish-dominant students in a large city 95%FL	Gates MacGinitie			+1.08
<i>Reading with Phonology</i> Hatcher et al. (1994)	Matched (S)	7 months	63 students: (32 E, 31C)	Ages 6–7	Schools in rural Northern England	BAS Word Reading Neale Analysis of Reading Ability	Accuracy Comprehension	+0.40 +0.62 +0.94	+0.65
<i>Intensive Reading Remediation</i> Blachman et al. (2004)	Randomized (S)	1 year	69 students (37E, 32C)	2, 3	School in 4 districts in upstate NY	Woodcock	Word ID Word Attack GORT Oral Reading Quotient Word Reading Word Reading Efficiency	+0.99 +0.96 +0.75 +0.77 +0.80	+0.85
<i>Targeted Reading Intervention</i> Vernon-Feagans et al. (2010)	Randomized quasi-experiment (S)	1 year	15 schools 310 students (194E, 116C)	K-1	Rural Title I schools; 50%W, 24%AA, 26% Other	Woodcock	Word Attack Letter–Word ID Passage Comprehension	+0.30 +0.45 +0.59	+0.45
Vernon-Feagans et al. (2009), Study 1	Randomized quasi-experiment (S)	1 year	6 schools 125 students (59E, 66C)	K-1	Rural impoverished counties in the southeast; 80%FL, 49% AA, 33%W, 10% AI	Woodcock Johnson	Letter–Word ID Word Attack	+0.24 +0.25	+0.25
Vernon-Feagans et al. (2009), Study 2	Randomized quasi-experiment (S)	1 year	4 schools 43 classrooms (26E, 17C) 151 students (97E, 54C)	K-1	Rural Texas and New Mexico; 37%W, 26% AA, 35% Other	Woodcock Johnson	Word Attack Letter–Word ID Passage Comprehension	+0.27 +0.24 +0.50	+0.34
<i>TEACH</i> Mantzicopoulos, Morrison, Stone, and Setrakian (1992)	Randomized (S)	2 years	116 students (59E, 57C)	1–2	Middle-class children in suburban San Francisco in lowest third of their class	SDRT-Comprehension SDRT-Phonetic Analysis Woodcock Word Attack K-ABC Reading/Decoding WRAT-Reading	+0.10 +0.09 +0.30 +0.29	+0.19	
Arnold (1977)	Matched (S)	7 months	3 schools 63 students (23E, 40C)	1	2 inner city and 1 middle class school				+0.34

Note: L=large study with at least 250 students; S=small study with less than 250 students; E=experimental; C=control; FL=free/reduced-price lunch; W=White; AA=African American; H=Hispanic; AI=American Indian; ELL=English language learner; RR=*Reading Recovery*; CTBS=Comprehensive Test of Basic Skills; ITBS=Iowa Test of Basic Skills; CAT=California Achievement Test; ADD=Auditory Discrimination in Depth; BAS=British Ability Scales; WRAPS=Word Reading and Phonics Scale; GORT=Gray Oral Reading Test; SDRT=Stanford Diagnostic Reading Test; K-ABC=Kaufmann Assessment Battery for Children; WRAT=Wide Range Achievement Test.



however, would only increase the weighted mean for *Reading Recovery* studies to +0.35, particularly because of two large studies that did not use Text Reading Level and found program effects near zero.

The two randomized studies found conflicting outcomes. An Ohio statewide study (Pinnell, Lyons, DeFord, Bryk, & Selzer, 1994) reported effect sizes on Gates MacGinitie of +0.19 for schools in which *Reading Recovery* teachers were experienced with the program, in comparison to students who did not receive tutoring or any other intervention. The effect size for a very similar intervention which gave new tutors extensive but accelerated training (compared to no tutoring) was –0.14 on the Gates. In contrast, an earlier randomized study by Pinnell, DeFord, and Lyons (1988) found an effect size of +0.52 on CTBS.

*Reading Recovery* effects were particularly weak in evaluations in large urban districts in the US. In large studies in Austin, Texas (Williams et al., 1995) and Detroit (Huggins, 1999), control groups scored slightly better than did *Reading Recovery*-tutored students. In a small California study, Acalin (1995) found that students with learning disabilities who received *Reading Recovery* performed no better than similar students who experienced *Project Read*, a phonetic small-group intervention. In contrast, all three of the non-US studies found strong positive effects. Burroughs-Lange (2008) found an effect size of +0.76 in London schools, Hurry and Sylva (2007) found an effect size of +0.85 in English primary schools, and Center, Wheldall, Freeman, and Outhred (1995) reported an effect size of +0.86 in urban Australian schools. The difference between the American studies and the UK and Australian studies may relate to the nature of the control groups, which were typically receiving small-group, phonetic instruction in the US but were less likely before the late 1990s to receive phonetic small-group assistance in the UK or Australia. Also, it is important to note that the measures in the Austin and Detroit studies were group-administered ITBS and CAT tests, which are probably much less sensitive than the individually administered tests used in other studies.

### 5.3. Other one-to-one tutoring by teachers

In the years since *Reading Recovery* was introduced, many other one-to-one tutoring programs have been developed and evaluated. Studies of those that used teachers as tutors are summarized in Table 1. These include programs such as *Auditory Discrimination in Depth*, *Early Steps/Howard Street Tutoring*, *Reading Rescue*, and *Targeted Reading Intervention*. The 12 studies of these programs had a weighted mean effect size of +0.56.

### 5.4. Conclusions: one-to-one tutoring by teachers

Across all studies of one-to-one tutoring by teachers, there were 20 qualifying studies (including 5 randomized and 3 randomized quasi-experiments). The overall weighted mean effect size was +0.39. Eight of these, with a weighted mean effect size of +0.23, evaluated *Reading Recovery*. Twelve studies evaluated a variety of other one-to-one approaches, and found a weighted mean effect size of +0.56. Discussions of the importance of phonics in tutoring, long-term effects, and comparisons of one-to-one tutoring by teachers to tutoring by paraprofessionals and volunteers, to small-group tutorials, and to classwide preventive interventions, appear later in this article.

## 6. Tutoring by paraprofessionals and volunteers

One-to-one tutoring by certified teachers is expensive, and in high-poverty communities with shortages of teachers, allocating scarce qualified teachers to small numbers of children may be hard to justify. For those reasons, many schools have long used paraprofessionals or volunteers as tutors, usually with materials specifically designed for this purpose. Ritter et al. (2009) and Wasik (1997) reviewed research on volunteer tutoring programs, and both reported substantial positive effects. Table 2 summarizes research on the reading outcomes of one-to-one tutoring programs in which the tutors are paraprofessionals or volunteers.

A total of 11 studies (6 randomized) evaluated one-to-one tutoring by paraprofessionals, with a weighted mean effect size of +0.38. Particularly promising effects were found for *Sound Partners*, *SMART*, *Reading Rescue*, and *Howard Street Tutoring*. Effect sizes across 7 studies of volunteer tutoring (4 fully randomized) averaged +0.16. However, this mean was brought down by the large sample sizes and low effect sizes of the studies of *Experience Corps* and the *West Philadelphia Tutoring Program*, which provided tutoring only 1–2 times a week. Excluding these, the weighted mean effect size for 5 studies of volunteer tutoring was +0.51. Especially positive effects were reported for *Book Buddies*, and in small studies by Pullen, Lane, and Monaghan (2004), Rimm-Kaufman, Kagan, and Byers (1999), and Allor and McCathren (2004). These five small studies had better-qualified and better-trained paraprofessionals and volunteers than might be common.

The effects seen for paraprofessional tutors and for volunteer tutors using structured and intensive programs pose a real challenge to the idea that only certified teachers can be effective tutors. The overall effect size for the 18 studies,  $ES = +0.24$ , is about the same as the mean for *Reading Recovery*, although much lower than the effects for phonetic tutoring by teachers ( $ES = +0.62$ ). It is important to note that studies by Brown, Morris, and Fields (2005) and Ehri, Dreyer, Flugman, and Gross (2007) compared certified and paraprofessional tutors and found the certified tutors to be much more effective, but both studies still found substantial positive effects for the paraprofessional tutors. What these findings imply is that schools might use a mix of certified, paraprofessional, and volunteer tutors, using the certified teachers as leaders and to work with the most difficult children. One-to-one tutoring is clearly very effective, and when resources are limited, well-structured

**Table 2**

Tutoring by paraprofessionals and volunteers.

