A Synthesis of Quantitative Research on Programs For Struggling Readers in Elementary Schools

Amanda J. Neitzel

Cynthia Lake

Johns Hopkins University

Marta Pellegrini

University of Florence, Italy

Robert Slavin

Johns Hopkins University

Reading Research Quarterly (in press)

Abstract

This article reviews research on the outcomes of diverse reading programs on the achievement of struggling readers in elementary schools. Sixty-five studies of 51 different programs met rigorous standards. Eighty-three percent were randomized experiments and 17% quasi-experiments. Outcomes were positive for one-to-one tutoring and were positive but not as large for one-to-small group tutoring. There were no differences in outcomes between teachers and teaching assistants as tutors. Whole-class approaches (mostly cooperative learning) and whole-class/whole-school approaches incorporating tutoring for struggling readers obtained outcomes for struggling readers as large as those found for all forms of tutoring, on average, and benefitted many more students. Technology-supported adaptive instruction did not have significant positive outcomes for struggling readers, however. In agreement with previous reviews, this synthesis found that substantial impacts can be obtained for struggling readers with interventions aligned within a response to intervention network.

A Synthesis of Quantitative Research on Programs for Struggling Readers in Elementary Schools

Recent policy shifts in favor of encouraging schools to use proven programs offer a new impetus for solving reading problems using evidence. These evidence-based reforms include policies and practices in which educators and policymakers use evidence of effectiveness as a criterion for choosing education programs, products, and practices. This movement has been supported by the passage of the Every Student Succeeds Act (ESSA) of 2015 (U.S. Congress, 2015), which defines strong, moderate, and promising levels of evidence. The law encourages schools and districts to use evidence to help them decide which programs to select. In particular, very low achieving schools seeking federal school improvement funds must use programs that meet one of the top three ESSA standards. The ESSA evidence standards could make evidence a determining factor in how schools make decisions about what is implemented in their schools. This has potential consequences for the many struggling readers in our schools. If schools opted to use evidence-based interventions (those with rigorous research to support their efficacy), they could have meaningful positive impacts on students not currently reading at grade level.

Prior Reviews

There have been several previous reviews of research on the outcomes of programs for struggling readers. A series of eight reviews has been published by a group at the University of Texas, including Jeanne Wanzek, Sharon Vaughn, and their colleagues (Austin et al., 2017; Scammacca et al., 2015; Vaughn et al., 2009; Wanzek et al., 2007, 2013a, 2013b, 2016, 2018, 2019). These reviews used very similar methods, breaking the literature into studies of K-3 Tier 3 interventions, 4+ Tier 3 interventions, K-3 Tier 2 interventions, 4+ Tier 2 interventions, interventions of less than 100 hours, and interventions of more than 100 hours. All the reviews

included tests of the moderating impacts of substantive factors, such as group size, duration, grade level, and the nature of the implementer (e.g., teacher vs. paraprofessional). Across all of these reviews, the findings were very similar. Tier 2 and Tier 3 interventions were found to be quite effective, and there were few substantive or methodological factors that significantly mediated these effects. Effects were stronger on alphabetics than on comprehension, and were stronger on researcher-made tests than on standardized tests.

A group at the University of Munich (Galuschka, Ise, Krick, & Schulte-Korne, 2014) reviewed effects of various treatment approaches for children and adolescents with reading disabilities. In this case, the focus was mostly on content of instruction rather than structure. The review found that interventions emphasizing phonics are the only ones that produce significantly positive effects.

Slavin, Lake, Davis, & Madden (2011) reviewed research on programs for struggling readers in elementary schools, grades K-5. Unlike all other reviews, this one included effects of Tier 1 approaches (i.e., whole-class teaching) on outcomes for struggling readers, and placed a great deal of emphasis on the evidence supporting specific programs, not just overall effects of interventions. This review reported strong, positive outcomes of one-to-one and small-group instruction, and of tutoring by teaching assistants as well as certified teachers, in common with the conclusions of the Wanzek/Vaughn and Gersten et al. (2020) reviews. However, this review also reported positive effects of whole-class approaches emphasizing cooperative learning, and of multi-tier approaches combining whole-class instruction for all and tutoring for struggling readers. It also reported small and non-significant positive impacts for computer-assisted instruction.