Study	Design large/small	Duration	N	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size	
<i>Tutoring by Paraprofessionals</i>									
<i>Sound Partners</i>									
Jenkins, Peyton, Sanders, and Vadasy (2004)	Randomized (S)	1 year	11 schools 99 students (79E, 20 C)	1	Urban schools in the northwest	Woodcock Johnson	Word ID Word Attack Passage Comprehension	+0.50 +0.77 +0.81	+0.69
						TOWRE	Phonemic Decoding Sight Word Reading	+0.38 +0.52	
						Bryant WRAT-Reading Woodcock Johnson		+1.13 +0.74	+0.34
Mooney (2003)	Randomized (S)	1 year	7 schools 47 students (28E, 19C)	1	Students at risk for emotional and behavioral disorders in midwestern city; 55% FL, 68%W, 21%AA, 9%H	Woodcock Johnson	Basic Skills Comprehension	+0.26 +0.22	
						DIBELS	Nonsense Word Fluency Oral Reading Fluency	+0.54 +0.35 +0.49	+0.52
Vadasy, Sanders, and Tudor (2007)	Randomized (S)	15 weeks	43 students (23E, 20C)	2–3	Urban schools in the northwest	Woodcock Word Attack and Word ID composite		+0.55	+0.71
Vadasy, Sanders, and Peyton (2005)	Matched (S)	1 year	57 students (38E, 19C)	1	Urban schools in the northwest	Woodcock Johnson	Word ID Word Attack Passage Comprehension	+0.86 +1.06 +0.66	
						TOWRE	Phonemic Decoding Sight Words Reading Rate Reading Accuracy	+0.55 +0.56 +0.33 +0.66 +1.02	
						WRAT			
<i>The Reading Connection</i>									
Compton (1992)	Randomized (L)	1 semester	483 students (266E, 217C)	1	Kalamazoo, MI 53% minority	ITBS			+0.22
<i>Start Making a Reader Today (SMART)</i>									
Baker, Gersten, and Keating (2000)	Randomized (S)	2 years	6 schools 24 classrooms 84 students (43E, 41C)	1–2	Oregon	Woodcock Johnson	Word ID Passage Comprehension Word Comprehension Oral Reading Fluency	+0.62 +0.36 +0.46 +0.54	+0.50
<i>Edmark Reading Program</i>									
Mayfield (2000)	Randomized (S)	1 semester	3 schools 60 students (31E, 29C)	1	Low-achieving non-special-ed students in rural Louisiana	Woodcock Johnson	Passage Comprehension Letter–Word ID Word Attack	+0.51 +0.14 +0.03	+0.23
<i>Wallach and Wallach</i>									
Dorval, Wallach, and Wallach (1978)	Matched (S)	1 year	58 students (20E, 38C)	1	Rural Roanoke Rapids, NC	CTBS Total Reading	Same-school controls Different-school controls	+0.66 +0.77	+0.71
<i>Programmed Tutorial Reading</i>									
McGrady (1984)	Matched (S)	1 year	6 schools 69 students (35E, 34C)	4	Rural/suburban middle class schools in Indiana. Students below 37th percentile	ITBS Comprehension			+0.20

<i>Reading Rescue</i> Ehri et al. (2007) (Paraprofessional tutors)	Matched (S)	6 months	96 students (26E, 70C)	1	Spanish-dominant students in a large city; 95% FL	Gates MacGinitie		+0.89
<i>Howard Street Tutoring</i> Brown et al. (2005) (Paraprofessional tutors)	Matched (S)	1 year	63 students (21E, 42C)	1	Urban schools in the intermountain west. 46% FL; 42% minority, 25% ELLs	Woodcock Johnson	Word Attack Passage Comprehension Word Recognition Passage Reading	+0.55 +0.07 +0.95 +0.60 +0.58
Tutoring by Volunteers <i>Experience Corps</i> Morrow-Howell, Jonson-Reid, McCrary, Lee, and Spitznagel (2009)	Randomized (L)	1 year	881 students (430E, 451C)	1–3	Schools in Boston, New York, Port Arthur, Texas 94%FL, 58%AA, 36%H, 24%ELL	Woodcock Johnson	Word Attack Passage Comprehension	+0.11 +0.10 +0.13
<i>Book Buddies</i> Meier and Invernizzi (2001)	Randomized (S)	4 months	55 students (28E, 27C)	1	South Bronx, NY Students in lowest 25% of grade at end of K; 99% FL, 69%H, 30%AA	WRAT-Reading Words read correctly per minute		+1.00 +0.78
<i>HOSTS</i> Ramey (1991)	Matched (S)	1 year	238 students (18E, 220C)	2–5	Low-achieving students in Seattle	CAT-Reading		+0.05
<i>Other Volunteer Tutoring</i> Ritter (2000)	Randomized (L)	1 year	11 schools 385 students (196E, 189C)	2–5	Philadelphia, PA. 87% FL, 96%AA	SAT-9		–0.10
Pullen et al. (2004)	Randomized (S)	3 months	10 schools 47 students (23E, 24C)	1	North-central Florida 53%W, 38%W, 58%FL	Woodcock Johnson	Letter ID Word Attack	+0.52 +0.23 +0.80
Rimm-Kaufman et al. (1999)	Randomized (S)	1 year	42 students (21E, 21C)	1	Cambridge, MA 29%AA, 26% Haitian Creole, 26% W, 60% FL	Clay Observational Survey	Reading Level Word Knowledge	+0.27 +0.35 +0.18
Allor and McCathren (2004)	Cohort 1 Matched Cohort 2 Randomized	6 months	Cohort 1 8 schools 86 students (61E, 25C) Cohort 2 10 schools 157 students (76E, 81C)	1	Cohort 1: 100%FL, 100%AA Cohort 2: 94%FL, 96% AA	Cohort 1 Woodcock Johnson	Word ID Word Attack Passage Comprehension	+0.54 +0.59 +0.93 +0.49
						TOWRE Real Words TOWRE Non-Words Cohort 2 Woodcock Johnson		+0.41 +1.44
						Word ID Word Attack Passage Comprehension		+0.11 +0.80 –0.16
						TOWRE Real Words TOWRE Non-Words DIBELS Non-sense Word Fluency		+0.14 +0.61 +0.31

Note: L = large study with at least 250 students; S = small study with less than 250 students; E = Experimental; C = Control; FL = Free/reduced-price lunch; W = White; AA = African American; H = Hispanic; ELL = English language learner; TOWRE = Test of Word Reading Efficiency; WRAT = Wide Range Achievement Test; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; ITBS = Iowa Test of Basic Skills; CAT = California Achievement Test.

programs making use of paraprofessionals and volunteers may reach more struggling readers for the same cost as serving many fewer children with certified teachers. The comparison between certified and paraprofessional and volunteer tutors is discussed further later in this article.

## 7. Small-group tutorials

The most common form of remedial or supplementary instruction for struggling readers is additional teaching in small groups, typically 30–45 min daily. Since the 1960s, small-group remediation has been the predominant use of Title I funds, and it is the most common format for children with reading disabilities. Small-group tutorials are potentially more cost-effective than one-to-one tutoring from teachers, because several children are taught at the same time, and the group setting creates possibilities for children to learn from each other as well as from the teacher. On the other hand, small-group teaching can be more of the same type of instruction that did not work the first time, can be difficult to coordinate with regular classroom instruction, and does not allow teachers to tailor instruction to students' needs as much as one-to-one instruction does.

Table 3 summarizes research on the reading outcomes of participation in small-group tutorials. A total of 20 studies evaluated 18 different models of small-group tutorials. 16 of these used random assignment to conditions. The overall mean effect size was +0.31.

It is important to note that the studies that met the inclusion criteria were not run-of-the-mill Title I pullout programs. Instead, all were named programs with extensive training, materials, and other supports, and a strong emphasis on phonics. Small-group models with promising effects include *Corrective Reading*, *QuickReads*, *Targeted Intervention*, *Proactive Reading*, *Responsive Reading*, *Read Naturally*, *Voyager Passport*, *Empower Reading*, *SHIP*, *Early Intervention in Reading*, *Read, Write, and Type*, and *LiPS*.

## 8. Classroom instructional process programs

One potential solution to the reading problems of many struggling readers is to enhance the quality of instruction in their regular classrooms. In previous reviews of beginning reading and upper-elementary reading programs (Slavin, Lake, et al., 2009; Slavin, Madden, et al., 2009), classroom instructional process programs were clearly the most effective and well-evaluated approaches for children in general. Introduction of training in programs that have been successfully evaluated with students in general, such as various forms of cooperative learning and phonics-oriented class programs, could be particularly beneficial for children who would otherwise have difficulty in learning to read. From the perspective of response to intervention, improving classroom instruction is Tier 1 of a plan to improve learning for struggling readers and reduce special education placements. Use of effective classroom strategies does not preclude individually targeted interventions for the hopefully small number of children who still need them, despite high-quality classroom teaching.

Table 4 summarizes research on the outcomes of these types of programs for children in the lowest performance levels of their classes (usually defined as the lowest 25%).

The effect sizes across 16 studies of classroom instructional process programs (1 randomized and 3 randomized quasi-experiments) were very positive for students at the lowest performance levels in their classes. The weighted mean was +0.56, similar to the findings for one-to-one phonetic tutoring. These effects are markedly more positive for low achievers than they were for the overall samples of students of all performance levels in reading, reported by Slavin, Lake, et al. (2009). Eight of the 16 studies involved forms of cooperative learning, *CIRC* (ES = +0.46 in 3 studies), *PALS* (ES = +0.49 in 4 studies), and same-age tutoring (ES = +1.55 in one small study). The mean for all 8 studies of cooperative learning was +0.58. Other particularly promising effects were found for programs that utilize structured, systematic, phonetic approaches to reading instruction: *Direct Instruction*, *Project Read*, *RAILS*, and *Precision Teaching*.

## 9. Classroom instructional process with tutoring (*Success for All*)

This section presents research on a single program, *Success for All*, which provides extensive school staff training and materials to improve all aspects of school organization and functioning, especially those aspects relating to reading, and also provides tutoring to struggling children, mostly first graders. The classroom interventions use a structured, fast-paced approach with a strong emphasis on cooperative learning, phonics, metacognitive skills, and frequent assessment. In second grade and beyond, *Success for All* uses an adaptation of *CIRC*, described earlier. Parent involvement and interventions for behavior and other non-academic problems are also emphasized. In contrast to one-to-one tutoring programs such as *Reading Recovery*, which provide intensive tutoring in first grade but no intervention afterwards, *Success for All* continues to provide classroom-level interventions (though not tutoring) throughout the elementary grades. Research on the reading outcomes of *Success for All* for students in the lowest performing segments of their classes is summarized in Table 5.

The weighted mean effect size for the lowest achievers in *Success for All* across 9 qualifying studies was +0.55, similar to the effect size for phonetic tutoring programs. This includes a national cluster randomized evaluation, carried out over a 3-year period (Borman et al., 2007). Where the results are different from one-to-one tutoring, however, is in long-term outcomes. Most of the *SFA* studies provided the program over at least a 3-year period, and generally found stable or increasing effect sizes over the years (see, for example, Borman et al., 2007). Slavin et al. (1993) evaluated continued treatment over 6 years (K-

**Table 3**  
Small-group tutorials.

Study	Design large/small	Duration	N	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size
<i>Corrective Reading: Decoding</i>								
Torgesen et al. (2006, 2007)	Randomized (S)	1 year	16 schools 165 students (99E, 66C)	3 and 5	Schools around Pittsburgh 44% FL, 81% W, 19% AA	Average of Woodcock, TOWRE, AIMSweb, and GRADE 3rd grade 5th grade	+0.22 +0.09	+0.16
Hempenstall (2008)	Matched (S)	7 months	206 students (134E, 72C)	Ages 8–11	Disadvantaged students in a Melbourne, Australia suburb	Woodcock Word Attack		+1.22
<i>Spell Read</i>								
Torgesen et al. (2006, 2007)	Randomized (S)	1 year	16 schools 196 students (115E, 81C)	3 and 5	Schools around Pittsburgh 44%FL, 69%W, 31%AA	Average of Woodcock, TOWRE, AIMSweb, and GRADE 3rd grade 5th grade	+0.21 +0.12	+0.17
<i>Wilson Reading</i>								
Torgesen et al. (2006, 2007)	Randomized (S)	1 year	16 schools 158 students (103E, 55C)	3 and 5	Schools around Pittsburgh, 48%FL, 56%W, 44%AA	Average of Woodcock, TOWRE, AIMSweb, and GRADE 3rd grade 5th grade	+0.26 +0.08	+0.17
<i>Failure Free Reading</i>								
Torgesen et al. (2006, 2007)	Randomized (S)	1 year	16 schools 219 students (113E, 104C)	3 and 5	Schools around Pittsburgh; 44%FL, 80%W, 20%AA	Average of Woodcock, TOWRE, AIMSweb, and GRADE 3rd grade 5th grade	+0.10 0.00	+0.05
<i>Quick Reads</i>								
Vadasy and Sanders (2008)	Randomized (S)	15 weeks	162 students (82E, 80C)	2–3	Urban Washington State; 30%W, 28%AA, 23%H, 16% Asian	Woodcock Word ID TOWRE Sight Word GORT Comprehension GORT Fluency DIBELS Fluency	+0.27 +0.12 +0.16 +0.30 +0.27	+0.22
Vadasy and Sanders (2008)	Randomized (S)	20 weeks	12 schools 119 students (54E, 65C)	4–5	Urban northwest; 40%AA, 25%W, 12%H, 9% Asian	Woodcock Johnson  TOWRE Sight Word DIBELS Fluency	Word ID Word Comprehension Passage Comprehension +0.33 +0.09 +0.23 +0.21 +0.16	+0.20
<i>Targeted Intervention</i>								
Wang and Algozzine (2008)	Randomized (S)	1 year	6 schools (4E, 2C) 139 students (101E, 38C)	1	Urban schools, 80%FL, 89%AA Woodcock or H	Word ID Word Attack Passage Comprehension  DIBELS Nonsense Word	+0.32 +0.43 +0.13  –0.12	+0.19
<i>Proactive Reading</i>								
Mathes et al. (2005)	Randomized (S)	1 year	6 schools 162 students (80E 82C)	1	Large urban district in Texas	Woodcock Johnson  Word ID Word Attack Passage Comprehension	+0.76 +0.51 +0.21	+0.49

Table 3 (Continued)

Study	Design large/small	Duration	N	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/ measure	Overall effect size	
<i>Responsive Reading</i> Mathes et al. (2005)	Randomized (S)	1 year	165 students (83E, 82C)	1	Large urban district in Texas	Woodcock Johnson	Word ID Word Attack Passage Comprehension	+0.28 +0.36 +0.30	+0.31
<i>New Heights Reading Program</i> Lesnick (2006)	Randomized (S)	18 weeks	9 schools 59 classes (30E, 29C) 233 students (118E, 115C)	3 and 5	Philadelphia and suburban PA	DIBELS TOWRE DRP	–0.01 –0.02 +0.09	+0.02	
<i>Read Naturally</i> Heistad (2005)	Matched (S)	1 year	2 schools 102 students (51E, 51C)	3 and 5	Minneapolis, MN	NALT Minnesota Comprehensive Assessments	+0.21 +0.34	+0.27	
<i>Voyager Passport</i> Ehri et al. (2007)	Matched (S)	6 months	122 students (52E, 70C)	1	Spanish-dominant students in a large city 95%FL	Gates MacGinitie	Decoding Comprehension	+0.26 +0.36	+0.31
<i>Empower Reading</i> Lovett et al. (2000)	Randomized (S)	14 weeks	37 students (15 E, 22C)	1–4	Students with reading disabilities in Toronto	Woodcock Johnson	Word Attack Word ID Passage Comprehension	+1.23 +0.45 +0.49	+0.71
Lovett et al. (2008)	Randomized (S)	1 year	166 students (122E, 44C)	2–8	Below-grade-level students in Toronto	WRAT Woodcock Johnson	Word Attack Word ID Passage Comprehension	+0.68 +0.22 +0.37 +0.02	+0.25
<i>Schools and Homes in Partnership (SHIP)</i> Gunn et al. (2005)	Randomized (S)	2 years	211 students (105E, 106C)	K–3	Rural districts in central Oregon, 62%H, 38%W	WRAT Reading Woodcock Johnson	Letter–Word ID Word Attack Vocabulary Comprehension	+0.40 +0.31 +0.66 +0.20 +0.29	+0.34
<i>Gottshall Small-Group Phonics</i> Gottshall (2007)	Randomized (S)	15 weeks	64 students (35E, 29C)	1	Boys in rural Nacogdoches, TX, 42%AA, 34%H, 20%W	DIBELS Oral Reading Fluency TPRI	DIBELS Oral Reading Fluency	+0.24	–0.10
<i>Early Intervention in Reading</i> Taylor, Short, Frye, and Shearer (1992)	Matched (S)	1 year	60 students (30E, 30C)	1	Students in a suburban Midwestern district	Gates MacGinitie	Graphophonemic Knowledge Reading Accuracy Reading Comprehension	–0.05 –0.01 –0.24	+0.82

<i>Read, Write, and Type-Small Group</i>									
Torgesen et al. (2009)	Randomized (S)	1 year	73 students (34E, 39C)	1	Florida schools	Woodcock Johnson			+0.36
							Word ID	+0.41	
							Word Attack	+0.59	
							Passage Comprehension	+0.33	
						TOWRE			
							Non-word	+0.26	
							Word	+0.22	
<i>Lindamood Phoneme Sequence Program-Small Group</i>									
Torgesen et al. (2009)	Randomized (S)	1 year	74 students (35E, 39C)	1	Florida schools	Woodcock Johnson			+0.66
							Word ID	+0.63	
							Word Attack	+0.93	
							Passage Comprehension	+0.46	
						TOWRE			
							Non-word	+0.79	
							Word	+0.50	

Note: L = large study with at least 250 students; S = small study with less than 250 students; E = experimental; C = control; FL = free/reduced-price lunch; W = White; AA = African American; H = Hispanic; ELL = English language learner; TOWRE = Test of Word Reading Efficiency; GRADE = Group Reading Assessment and Diagnostic Examination; GORT = Gray Oral Reading Test; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; DRP = Degrees of Reading Power; NALT = Northwest Achievement Levels Test; WRAT = Wide Range Achievement Test; TPRI = Texas Primary Reading Inventory; CTOPP = Comprehensive Test of Phonological Processing.