The most recent review in this area was written by Gersten et al. (2020). This review, written under a contract from the U.S. Department of Education, used the inclusion criteria of the What Works Clearinghouse (2014), and included certified WWC reviewers on its authorship team. It focused on Tier 2 interventions for grades 1 to 3, explicitly excluding kindergarten as being of limited relevance.

Like the Wanzek/Vaughn series of meta-analyses, Gersten at al. (2020) chose to focus on overall effects, rather than individual programs. They found a substantial positive impact of qualifying interventions (ES=+0.39). Also like the Wanzek/Vaughn reviews, Gersten et al. failed to find significant effects of moderators, except for outcome measures (outcomes were stronger for alphabetics than for comprehension, for example). There were no significant differences for grade level, research design, one-to-one versus small-group settings, or teachers vs. paraprofessionals.

Need for the Current Review

The Wanzek/Vaughn series of reviews, and the recent Gersten et al. (2020) review, have established that Tier 2 and Tier 3 interventions are effective in accelerating the achievement of struggling readers. This result is satisfying to those who are seeking evidence for the effectiveness of RTI/MTSS. However, for educators and policy makers, this is not enough. They need more specific guidance about what particular interventions are likely to make a maximum difference for their students. To do this, it is important to look more precisely into the evidence bases for alternative programs and program types. This, in turn, requires confidence that each study of each program provides valid evidence for the programs.

Other than the Slavin et al. (2011) meta-analysis, the main prior reviews of research on programs for elementary struggling readers, the Wanzek/Vaughn series and Gersten et al.,

(2020), were designed to look for overall impacts of Tier 2 and Tier 3 interventions, but were not designed to investigate effects of particular programs or categories of programs. In fact, none of these reviews even mentioned the names of programs being evaluated, nor grouped the data to suggest which specific programs were particularly effective. Also, these meta-analyses included many studies that would not be appropriate in a review focused on the effectiveness of particular programs. For example, both the Wanzek/Vaughn series and Gersten et al. included studies in which the researcher or research team provided the tutoring (e.g., Case et al, 2010), and outcome measures that were over-aligned with the experimental treatment, or in which the researchers or developers created the outcome measures (e.g., Schwartz, 2005). These reviews also included studies in which tutors themselves administered the posttest assessments (e.g., Schwartz, 2005), and studies with very small sample sizes, as low as 21 (Vadasy, Sanders, & Peyton, 2006b, Study 2). When looking at the relative effectiveness of alternative programs, these factors become extremely important, as, for example, small sample size and use of developer- or researcher-made measures are known to greatly inflate effect sizes in meta-analyses (see Cheung & Slavin, 2016; deBoer et al., 2014). A program could appear to be particularly effective not because it is likely to excel in practical use, but because it happened to have been evaluated in a very small study or used a measure made by developers or researchers, or one closely aligned with the experiment treatment.

In addition to meta-analyses of research, the What Works Clearinghouse reviews reading programs, using inclusion criteria similar to those of the present article. The WWC program reviews are very helpful, but the WWC does not integrate the programs into a meta-analysis, which means that it does not provide information on categories of programs or on moderators of

program impacts. Also, the WWC is not up to date; its "Find What Works" literacy section was last updated in 2013, and its most recent tutoring article was published in 2013.

The Wanzek/Vaughn series of reviews, the Gersten et al. review, and the Galuschka et al. review are all focused on variables, not programs. This is a valid focus for some purposes, bur for practice, it is not enough. School leaders and teachers often want to know which specific programs have strong evidence of effectiveness. The Every Student Succeeds Act (ESSA), passed by the U.S. Congress in 2015, defines "strong," "moderate," and "promising" evidence of effectiveness for programs, not variables, and makes certain funding available for schools to adopt programs that meet these standards. The What Works Clearinghouse and England's Education Endowment Foundation (EEF) also define effectiveness in terms of programs, not variables. Also, tutoring is being widely proposed as a solution for student learning losses due to school closures in the 2020 Covid-19 pandemic. England and the Netherlands have begun large programs to provide such tutoring, and tutoring is part of post-COVID services in several U.S. states. To ensure that these efforts pay off, educators will need information on specific proven programs to make informed decisions.

The present review is designed to significantly update the review of research on effective programs published by Slavin et al. (2011). Because of the many studies that have appeared since that time, the present review uses much tougher inclusion standards than those that could have been used in 2011, and as a result, many of the articles accepted in 2011 were not accepted in the current review. Also, the current review uses modern forms of meta-analysis and meta-regression that were not in wide use in 2011, but are rapidly becoming expected (see Tipton et al., 2019). The present review does produce information on variables, both on categories of interventions and key moderators, as in the Wanzek/Vaughn and Gersten et al. reviews, but its

focus on programs provides a clearer guide to educators on what was actually implement in successful studies, and gives them information on practical means of implementing the findings of research.