**Table 4**  
Classroom instructional process approaches.

Study	Design large/small	Duration	N	Grade	Sample characteristics	Posttest		Effect sizes by subgroup/ measure	Overall effect size
<i>Cooperative Integrated Reading and Competition</i> Stevens and Slavin (1995a)	Matched (S)	2 years	137 students (72E, 65C)	2–6	Special Education students in working-class suburb of Baltimore 9%FL, 95%W	CAT	Vocabulary	+0.37	+0.35
Stevens and Slavin (1995b)	Matched (S)	2 years	76 students (40E, 36C)	2–6	Suburban Maryland 10%FL, 92%W	CAT	Comprehension	+0.32	
Bramlett (1994)	Matched (S)	1 year	149 students (82E, 67C)	3	Rural southern Ohio	CAT	Vocabulary	+0.81	
<i>PALS</i> Mathes and Babyak (2001)	Randomized quasi-experimental (S)	14 weeks	20 classes 56 students (27E, 29C)	1	Medium-sized district in Florida 50%W, 48% AA	Woodcock Johnson	Comprehension	+0.85	+0.38
							Total Reading	+0.33	
							Word Analysis	+0.56	
							Vocabulary	+0.30	+0.59
							Word Identification	+0.51	
							Word Attack	+0.89	
Mathes, Torgesen, and Allor (2001)	Matched (S)	16 weeks	24 classes (12E, 12C) 75 students (42E, 33C)	1	Southeastern district 47%W, 51%AA	TOWRE	Basic Skills	+0.71	+0.48
							Passage Comprehension	+0.23	
							Non-Word Efficiency	+0.48	
						Woodcock	Word Efficiency	+0.34	+0.46
							Word Identification	+0.43	
							Word Attack	+0.58	
Mathes, Howard, Allen, and Fuchs (1998)	Matched (S)	16 weeks	20 classes 56 students (28E, 28C)	1	Southeastern urban district 37%W, 63%AA	Woodcock	Basic Skills	+0.55	+0.43
							Passage Comprehension	+0.50	
							Word Identification	+0.51	
							Word Attack	+0.69	+0.43
							Passage Comprehension	+0.19	
							Non-Word	+0.48	
Mathes et al. (2003)	Matched (S)	16 weeks	15 teachers (7E, 8C) 59 students (31E, 28C)	1	Low achievers in a southeastern district 29%FL, 49%W, 42%AA	TOWRE	Word Efficiency	+0.13	+0.43
							Word ID	+0.41	
							Word Attack	+0.98	
						Woodcock	Passage Comprehension	+0.13	



<i>Same Age Tutoring</i> Eldredge, Quinn, and Quinn (1988)	Randomized quasi-experiment (S)	1 year	5 schools 64 students (32E, 32C)	2	Middle class schools in Provo, UT	Gates MacGinitie		+1.55
							Comprehension Vocabulary	+1.33 +1.77
<i>Reading and Integrated Literacy Strategies (RAILS)</i> Stevens et al. (2008)	Matched (S)	1 year	3 schools 87 students (51E, 36C)	K-5	Low-achieving schools in a small city in central PA. 71%FL, 94%W	MAT		+0.49
							Special Education Students K-2 Special Education Students 3–5	+0.47 +0.50
<i>Contextually-Based Vocabulary Instruction</i> Nelson and Stage (2007)	Randomized quasi-experiment (S)	3 months	73 students (41E, 32C)	3, 5	Midwestern schools 70%W, 24% H	Gates MacGinitie		+0.41
							Comprehension Vocabulary	+0.60 +0.23
<i>Reading Styles</i> LaShell (1986)	Matched (S)	1 year	90 students (47E, 43C)	2–6	Students with learning disabilities in rural Snohomish County, WA	Gray Oral Reading Test		+0.79
Brooks (1991)	Matched (S)	1 semester	2 schools 42 students (22E, 20C)	2–6	Title I students in Northwest Ohio	Spadafore Diagnostic Reading Test		+0.36
							Oral Reading Silent Reading	+0.21 +0.51
<i>Direct Instruction</i> Bowers (1972)	Randomized (S)	1 year	8 classes (4E, 4C) 123 students (60E, 63C)	1	Urban schools in Oklahoma 100%W	Gates MacGinitie		+0.25
							Comprehension Vocabulary	+0.15 +0.35
Davis (1995)	Matched (S)	1 year	2 schools 111 students (59E, 52C)	2	Title I students in southern Mississippi	SAT		+0.49
<i>Project Read</i> Greene (2001)	Matched (S)	1 year	224 students (112E, 112C)	1–3	Students below 25th percentile in Louisiana	CAT		+0.59
<i>Precision Teaching</i> Haring and Krug (1975)	Matched (S)	1 year	4 schools 54 students (24E, 30C)	Ages 9–12	Disadvantaged students with mental retardation; 70% AA	WRAT		+1.18

Note: L = large study with at least 250 students; S = small study with less than 250 students; E = experimental; C = control; FL = free/reduced-price lunch; W = White; AA = African American; H = Hispanic; ELL = English language learner; CAT = California Achievement Test; MAT = Metropolitan Achievement Test; TOWRE = Test of Word Reading Efficiency; SAT = Scholastic Achievement Test, WRAT = Wide Range Achievement Test.

**Table 5**  
Classroom instructional process with tutoring (*Success for All*).

Study	Design large/small	Duration	N	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size
Borman et al. (2007)	Randomized (L)	3 years	35 schools 703 students (362 E, 341 C)	K-2	Title I schools throughout the US, 72% FL, 57% AA, 31% W, 10% H	Woodcock	Word Identification +0.22 Word Attack +0.40 Passage Comprehension +0.22	+0.28
Madden et al. (1993); Slavin et al. (1993)	Matched (L)	6 years	330 students (165E, 165C)	1–5	African American students in high-poverty schools in Baltimore, MD scoring in the lowest 25%	Average of Woodcock and DORT/Gray	1st grade +1.18 2nd grade +0.92 3rd grade +1.32 4th grade +0.80 5th grade +1.03	+1.05
Ross, Nunnery, and Smith (1996)	Matched (S)	1 year	4 schools (2 E, 2 C) 138 students (42E, 96C)	1	Lowest performers in mostly Hispanic schools in Amphitheater District near Tucson, AZ	Woodcock	Word Identification +0.44 Word Attack +1.07 Passage Comprehension +0.30	+0.54
Ross and Casey (1998)	Matched (S)	2 years	8 schools 92 students (36E, 56C)	K-1	High-poverty schools in Ft. Wayne, IN; 75% FL, 45% minority	Durrell Oral Reading Average of Woodcock and Durrell	+0.37	+0.35
Ross, Smith, and Casey (1994)	Matched (S)	3 years	2 schools 93 students (58E, 38C)	1–3	Students in Caldwell, ID	Average of Woodcock and Durrell		–0.20
Casey, Smith, and Ross (1994)	Matched (S)	1 year	4 schools 65 students (49E, 16C)	1	Schools in Memphis, TN	Average of Woodcock and Durrell		+0.54
Ross et al. (1993)	Matched (S)	3 years	4 schools 47 students (23E, 24C)	1–3	AA schools in Montgomery, AL	Average of Woodcock and Durrell		+1.16
Ross et al. (1995)	Matched (S)	4 years	4 schools 41 students (24E, 17C)	1–4	Students in Ft. Wayne, IN	Average of Woodcock and Durrell		+0.45
Smith, Ross, and Casey (1994)	Matched (S)	4 years	2 schools 38 students (21E, 17C)	1–4	African American schools in Memphis, TN	Average of Woodcock and Durrell		+1.14

Note: L = large study with at least 250 students; S = small study with less than 250 students; E = experimental; C = control; FL = free/reduced-price lunch; W = White; AA = African American; H = Hispanic; ELL = English language learner; DORT = Durrell Oral Reading Test.

5), and then a follow-up study by Borman and Hewes (2003) followed SFA students to eighth grade. These long-term studies found that positive effects of *Success for All* maintained over time. This is in contrast to the disappointing findings of long-term follow-ups of one-to-one tutoring alone without classroom interventions after tutoring. These findings may suggest more broadly that even the most effective first-grade tutoring approaches require follow-up with ongoing intervention in the later grades. This issue is discussed further later in this article.

## 10. Instructional technology

Over the past 30 years, one of the most common solutions applied for children who are struggling to learn to read is to give them computer-assisted instruction (CAI) software. Modern CAI programs adapt to children's specific needs and give them activities with graphics and exciting elements that can supplement classroom instruction. However, previous reviews of research on elementary CAI applications in reading find few positive effects (Dynarski et al., 2007; Kulik, 2003; Slavin, Lake, et al., 2009; Slavin, Madden, et al., 2009). Table 6 summarizes research on outcomes of instructional technology for children who are at the lowest performance levels of their classes.

Across 14 qualifying studies (5 randomized), IT had minimal impacts on the achievement of struggling readers. The weighted mean effect size was only +0.09.

## 11. Findings on key issues

The review of programs for struggling readers examined many important questions beyond the overall outcomes of various approaches. The following sections address these questions. In each case, we consider both within-study and between-study comparisons. The within-study comparisons use the same study inclusion criteria as those applied for the main review, but in a few cases studies that compared alternative treatments but did not qualify for inclusion due to the lack of a control group representing ordinary practice are cited if they met all other inclusion criteria.

### 11.1. The importance of phonics

Across all categories of programs, almost all successful programs have a strong emphasis on phonics. As noted earlier, one-to-one tutoring programs in which teachers were the tutors had a much more positive weighted mean effect size if they had a strong phonetic emphasis (mean ES = +0.62 in 10 studies). One-to-one tutoring programs with less of an emphasis on phonics, specifically *Reading Recovery* and *TEACH*, had a weighted mean effect size of +0.23. Within-study comparisons support the same conclusion. Averaging across five within-study comparisons, the mean difference was +0.18 favoring approaches with a phonics emphasis.

### 11.2. Long-term effects of first-grade tutoring

A key part of the argument for intensive one-to-one tutoring for struggling first graders depends on the idea that tutoring can bring most struggling readers up to grade level and then they will remain normal readers from then on. The cost-effectiveness argument for very expensive programs such as *Reading Recovery* make the case that the expense is justified by reductions in the need for remedial services or retentions in later years.

Surprisingly, few long-term follow-ups of outcomes of early tutoring have been done, but the existing evidence does not support the contention that successful first-grade tutoring has long-lasting effects. The best studies of these long-term outcomes both involved children in high-poverty schools in the UK. Hurry and Sylva (2007) followed up London 6 and 7 year olds who received *Reading Recovery*. At the end of their tutoring year, the *Reading Recovery*-tutored children scored substantially better than matched children who were not tutored (ES = +0.85,  $p < .001$ ). However, a year later, at age 8, the effect size dropped to +0.40 ( $p < .01$ ), and then to +0.15 (n.s.) at age 10. Burroughs-Lange (2008) similarly found an effect size averaging +0.76 at the end of year 1, but by year 4 the effect size was only +0.20.

Studies by Hurry and Sylva (2007) and Hurry and Holliman (2009) found strong effects on treatment-independent measures initially, but these effects dissipated over time. Pinnell et al. (1988) found a similar pattern on *Reading Recovery*'s Text Reading Level measure. Effect sizes were +0.80 in first grade, +0.47 in second grade, and +0.26 in third grade. Another 5-year follow-up study in Detroit by Huggins (1999) found no differences on state accountability tests between former *Reading Recovery* and control students in fifth grade (ES = +0.13, n.s.), but this study also found no differences at first grade (ES = -0.09, n.s.). A follow-up study of *Reading Recovery*-tutored students in North Carolina by Baenen, Bernholz, Dulaney, and Banks (1997) found no differences among third graders who had been tutored in first grade with *Reading Recovery* and matched control students on North Carolina End-of-Grade tests.

While there is little evidence to suggest that first-grade tutoring alone is sufficient to maintain struggling students at high levels, an interesting point of contrast comes from a long-term Baltimore evaluation of *Success for All*. Students in the lowest 25% of their schools were followed by Slavin et al. (1993) from first to fifth grades, and were tested each year on Woodcock and Durrell measures. The initially lowest-achieving students in *Success for All* scored substantially better than matched controls at the end of first grade (ES = +1.18,  $p < .001$ ), and they maintained these gains through fifth grade (ES = +1.03,  $p < .001$ ). A follow-up to eighth grade by Borman and Hewes (2003) found that the former SFA students continued to score significantly

**Table 6**  
Instructional technology.

Study	Design large/small	Duration	N	Grade	Sample characteristics	Posttest	Effect sizes by subgroup/measure	Overall effect size
<i>Jostens (early form of Compass Learning)</i>								
Becker (1994)	Randomized (S)	1 year	1 school 56 students	2–5	Low performing students in inner city Baltimore	CAT		+0.41
Sinkis (1993)	Matched (S)	1 year	422 students (228E, 194C)	3, 5, 6	Urban northeast	MAT Comprehension	Grade 3 Grade 5 Grade 6	+0.14 +0.22 –0.01
Standish (1995)	Matched (S)	1 year	43 students (22E, 21C)	2	Title I students in suburban Delaware	MAT6 Reading		+0.55
<i>Fast ForWord</i>								
Rouse and Krueger (2004)	Randomized (L)	1 year	4 schools 454 students (237E, 217C)	3–6	High-poverty Hartford (CT) schools 59% FL, 66% H, 27% AA, 61% ELL	Connecticut Mastery Test		+0.05
Marion (2004)	Matched (S)	1 year	63 students (34E, 29C)	5–6	Schools in Appalachian TN 52% FL, 100% W	Terra Nova		+0.15
<i>Lexia</i>								
Macaruso et al. (2006)	Matched (S)	1 year	10 schools 167 students (83E, 84C)	1	Title I students in schools in Boston, MA	Gates MacGinitie		+0.67
<i>Other Supplemental CAI</i>								
Dynarski et al. (2007); Campuzano et al. (2009) - Destination Reading - Waterford - Headsprout - Plan Focus -Academy of Reading	Randomized (L)	1 year	Cohort 1: 872 students (505E, 367C) Cohort 2: 232 Students (130E, 102C)	1	National. 49% FL, 44%W, 31%AA, 22%H	SAT-9	Cohort 1 Cohort 2	+0.02 –0.39
Dynarski et al. (2007); Campuzano et al. (2009); -LeapTrack - Academy of Reading -Read 180 -Knowledge Box (cohort 1)	Randomized (L)	1 year	Cohort 1: 755 students (410E, 345C) Cohort 2: 95 Students (52E, 43C)	4	National. 64% FL, 17%W, 57%AA, 23%H	SAT-10	Cohort 1 Cohort 2	–0.01 +0.48
Becker (1994)	Randomized (S)	1 year	60 students	2–5	Schools in Baltimore, MD; 50% FL	CAT-Reading		+0.10
Ramey (1991)	Matched (L)	1 year	282 students (62E, 220C)	2–5	Urban Washington State	CAT-Reading		+0.22
Bass, Ries, and Sharpe (1986)	Matched (S)	1 year	2 schools (1E, 1C) 145 students (73 E, 72 C)	5–6	High-poverty schools in rural VA	SRA Virginia Basic Learning Skills Test		+0.22 +0.13
Chiang et al. (1978)	Matched (S)	1 year	4 schools 137 students (65E, 72C)	7–12-year-olds	Cupertino, CA, a middle class suburb of San Francisco	PIAT	Reading Recognition Reading Comprehension	+0.18 +0.26
Roth and Beck (1987)	Matched (S)	1 year	6 classes (3E, 3C) 37 students (20E, 17C)	4	Low SES urban school	CAT Vocabulary CAT Reading Comprehension		+0.98 –0.10
Coomes (1985)	Matched (S)	1 year	4 schools 36 students (18E, 18C)	4	Middle class schools in TX. 90% W.	CTBS		+0.30

Note: L = large study with at least 250 students; S = small study with less than 250 students; E = experimental; C = control; FL = free/reduced-price lunch; W = White; AA = African American; H = Hispanic; ELL = English language learner; CAT = California Achievement Test; MAT = Metropolitan Achievement Test; SAT = Scholastic Achievement Test; PIAT = Peabody Individual Achievement Test; CTBS = Comprehensive Test of Basic Skills.

better than controls on CTBS Reading ( $ES = +0.34, p < .01$ ), and were significantly less likely to have been retained or assigned to special education. Numerous 3- and 4-year longitudinal studies have also found continuing positive effects of *Success for All* over time (e.g., Borman et al., 2007; Correnti, 2009; Ross, Smith, & Casey, 1995; Ross, Smith, Bond, Casey, & Johnson, 1993). The comparison of these longitudinal studies suggests that tutoring in first grade can have substantial and lasting effects if the tutoring is followed up with improvements in classroom instruction throughout elementary school.

### 11.3. Teachers vs. paraprofessionals and volunteers as tutors

The cost of having certified teachers tutor struggling readers is prohibitive for many schools, so tutoring is often provided by paraprofessionals. Volunteers are also often used as tutors, and volunteer tutoring was the focus of the Clinton Administration's America Reads initiative. Can paraprofessionals and volunteers do as well as teachers in tutoring struggling readers?

The outcomes summarized in Table 2 suggest that paraprofessionals and volunteers can obtain good results with struggling readers, with a mean effect size across 18 studies of +0.24. The mean effect size for paraprofessionals was +0.38 in 11 studies, and for volunteers it was +0.16 in 7 studies (but excluding two studies in which volunteers tutored only once or twice a week, the mean was +0.50). The overall effects for paraprofessionals are similar to those found for all studies of one-to-one tutoring by teachers ( $ES = +0.39$ ). However, almost all of the paraprofessional and volunteer tutoring studies involved programs with a strong phonetic emphasis, so a better comparison is between phonetic tutoring by teachers ( $ES = +0.62$ ) and phonetic tutoring by paraprofessionals ( $ES = +0.39$ ). One small study, by Brown et al. (2005), compared teachers and paraprofessionals as tutors using the same program, and found much better effects for teachers ( $ES = +0.47$ ). Still, children tutored by paraprofessionals scored much better than similar children who were not tutored ( $ES = +0.53$ ). Ehri et al. (2007) found much better outcomes for teachers than for paraprofessionals using the same program ( $ES = +0.52$ ), but again, students tutored by paraprofessionals obtained much better outcomes than non-tutored controls ( $ES = +0.89$ ).

### 11.4. One-to-one vs. small-group tutorials

For many years, the dominant form of assistance for struggling readers has been small-group tutorials, in which students are taught in groups of two or more. In the US, this has been the main use of Title I/Chapter 1 funds since Title I began in 1965, and was emphasized in the Bush Administration's Reading First and Supplemental Educational Services programs. Current US policies on response to intervention (RTI) also suggest small-group tutorials as Tier 2 interventions for struggling readers (see Gersten et al., 2009).