Conceptual Framework

The conceptual framework for the present review builds on response to intervention (RTI) (Fuchs & Fuchs, 2006; Fuchs, Mock, Morgan, & Young, 2003). In recent years, the RTI framework has been extended to behavioral interventions, using the collective term multi-tier systems of support (MTSS; Brown-Chidsey & Bickford, 2015). RTI (or MTSS as applied to academic learning), has an initial focus on Tier 1, prevention within the general education class. Students who are not meeting pre-established standards may then be considered for more intensive services, either interventions of moderate intensity (Tier 2), or if problems are very serious, intensive interventions (Tier 3).

The present review does not take a position on how RTI policies are used in practice, or on outcomes of RTI. Such outcomes are very difficult to determine, because in a nationally mandated program, it is impossible to identify control groups not using RTI (see Balu et al, 2015; Fuchs & Fuchs, 2017). However, this review uses RTI as a familiar framework to organize evidence on programs for struggling readers in elementary schools. As a framework, RTI provides a means of categorizing reading interventions as prevention (Tier 1), moderate intensity intervention (Tier 2), or intensive intervention (Tier 3). This categorization is widely used in previous U.S. reviews, such as those by Gersten et al. (2020), and Wanzek et al. (2016). Using the same RTI/MTSS framework enables comparisons across these reviews, as well as individual studies that also place themselves in the RTI/MTSS framework. Also, RTI/MTSS remains U.S. policy, and the 2015 Every Student Succeeds Act (ESSA) strongly reinforced the federal

commitment to RTI/MTSS policies, so organizing the present review within an RTI/MTSS framework enhances the review's relevance to policy.

Methods

Inclusion Criteria

We developed inclusion criteria based on the ESSA evidence standards, prior reviews, as well as accepted standards for synthesizing rigorous research (e.g., Borenstein et al., 2009; Lipsey & Wilson, 2001; Valentine, Hedges, & Cooper, 2019). The present synthesis used inclusion standards similar to those of the What Works Clearinghouse (WWC, 2020), except where otherwise noted. The inclusion criteria are described below:

1. Studies had to evaluate a reading program, defined as a specific, replicable school day/school year approach combining materials, software, assessments, professional development, and other elements designed to improve the reading achievement of struggling readers.
2. Students in included studies had to be struggling readers, variously defined in studies as students reading below grade level or reading in the lowest half, third, or quarter of their grade. Students with IEPs and English learners were included if they were in regular reading classes.
3. Students in included studies had to be in kindergarten or grades one to five, or six if sixth graders were in elementary schools.
4. Studies needed to include a similar comparison group of children receiving the typical instruction that would have occurred without the intervention ("business as usual"). Control students may have received remedial or supplementary services if these were

what the school would have routinely offered. In studies in which struggling readers were selected from within diverse classes, selection had to use the same processes in both treatments (e.g., lowest quarter), and the pretests of the selected groups had to be equivalent.

5. Studies were sought in countries whose education systems are most similar to those in the U.S. These include North America, Europe, Israel, Australia, and New Zealand. However, the reports had to be available in English. In practice, 59 qualifying studies took place in the U.S., plus one each in the Republic of Ireland, the U.K., and Sweden.
6. The studies had to use either random assignment or quasi-experimental (matched) methods, with adjustments for pretest differences. The level of assignment could be schools, teachers, or students.
7. Studies had to demonstrate baseline equivalence between groups based on the analytic sample (after attrition). Studies with pretest differences greater than 0.25 SD were excluded. Also, studies with differential attrition of 15% or more were excluded. These criteria are essentially the same as those used by the WWC (2020).
8. Studies' outcome measures needed to be independent, quantitative measures of reading, such as individually administered or group-administered reading assessments created independently of the program. Experimenter- or developer-made measures and measures aligned to experimental but not control groups were excluded. WWC (2020) standards exclude "overaligned" measures, but not all measures made by developers or researchers. The rationale for our exclusion of such measures is that experimenter- and developer-made measures have been shown to have substantially inflated effect sizes as compared to independent measures (Cheung & Slavin, 2016; de Boer et al., 2014).