The renewed emphasis on small-group tutorials in US policies was motivated in part by a review of research by Elbaum et al. (2000) that concluded that small-group and one-to-one instruction had similar effects. However, this conclusion was based on just two small dissertations. One, by Evans (1996), did not meet the inclusion requirements of the present review, as it involved only 8 students within a single class and used only *Reading Recovery* Diagnostic Survey measures. The other, by Acalin (1995), compared *Reading Recovery* to *Project Read*, a small-group approach that provides extensive training in phonics and phonemic awareness, hardly a run-of-the-mill small-group treatment.

The data summarized in Table 3 suggest that small-group tutorials with a strong phonetic emphasis and extensive training and follow-up can indeed be effective for struggling readers (weighted mean  $ES = +0.35$  in 22 studies), but the effects are less than those for phonetic tutoring by teachers (mean  $ES = +0.62$ ) and similar to the effect size for tutoring by paraprofessionals ( $ES = +0.38$ ). Within-study comparisons by Ehri et al. (2007) found much larger effects for one-to-one tutoring than for a small-group treatment that used a similar curriculum ( $ES = +0.57$ ).

A study by Vaughn et al. (2003) compared 1-1, 1-3, and 1-10 groupings for struggling second graders. Instructional methods were held constant across groupings. Averaging Woodcock Word Attack, Woodcock Passage Comprehension, and DIBELS Fluency, adjusted posttest effect sizes for monolingual English speakers were +0.32 for 1-1 vs. 1-3 and +0.71 for 1-1 vs. 1-10. For English language learners there were no differences between 1-1 and 1-3 ( $ES = +0.06$ ), but the effect size for 1-1 vs. 1-10 was +0.43.

Simply multiplying the effect sizes for small-group teaching by the number of students involved would suggest that small-group tutoring is more cost-effective, although it is important to note that most small-group tutoring programs provide 30–40 min of daily instruction all year, while one-to-one tutoring is usually given less time per day and for a few months, so differences in teacher time per child are not as large as they might appear. For example, a teacher could equally teach a group of three all year or provide 60 tutoring sessions to each of three children. These findings are important in considering response to intervention, for example, because it suggests that if they can afford it, schools should try to arrange one-to-one tutoring for students in the greatest difficulty. There may also be a broader set of students with milder difficulties for whom small-group tutorials are sufficient. The tradeoff between the number of students served and the effectiveness of tutoring needs to be carefully considered.

### 11.5. Classroom instructional process approaches vs. tutoring

One of the most surprising findings in the present review is the effectiveness of classroom instructional process approaches. The average effect size across 16 studies was +0.56 for students in the lowest performing segment of their

classes, similar to the effect sizes found for one-to-one phonetic tutoring. Most of these instructional process programs are forms of cooperative learning with a strong focus on phonics (e.g., *CIRC* and *PALS*), and other structured phonetic models (e.g., *Direct Instruction*, *RAILS*, and *Project Read*).

What these findings suggest is that it is critical to focus first on core classroom instructional strategies, using methods for the whole class that improve reading performance for all, but particularly for low achievers. These same instructional process programs also work with students in general, but the effect sizes for lowest-achieving students are about twice those reported for students in general (see Slavin, Lake, et al., 2009; Slavin, Madden, et al., 2009). The finding that improvements in classroom instruction are associated with effect sizes like those of tutoring does not imply that tutoring is unnecessary, but rather that professional development for classroom teachers in proven methods should be a major emphasis of programming for struggling readers. There will always be individual children who continue to struggle despite excellent classroom instruction, but the numbers should be much smaller and the remaining difficulties more tractable when initial classroom instruction has used effective instructional process approaches.

### 11.6. Implications for response to intervention

*Response to intervention*, or *RTI* (Allington & Walmsey, 2007; Fuchs & Fuchs, 2006; Gersten et al., 2009) is a policy, currently dominant in the US, that emphasizes the need to provide struggling students with increasing levels of support to attempt to solve their learning problems in the context of general education, before involving the special education system. *RTI* describes three “tiers” of intervention. Tier 1 is regular classroom instruction, Tier 2 is almost always small-group instruction, and Tier 3 may be one-to-one tutoring, other intensive services, or possibly assignment to special education.

The findings of the current review have important implications for *RTI*. First, they emphasize the potential of Tier 1 instruction to enhance the learning of at-risk students. If teachers can significantly enhance the learning of low achievers by adopting cooperative learning or structured phonetic classroom models, this avoids a great deal of frustration, demotivation, and possibly stigmatization for the children themselves, as well as greatly reducing difficulties, expense, and disruption inherent to providing supplemental small-group or tutoring services.

Secondly, the conclusions of this review differ from those of Elbaum et al. (2000), who reported that there were few differences between one-to-one tutoring and small-group tutorials. More than a decade later, there are now many more studies of both types of intervention as well as more within-study comparisons, which show clear differences between phonetic tutoring and phonetic small-group instruction. One-to-one, not small-group instruction, appears to be the most effective Tier 2 or Tier 3 intervention. Because one-to-one tutoring by teachers is expensive, there may be a rationale for Tier 2 interventions that provide students with mild reading problems specific small-group tutoring with extensive professional development, such as *QuickReads*, *Corrective Reading*, *Empower Reading*, *Voyager*, *Early Intervention in Reading*, *LiPS*, *Read, Write, and Type*, or *Read Naturally*. Certainly, the findings of this review would suggest that children who have failed to respond adequately to Tier 1 instruction using proven models and then have failed to respond adequately to proven small-group tutorials should receive one-to-one tutoring using proven models before long-term special education services are considered. The evidence does not support the idea that a relatively brief tutoring experience in first grade is enough to ensure success throughout elementary school and beyond, but it does suggest that with a continuing focus on effective classroom instructional models, most children who receive effective tutoring interventions in first grade can be kept on track in reading.

## 12. Discussion

A total of 97 studies met the inclusion criteria for this review. They compared alternative strategies for helping struggling students in the elementary grades to succeed in reading. Remarkably, 38 of these studies used random assignment to treatments, and 6 used randomized quasi-experiments. Collectively, the studies involved more than 14,000 students. Ninety of the qualifying studies took place in the US, three in England, two in Australia, and two in Canada.

Key findings of the review were as follows.

1. One-to-one tutoring works. Teachers are more effective as tutors than paraprofessionals or volunteers, and an emphasis on phonics greatly improves tutoring outcomes.
2. Although one-to-one phonetic tutoring for first graders is highly effective, effects last into the upper elementary grades only if classroom interventions continue past first grade.
3. Small-group tutorials can be effective, but are not as effective as one-to-one instruction by teachers or paraprofessionals.
4. Classroom instructional process approaches, especially cooperative learning and structured phonetic models, have strong effects for low achievers (as well as other students).
5. Traditional computer-assisted instruction programs have little impact on reading.

These findings support the idea, central to current response to intervention models popular in the US (Fuchs & Fuchs, 2006; Gersten et al., 2009), that the best approach for struggling readers is to provide high-quality instruction in the first place, followed up with intensive instruction to the hopefully small number of students who continue to have difficulties despite high-quality classroom instruction. However, the findings point to a particular focus on improving regular classroom

teaching to prevent reading problems and on the use of one-to-one rather than small-group instruction for students with the most serious difficulties. The findings support the idea that high quality intervention over many years is needed for lasting impacts, in contrast to the expectation that brief, intensive tutoring will put struggling readers permanently on track. Finally, the findings are consistent with those of previous reviews of classroom instructional programs for elementary reading by Slavin, Lake, et al. (2009), which found more positive effects of programs that provide extensive professional development to teachers in proven models than they did for programs that provide technology, alternative curricula, or other interventions that do not change daily teaching practices.

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 of 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 the control group, even though results on such measures may also be of importance to researchers or educators.

Although all studies that met the inclusion criteria took place in the US, UK, Canada, or Australia, the policy implications should apply to all countries in the developed world. Reading difficulties, especially among students in deprived areas, are common in all countries, and local as well as national governments spend enormous amounts on remediation or special services for these children. Knowing the relative impacts of alternative approaches to prevention and intervention should be useful anywhere.

The message of this review is optimistic. There are many proven and promising approaches for struggling readers. It is no longer possible or responsible to do less than what we know how to do to be able to greatly reduce the numbers of children who fail to learn to read adequately. We have both effective and cost-effective tools at hand. While more research is always needed, we already know enough to make a substantial difference in the reading performance of at-risk children.

### For more information

In order to meet the page limitations of this journal, much information about the individual studies and other ancillary material appear only in a web version, available at [www.bestevidence.org](http://www.bestevidence.org). The web version includes descriptions of all included studies, an appendix listing all excluded studies considered germane to the review and the reason for exclusion, and a table providing ratings of the strength of the evidence supporting particular programs. The web version will be continuously updated as additional research is reported.

### Disclosure statement

Some of the programs reviewed, specifically *Success for All* and Cooperative Integrated Reading and Composition, are disseminated by the nonprofit *Success for All* Foundation. Robert Slavin is the unpaid chairman of SFAF, Nancy Madden is the President, and Susan Davis is an employee. None of these people receives any personal royalties or other income based on the levels of use of these programs.

### Acknowledgements

This research was funded by the Institute of Education Sciences, U.S. Department of Education (Grant No. R305A040082). However, any opinions expressed are those of the authors and do not necessarily represent IES positions or policies.