9. Individually administered measures given by the teachers who taught the program were not accepted, as such measures have potential for bias.
10. Studies needed to have a minimum duration of 12 weeks from pretest to posttest. Cheung & Slavin (2016) found that studies with brief durations have inflated effect sizes, perhaps because in brief experiments, researchers can provide unrealistic levels of support that could not be maintained for a semester or more. For example, in the Wanzek et al. (2016) review, 11 of 72 included studies provided 10 hours of instruction or less, yet these studies had the second-highest effect size among five categories of dosage. This standard differs from that of the WWC, which does not specify minimum durations.
11. Studies had to evaluate programs delivered by teachers or teaching assistants, not research staff. If programs were delivered by research staff (e.g., Case et al., 2010), studies were excluded. The WWC standards do not mention this issue.
12. Studies needed to include at least 30 students and two tutors in each treatment condition. Evaluations of programs with very small sample sizes produce greatly inflated effect sizes (Cheung & Slavin, 2016). The WWC requires at least two teachers or schools per condition, but does not have a minimum number of students.
13. Studies must have been published between 1990 and 2020, with the exception of technology programs, which must have been published between 2000 and 2020 (because technology has changed so much since the 1990s).

Literature Search Procedure

A broad literature search was conducted to locate as many studies as possible that might meet the inclusion criteria. Electronic searches were made of educational databases including JSTOR, ERIC, EBSCO, PsycInfo, and Dissertations Abstracts International. Various

combinations of key words, such as “reading,” “primary students,” “struggling readers,” and “tutoring” were used to identify studies.

Google Scholar and other internet search engines and educational publisher websites were also searched. Because of the known difference between the effect sizes of published and unpublished research (Cheung & Slavin, 2016; Polanin et al., 2016), particular efforts were made to locate this "gray literature".

Studies were also sought among those reviewed by the WWC, as well as those reported by i3, IES, the Education Endowment Foundation (EEF) in England, and other funders and research organizations. Citations from identified studies and previous reviews of reading interventions were examined for possible inclusion. We also conducted hand searches of the tables of contents of recent issues of key journals. This process continued until no new studies were found. This process is summarized in a PRISMA diagram, shown in Figure 1.

Review and coding procedures

Once potential studies were identified, a three-stage process was followed that screened studies for relevance, reviewed the studies against the inclusion criteria, and finally extracted details of the included studies by coding them into a standardized database.

During the screening stage, a first review of each study was conducted by a single reviewer, who examined the title and abstract using the inclusion criteria. This process was completed using Covidence, an online platform for systematic review. In initial screening, studies were excluded if their titles and abstracts made it clear that the study was not relevant. If there was any possibility that a study might meet standards, it was retained for a more thorough second stage review.

Studies that were not eliminated in the screening phase were located and the full text read by one of the study team members in the second stage of the process. The study was assessed against the inclusion criteria, to determine if it met all necessary requirements. All the studies retained were examined by a second author to confirm that they met the inclusion criteria.

Studies that met the inclusion criteria were coded by one of the study team members. Codes were verified by another study team member. Data to be coded beyond outcome measures, sample sizes, and effect sizes included substantive factors, methodological factors, and extrinsic factors (Lipsey, 2019). These are described below. A complete set of data are available at Neitzel, Lake, Pellegrini, & Slavin, 2020.

Substantive factors. Substantive factors describe the intervention, population, and context of the study. These coded factors included category of intervention (multi-tier whole-class/whole-school approach, whole-class Tier 1 approach, technology-supported adaptive instruction, tutoring), duration of intervention, student grade level, and population description (race, ethnicity, special education status, and free/reduced price meals status). For tutoring studies, additional factors were coded, including frequency and length of tutoring sessions, whether this was extra time for literacy (replacement or supplement to literacy instruction), group size, and provider (teacher, paraprofessional, paid volunteer, or unpaid volunteer).

Methodological factors. Methodological factors included the research design (randomized or quasi-experimental design), the level of assignment (student or cluster), and the type of outcome. Outcomes were categorized into four groups: general reading, fluency, comprehension, or alphabetics (WWC, 2014). Alphabetics included subskills of reading such as letter identification and phonics outcomes, fluency included reading accuracy and reading with expression, comprehension included understanding meaning of text, and general reading

included outcomes combining other categories, such as standardized reading tests providing a "total" score.

Extrinsic factors. Extrinsic factors are those which may not directly influence the results, but may be related to the outcomes. Extrinsic factors coded were publication status and year of publication.