We thank Steve Ross, Yola Center, and Jean Gross for comments on an earlier draft. We also thank Kate Misso and Michele Victor for assistance in locating references.

### References

- Acalin, T. (1995). *A comparison of Reading Recovery to Project READ*. M.S. Dissertation, California State University, Fullerton, United States—California. Retrieved September 5, 2007, from ProQuest Digital Dissertations database (Publication No. AAT 1361908).
- Allington, R. A., & Walmsey, S. (Eds.). (2007). *No quick fix, the RTI edition: Rethinking literacy programs in America's elementary schools*. New York: Teacher's College Press.
- Allor, J., & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disabilities Research & Practice*, 19(2), 116–129.
- Arnold, L. (1977). Prevention by specific perceptual remediation for vulnerable first-graders: Controlled study and follow-up of lasting effects. *Archives of General Psychiatry*, 34(11), 1279–1294.
- Baenen, N., Bernholm, A., Dulaney, C., & Banks, K. (1997). Reading Recovery: Long-term progress after three cohorts. *Journal of Education for Students Placed at Risk*, 2(2), 161–181.

- Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, 35(4), 494–515.
- Bass, G., Ries, R., & Sharpe, W. (1986). Teaching basic skills through microcomputer assisted instruction. *Journal of Educational Computing Research*, 2(2), 207–219.
- Becker, H. J. (1994). Mindless or mindful use of integrated learning systems. *International Journal of Educational Research*, 21(1), 65–79.
- Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., et al. (2004). Effects of intensive reading remediation for second and third grades and a 1-year follow-up. *Journal of Educational Psychology*, 96(3), 444–461.
- Borman, G., & 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., 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.
- Bowers, W. M. (1972). *An evaluation of a pilot program in reading for culturally disadvantaged first grade students*. Unpublished doctoral dissertation, University of Tulsa.
- Bramlett, R. (1994). Implementing cooperative learning: A field study evaluating issues for school-based consultants. *Journal of School Psychology*, 32(1), 67–84.
- Brooks, J. D. (1991). *Teaching to identified learning styles: The effects upon oral and silent reading and listening comprehension*. Doctoral Dissertation, University of Toledo.
- Brooks, G. (2007). *What works for children with literacy difficulties?: The effectiveness of intervention schemes* (3rd ed.). Slough, England: DfCSF/NFER.
- Brown, K., Morris, D., & Fields, M. (2005). Intervention after grade 1: Serving increased numbers of struggling readers effectively. *Journal of Literacy Research*, 37(1), 61–94.
- Burroughs-Lange, S. (2007). *Evaluation of Reading Recovery in London Schools: Every child a reader 2005–2006*. University of London. Available from: <http://www.dcsf.gov.uk/aboutus/reports/pdfs/deptreport2007.pdf>
- Burroughs-Lange, S. (2008). *Comparison of literacy progress of young children in London schools: A Reading Recovery follow-up study*. Retrieved from <http://ioewebsserver.ioe.ac.uk/ioe/cms/get.asp?cid=9263>
- Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts*. Washington, DC: Institute of Education Sciences.
- 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.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., et al. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly*, 30(2), 240–263.
- Cheung, A., & Slavin, R. E. (2005). Effective reading programs for English language learners and other language minority students. *Bilingual Research Journal*, 29(2), 241–267.
- Chiang, A., RMC Research Corp., A., & Others, A. (1978, September 1). *Demonstration of the use of computer-assisted instruction with handicapped children*. Final report. (ERIC Document Reproduction Service No. ED166913).
- Clay, M. (1985). *The early detection of reading difficulties* (3rd ed.). (ERIC Document Reproduction Service No. ED263529).
- Compton, G. (1992). *The Reading Connection: A leadership initiative designed to change the delivery of educational services to at-risk children*. Unpublished doctoral dissertation, Western Michigan University.
- Coomes, P. (1985). *The effects of computer assisted instruction on the development of reading and language skills*. Unpublished doctoral dissertation, North Texas State University.
- Cooper, H. (1998). *Synthesizing research* (3rd ed.). Thousand Oaks, CA: Sage.
- Correnti, R. (2009). Examining CSR program effects on student achievement: Causal explanation through examination of implementation rates and student mobility. In *Paper presented at the annual meetings of the Society for Research on Educational Effectiveness*, Crystal City, VA.
- Curry, J., Griffith, J., & Williams, H. (1995). *Reading Recovery in AISD*. Austin Independent School District, Department of Audit and Evaluation.
- Davis, C. E. (1995). *Comparing Direct Instruction reading with a basal reading program in relation to achievement, attitude toward school, and self-concept*. Unpublished doctoral dissertation, University of Southern Mississippi.
- Dorval, B., Wallach, L., & Wallach, M. A. (1978). Field evaluation of a tutorial reading program emphasizing phoneme identification skills. *The Reading Teacher*, 31, 784–790.
- Dynarski, M., Agodini, R., Heavise, 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.
- Ehri, L., Dreyer, L., Flugman, B., & Gross, A. I. (2007). Reading Rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, 44(2), 414–448.
- Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, 92, 605–619.
- Eldredge, J. L., Quinn, D., & Quinn, D. (1988). Increasing reading performance of low-achieving second graders with dyad reading groups. *Journal of Educational Research*, 82(1).
- Escamilla, K. (1997). Describiendo la lectura: An early intervention literacy program in Spanish. In S. Swartz, & A. Klein (Eds.), *Research in Reading Recovery* (pp. 109–121). Portsmouth, NH: Heinemann.
- Evans, T. L. P. (1996). *I can read deze books: A qualitative comparison of the Reading Recovery Program and a small-group reading intervention*. Unpublished doctoral dissertation, Auburn University.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 92–128.
- Gersten, R., Compton, D., Connor, C., Cimino, J., Santoro, L., Linan-Thompson, S., et al. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades. A practice guide (NCEE 2009-4045)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides>
- Gottshall, D. L. (2007). *Gottshall Early Reading Intervention: A phonics based approach to enhance the achievement of low performing, rural, first grade boys*. Unpublished doctoral dissertation, University of North Texas.
- Greene, J. (2001). Systematic phonology: The critical element in teaching reading and language to dyslexics. In C. McIntyre, & J. Pickering (Eds.), *Clinical Studies of Multisensory Structured Language Education for Students with Dyslexia and Related Disorders* (pp. 173–188). Dallas, TX: International Multisensory Structured Language Education Council.
- Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and non-Hispanic students. *Journal of Special Education*, 39(2), 66–85.
- Haring, N. G., & Krug, D. A. (1975). Evaluation of a program of systematic instruction procedures for extremely poor retarded children. *American Journal of Mental Retardation*, 79(6), 627–631.
- Harrison, C. (2000). Reading research in the United Kingdom. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research (Vol. III)*. Mahwah, NJ: Erlbaum.
- Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, 65, 41–57.
- Heistad, D. (2005). *A Minneapolis study of the effects of Read Naturally on fluency and reading comprehension: A supplemental service intervention*. Minneapolis: Minneapolis Public Schools.
- Hempenstall, K. (2008). Corrective Reading: An evidence-based remedial reading intervention. *Australasian Journal of Special Education*, 32(1), 23–54.
- Huggins, R. (1999 April). *Longitudinal study of the Reading Recovery program, 1994–1998*. Detroit Public Schools, Office of Research, Evaluation and Assessment.