Effect size and variance calculation

The effect sizes of interest in this study are standardized mean differences. These are effect sizes that quantify the difference between the treatment and control group on outcome measures, divided by standard deviations. This allows the magnitude of impacts to be compared across interventions and outcome measures.

In meta-analysis models, studies were weighted to give more weight to studies with the greatest precision (Hedges, Tipton, & Johnson, 2010). In practice, this primarily involves weighting for sample size. Weights for each study were calculated according to the following formula:

$$W_j = \frac{1}{k_j(\bar{v}_j + \tau^2)}$$

where W_j is the weight for study j, k_j is the number of findings in study j, $\bar{v}_j$ is the average finding-level variance for study j, and τ^2 is the between-study variance in the study-average effect sizes (Hedges et al., 2010; Tipton, 2015). Variance estimates were adjusted for clustering as described by Hedges (2007).

Meta-regression

We used a multivariate meta-regression model with robust variance estimation (RVE) to conduct the meta-analysis (Hedges et al., 2010). First, our data included multiple effect sizes per study, and RVE accounts for this dependence without requiring knowledge of the covariance

structure (Hedges et al., 2010). Second, this approach allows for moderators to be added to the meta-regression model and calculates the statistical significance of each moderator in explaining variation in the effect sizes (Hedges et al., 2010). We estimated two meta-regression models. First, we estimated a null model to produce the average effect size without adjusting for any covariates. Second, we estimated a meta-regression model with the identified moderators of interest and covariates. This model took the general form:

$$T_{ij} = \beta_0 + \beta_k X_{ij} + \beta_m X_j + \eta_j + \varphi_{ij} + \varepsilon_{ij}$$

where T_{ij} is the effect size estimate i in study j, β_0 is the grand mean effect size for all studies, β_k is a vector of regression coefficients for the covariates at the effect size level, X_{ij} is a vector of covariates at the effect size level, β_m is a vector of regression coefficients at the study level, and X_j is a vector of covariates at the study level, η_j is the study-specific random effect, and φ_{ij} is the effect size specific random effect. The X_{ij} and X_j included substantive, methodological, and extrinsic factors, as outlined above. All moderators and covariates were grand-mean centered to facilitate interpretation of the intercept. All reported mean effect sizes come from this meta-regression model, which adjusts for potential moderators and covariates. The packages *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2020) were used to estimate all random-effects models with RVE in the R statistical software (R Core Team, 2020).

Program Categories

Qualifying studies were organized into five categories according to main program features. In several cases, categories were further subdivided. The categories and their corresponding results and tables were presented according to response to intervention tiers: Whole-class and whole-school approaches (Tier 1), multi-tier (combining Tier 1 and Tiers 2/3),

technology-supported adaptive instruction (Tier 2), and tutoring, both one-to-small group (Tier 2) and one-to-one (Tier 3).

Whole-class and whole-school programs. Two categories emphasize prevention approaches to improving reading among struggling readers. Both involve improving the effectiveness of core reading instruction for all students, whether or not they are known to be struggling readers (although in this review, only the performance of the struggling readers is analyzed). In an RTI framework, this is the essence of Tier 1 (Fuchs & Fuchs, 2006). In concept, the advantage of such approaches is that entire classes or schools can receive cost-effective interventions. Students who are at risk may succeed in initial instruction, likely reducing the need for instructional intervention or special education, and students' actual performance, rather than tests alone, inform each student's need for supplemental services.

Multi-tier whole-class/whole school approaches. One category of preventive programs combines professional development in proven whole-class strategies with tutoring for students who need it. The classroom programs emphasize cooperative learning, in which students work in small teams to help each other learn, a strong emphasis on phonics in the early grades, or (usually) both cooperative learning and phonics. In addition, those students who are struggling in reading receive either small-group tutoring as a Tier 2 intervention, or one-to-one tutoring as a Tier 3 intervention. These interventions are closely linked to the classroom approaches.

The theory behind the use of cooperative learning to improve the reading skills of struggling readers depends on two major factors (see Slavin, 2017). One is motivational; cooperative learning in small teams trying to achieve team goals provides students with incentives to learn and to help their teammates learn, to help themselves and teammates achieve success as a team. The second factor is peer teaching, in which students working with teammates

have opportunities to learn from each other and teach each other (Webb, 2008). Peer work is engaging, social, and exciting to most students. Peer teaching draws on Vygotsky's (1978) theories, which emphasize the gains students can make from interacting with peers operating in their zone of proximal development.

The theory behind an emphasis on phonics in beginning reading depends on evidence that most students struggling to learn to read need a systematic approach to learning alphabet sounds and fluently combining them to form words (Shaywitz & Shaywitz, 2020). Phonics is a key element of the "five pillars of reading" emphasized by the National Reading Panel (2000) and the National Research Council review (Snow et al., 1998).

Two programs fell in the multi-tier category. Success for All (Cheung et al., in press), evaluated in three qualifying studies, is a whole-school program whose instructional programs emphasize cooperative learning and phonics. Students in grades 1-3 who are struggling in reading may receive one-to-one or one-to-small group instruction. Another program, Enhanced Core Reading Instruction, evaluated in one study, focuses only on first grade. Students receive whole-class teaching emphasizing phonics, and those who need assistance receive one-to-small group pre-teaching closely aligned with upcoming class instruction.

Whole-class Tier 1 approaches. A second category consists of programs that use classroom methods very similar to those used in multi-tier models, emphasizing cooperative learning, phonics, or both. However, unlike multi-tier models, these Tier 1 approaches do not link to any specific tutoring programs, and their evaluations focus only on the classroom teaching methods, not tutoring or other Tier 2 or Tier 3 interventions. The theoretical bases for the classroom programs are the same as those for the multi-tier models.

The whole-class Tier 1 category included three programs. Cooperative Integrated Reading and Composition (CIRC) emphasizes cooperative learning, phonics, and meta-cognitive comprehension approaches (Stevens et al., 1987). An adapted form of CIRC is used as the upper-elementary component of Success for All, but does not incorporate the use of tutoring. PALS is another program that emphasizes cooperative learning and phonics in first grade. Ladders to Literacy focuses on phonics teaching in kindergarten.

Technology-supported adaptive instruction (Tier 2). Technology is used in education both as whole-class instruction (Tier 1) and as intervention, for students who are struggling in reading (Tier 2). We found several studies of applications of technology as intervention, but no whole-class technology approaches that met inclusion standards, provided whole-class technology-supported instruction, and reported outcomes for a subgroup consisting of struggling readers. For this reason, technology applications appear in this review only as Tier 2 interventions for groups composed entirely of struggling readers.

Most remedial reading technology interventions (e.g., System 44 and SuccessMaker) primarily involve computer-assisted instruction, in which students work at their own levels (assessed by the computers) on exercises designed to build their reading skills in a step-by-step fashion. The exercises mostly consist of questions, ending with a summative unit test. If students pass the unit test, they are advanced to the next set of exercises, building gradually toward well-specified objectives. If students make errors, the software is likely to offer prompts or hints, and may provide further questions at the same level before advancing to the next skills in a sequence. As students obtain a certain number of correct answers, they may receive symbolic rewards of some kind, such as the ability to earn points, prizes, or fun activities on the computer. Computer-

assisted instruction (CAI) programs typically provide teachers with detailed data on student progress and success rates, and teachers may use these data to modify assignments.

Many CAI programs cycle students through multiple types of activities. In addition to work on computers, students may receive instruction from teachers and/or work in small groups. Elementary studies of technology approaches in reading have generally found positive but modest impacts (Cheung & Slavin, 2012; Campuzano et al., 2009; Dynarski et al., 2007).

Among technology studies meeting the standards of this review, there was one major exception to this description. It was a study of a program called New Heights (Lesnick, 2006), in which students choose a text at their level. After a teacher introduction, students receive audio tapes that repeatedly read the text to them, until they can pass fluency and comprehension tests.

One-to-small group tutoring (Tier 2). The largest number of programs evaluated with struggling readers have involved tutoring. We categorized one-to-small-group tutoring as Tier 2, and one-to-one tutoring as Tier 3.

One-to-small group tutoring is typically supplemental instruction (usually 30 minutes a day for a semester or more) provided to groups of from 2 to 6 students at a time. Typically, students in one-to-small group tutoring are all reading at the same level. Tutors may be certified teachers or teaching assistants (paraprofessionals), usually with extensive training. Small group tutors serve many more students each day than do one-to-one tutors, of course. For example, in seven half-hour sessions each day, a one-to-four tutoring model would serve 28 students daily, while the same tutor working one-to-one would serve only seven.