- Hurry, J., & Holliman, A. (2009). *The impact of Reading Recovery three years after intervention*. Retrieved June 3, 2010 from [www.everychildareader.org/pub/index.cfm](http://www.everychildareader.org/pub/index.cfm)
- Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading*, 30(3), 227–248.
- Jenkins, J. R., Peyton, J. A., Sanders, E. A., & Vadasy, P. F. (2004). Effects of reading decodable texts in supplemental first-grade tutoring. *Scientific Studies of Reading*, 8, 53–85.
- Kamil, M., Mosenthal, P., Pearson, P., & Barr, R. (Eds.). (2000). *Handbook of reading research (Vol. III)*. Mahwah, NJ: Erlbaum.
- Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say*. SRI Project Number P10446.001. Arlington, VA: SRI International.
- LaShell, L. (1986). *An analysis of the effects of reading methods upon reading achievement and locus of control when individual reading style is matched for learning-disabled students*. Doctoral Dissertation, Fielding University.
- Lesnick, J. (2006). *A mixed-method multi-level randomized evaluation of the implementation and impact of an audio-assisted reading program for struggling readers*. Unpublished doctoral dissertation, University of Pennsylvania.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Lovett, M. W., Lacerenza, L., Borden, S., Frijters, J., Steinbach, K., & De Palma, M. (2000). Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology*, 92(2), 263.
- Lovett, M., Frijters, J., Steinbach, K., Temple, M., Benson, N., et al. (2008). Interventions for reading difficulties: A comparison of Response to Intervention by ELL and EFL struggling readers. *Journal of Learning Disabilities*, 41(4), 333–352.
- 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.
- 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.
- Mantzicopoulos, P., Morrison, D., Stone, E., & Setrakian, W. (1992). Use of the search/teach tutoring approach with middle-class students at risk for reading failure. *The Elementary School Journal*, 92, 573–586.
- 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.
- Mathes, P. G., & 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. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). 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., & Allor, J. H. (2001). The effects of Peer Assisted Learning Strategies for first-grade readers with and without additional computer assisted instruction in phonological awareness. *American Educational Research Journal*, 38, 371–410.
- Mathes, P. G., Torgesen, J., 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. *The Elementary School Journal*, 103(5), 459–479.
- Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, 40(2), 148–182.
- Mayfield, L. G. (2000). *The effects of structured one-on-one tutoring in sight word recognition of first-grade students at-risk for reading failure*. Unpublished doctoral dissertation, Louisiana Tech University.
- McGrady, D. S. (1984). The effect of programmed tutoring upon the reading comprehension of fourth-grade students enrolled in a Chapter 1 reading program. In *Paper presented at the annual meeting of the Association for Educational Communications and Technology*, Dallas, TX.
- Meier, J., & Invernizzi, M. (2001). Book buddies in the Bronx: Testing a model for America Reads. *Journal of Education for Students Placed at Risk*, 6, 319–333.
- Mooney, P. J. (2003). *An investigation of the effects of a comprehensive reading intervention on the beginning reading skills of first graders at risk for emotional and behavioral disorders*. Ph. D. Dissertation, The University of Nebraska – Lincoln.
- Morris, D., Tyner, B., & Perney, J. (2000). Early Steps: Replicating the effects of a first-grade reading intervention program. *Journal of Educational Psychology*, 92(4), 681–693.
- Morrow-Howell, N., Jonson-Reid, M., McCrary, S., Lee, Y., & Spitznagel, E. (2009). *Evaluation of Experience Corps: Student reading outcomes (CSD Research Report 09-01)*. Washington University, Center for Social Development, St. Louis, MO.
- National Assessment of Educational Progress. (2007). *The nation's report card*. Washington, DC: National Center for Education Statistics.
- Nelson, J., & Stage, S. (2007). Fostering the development of vocabulary knowledge and reading comprehension through contextually-based multiple meaning vocabulary instruction. *Education and Treatment of Children*, 30(1), 1–22.
- 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.
- Pinnell, G., DeFord, D., & Lyons, C. (1988). *Reading Recovery: Early intervention for at-risk first graders*. Arlington: ERS.
- Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 29(1), 8–39.
- Policy Exchange. (2009). *Every Child a Reader: An example of how top-down education reforms make matters worse*. Retrieved March 24, 2009, at [www.policyexchange.org.uk](http://www.policyexchange.org.uk)
- Pullen, P., Lane, H., & Monaghan, M. (2004). Effects of a volunteer tutoring model on the early literacy development of struggling first grade students. *Reading Research and Instruction*, 43(4), 21–40.
- Ramey, M. (1991). *Compensatory education sustained gains from spring 1988 to spring 1990*. Seattle, WA: Seattle Public Schools.
- Rimm-Kaufman, S., Kagan, J., & Byers, H. (1999). The effectiveness of adult volunteer tutoring on reading among “at risk” first grade children. *Reading Research and Instruction*, 38(2), 143–152.
- Ritter, G. (2000). *The academic impact of volunteer tutoring in urban public elementary schools: Results of an experimental design evaluation*. Unpublished doctoral dissertation, University of Pennsylvania.
- Ritter, G., Barnett, J., Denny, G., & Albin, G. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research*, 79(1), 3–38.
- Ross, S. M., & Casey, J. (1998). *Success for All evaluation, 1997–98 Tigard-Tualatin School District*. Memphis: University of Memphis, Center for Research on Educational Policy.
- Ross, S., Smith, L., Bond, C., Casey, J., & Johnson. (1993). *Final report: 1994–94 Success for All program in Montgomery, AL*. Memphis, TN: Center for Research on Educational Policy.
- Ross, S. M., Smith, L. J., & Casey, J. (1994). *Final report: 1993–94 Caldwell, Idaho. SFA results*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.
- Ross, S. M., Smith, L. J., & Casey, J. (1995). *Final report: 1994–95 Success for All program in Fort Wayne*. Memphis, IN: University of Memphis, Center for Research in 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.
- 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.

- Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a “scientifically-based” reading program. *Economics of Education Review*, 23(4), 323–338.
- Santa, C. M., & Høien, T. (1999). An assessment of early steps: A program for early intervention of reading problems. *Reading Research Quarterly*, 34, 54–79.
- Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, 97(2), 257–267.
- 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.
- Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly*, 30, 958–995.
- Sinkis, D. (1993). *A Comparison of Chapter One student achievement with and without computer-assisted instruction*. Unpublished doctoral dissertation, University of Massachusetts.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15(9), 5–11.
- Slavin, R. (2008). What works? Issues in synthesizing education program evaluations. *Educational Researcher*, 37(1), 5–14.
- Slavin, R. E., & Madden, N. A. (in press). Bias due to measures inherent to treatments in systematic reviews in education. *Journal of Research on Educational Effectiveness*.
- Slavin, R. E., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31(4), 500–506.
- 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, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.
- 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., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, 79(4), 1391–1465.
- Slavin, R. E., Madden, N. A., Chambers, B., & Haxby, B. (2009). *Two million children: Success for All*. Thousand Oaks, CA: Corwin.
- Smith, L. J., Ross, S. M., & Casey, J. (1994). *Final report: 1993–94 Success For All program in Memphis*. Tennessee, Memphis, TN: University of Memphis, Center for Research on Educational Policy.
- Snow, C. E., Burns, S. M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- 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).
- Stevens, R. J., & Slavin, R. E. (1995). The cooperative elementary school: Effects on students’ achievement, attitudes and social relations. *American Educational Research Journal*, 32(2), 321–351.
- Stevens, R., & Slavin, R. E. (1995). Effects of a cooperative learning approach in reading and writing on academically handicapped and non-handicapped students. *The Elementary School Journal*, 95(3), 241–261.
- Stevens, R., Van Meter, P., Garner, J., Warcholak, N., Bochna, C., & Hall, T. (2008). Reading and Integrated Literacy Strategies (RAILS): An integrated approach to early reading. *Journal of Education for Students Placed at Risk*, 13(4), 357–380.
- Taylor, B. M., Short, R. A., Frye, B. J., & Shearer, B. A. (1992). Classroom teachers prevent reading failure among low-achieving first-grade students. *Reading Teacher*, 45(8), 592–597.
- 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*. London: DfES Research Report, p. 711.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading*, 1, 217–234.
- Torgesen, J. K., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., et al. (2006). *National Assessment of Title I interim report: Volume II: Closing the reading gap: First year findings from a randomized trial of four reading interventions for striving readers*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.
- Torgesen, J., Schirm, A., Castner, L., Vartivarian, S., Mansfield, W., Myers, D., et al. (2007). *National assessment of Title I final report. Volume II: Closing the reading gap: Findings from a randomized trial of four reading interventions for striving readers*. Washington, DC: IES.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2009). *Computer assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches*. Tallahassee, FL: Florida State University.
- Vadasy, P. F., & Sanders, E. A. (2008). Repeated reading intervention: Outcomes and interactions with readers’ skills and classroom instruction. *Journal of Educational Psychology*, 100(2), 272–290.
- Vadasy, P. F., Sanders, E. A., & Peyton, J. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring: How text matters. *Journal of Learning Disabilities*, 38(4), 364–380.
- Vadasy, P. F., Sanders, E. A., & Tudor, S. (2007). Effectiveness of paraeducator-supplemented individual instruction: Beyond basic decoding skills. *Journal of Learning Disabilities*, 40(6), 508–525.
- Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryant, D. P., Dickson, S., & Blozis, S. A. (2003). Reading instruction groups for students with reading difficulties. *Remedial and Special Education*, 24(5), 301–315.
- Vernon-Feagans, L., Amendum, S., Kainz, K., Ginsberg, M., Wood, T., & Bock, A. (2009 March). The Targeted Reading Intervention (TRI): A classroom teacher Tier 2 intervention to help struggling readers in early elementary school. In *Paper presented at the annual meetings of the Society for Research on Effective Education*, Crystal City, VA.
- Vernon-Feagans, L., Kainz, K., Hedrick, A., Ginsberg, M., & Amendum, S. (2010 March). The Targeted Reading Intervention: A classroom teacher professional development program to promote effective teaching for struggling readers in kindergarten and first grade. In *Paper presented at the annual meetings of the Society for Research on Effective Education*, Washington, DC.
- Wang, C., & Algozzine, B. (2008, June 1). Effects of targeted intervention on early literacy skills of at-risk students. *Journal of Research in Childhood Education*, 22(4), 425–439.
- Wasik, B. A. (1997). Volunteer tutoring programs: Do we know what works? *Phi Delta Kappan*, 79(4), 282–287.
- Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 28, 178–200.
- What Works Clearinghouse. (2009). *Beginning reading. What Works Clearinghouse Topic Report*. Retrieved March 24, 2009, from <http://ies.ed.gov/ncee/wwc/>
- Williams, H., Curry, J., & Griffith, J. (1995). *Reading recovery in AISD*. Austin Independent School District, Department of Audit and Evaluation.
- Ysseldyke, J. E., & Taylor, B. M. (2007). Understanding the factors that allegedly contribute to students’ reading difficulties. In B. M. Taylor, & J. E. Ysseldyke (Eds.), *Effective instruction for struggling readers K-6*. New York: Teachers College Press.