The theoretical basis for all tutoring is straightforward. Tutoring is expected to enhance student reading achievement by providing instruction adapted to students' needs to help them through a planned sequence of skills known to be essential in reading: phonemic awareness,

phonics, comprehension, fluency, and vocabulary (Elbaum, Vaughn, Hughes, & Moody, 2000; National Reading Panel, 2000; Shaywitz & Shaywitz, 2020; Snow et al., 1998). Well-trained tutors with structured materials are able to assess students' current reading levels, diagnose reading deficits, deliver instruction building on students' strengths and needs, provide immediate feedback to students, and constantly assess students' progress and adjust their teaching to consider each student's unique progression.

In addition to individualizing instruction, tutoring creates an environment in which one tutor and one or a small group of children can establish strong relationships. Tutors can provide a great deal of personal attention and praise craved by all students, but perhaps especially by those who have not been successful in school.

One-to-One Tutoring (Tier 3). Previous reviews (Austin et al., 2017; Slavin et al., 2011; Elbaum et al., 2000; Gersten et al., 2020; Wanzek et al., 2007, 2013a, 2013b, 2016, 2018, 2019), have concluded that one-to-one tutoring is the most effective intervention known to improve the reading achievement of struggling students in elementary school. The problem with one-to-one tutoring is that it is resource intensive, especially if the tutors are certified teachers. As a result, few students can typically be served in a one-to-one format, and if students need tutoring over longer-than-usual time periods to achieve success, they are unlikely to receive it. Because of the cost of one-to-one tutoring and the limits it places on the numbers of struggling students it can serve, tutoring approaches are increasingly being designed to use well-trained teaching assistants (about half the salary cost of certified teachers), "paid volunteers" such as AmeriCorps members*, unpaid volunteers, and one-to-small group tutoring. All of these variations are discussed in this article, and effectiveness of each are examined.

* AmeriCorps is a U.S. program that trains volunteers to serve their communities, often as tutors. They receive stipends and other benefits.

Resolving Disagreements

All decisions about study inclusion, categorization, or methodological features, were reviewed by at least two authors. If there were disagreements after thorough review by the initial pair of reviewers, additional authors were consulted, and discussions were continued until all authors reached consensus.

Nature of Schools and Students

In all included studies, the students in the study samples were below-level readers (see inclusion criterion #1). In the lists of included studies in Tables 3 to 6B, sample descriptions are provided.

Limitations

No single review can cover all issues in a field as diverse as reading. Each must justify a purpose and audience, and then make inclusion criteria that correspond to them.

The present review focuses on specific, replicable programs that school leaders might choose to implement in order to improve the achievement of struggling readers. The review does not include studies focused on variables, such as phonetic vs. non-phonetic emphasis (but see Galuschka et al. 2014, which does make this comparison). All programs are composed of variables, but each program is a specific combination of variables that could be replicated as such, with particular materials, software, and professional development. The restriction to evaluations of programs, not variables, is consistent with the policies of the What Works Clearinghouse (2020).

This synthesis did not review programs delivered outside of the school day and year, such as after school or summer programs. These can be important resources for struggling readers, but they are complex to evaluate and compare to school-day/year programs. For reviews of research

on summer school programs, see Allington & McGill-Franzen, 2018; and Kim & Quinn, 2013. For reviews of research on after-school programs, see Chappell et al. (2011) and Muñoz, Chang, & Ross (2012).

The present review excludes outcome measures made by program developers or researchers, because use of such measures have been found to greatly inflate effect size estimates, in comparison to use of independent measures (Cheung & Slavin, 2016; deBoer et al., 2014). Findings of developer/researcher-made measures can be useful for theory building and implementation monitoring, but educators need to know the effects of programs on independent measures.

This review focuses only on quantitative research. Qualitative research plays an important role in theory building and in understanding of schools, students, and society, but it is less helpful in asking the "what works" questions at the heart of this review.

Results

A total of 65 studies of 51 interventions for elementary struggling readers met the inclusion criteria (see Table 1 for characteristics of these studies). Overall, these studies exhibited a high level of methodological rigor, with 85% (*k*=54) of the studies randomized at either the student or cluster level and only 17% (*k*=11) using quasi-experimental methods.

The full meta-regression model is shown in Table 2. For all studies, this model controlled for program type, research design, level of assignment, race/ethnicity, poverty, and outcome type. Among tutoring programs, three tutoring-specific factors were controlled: extra time, provider type, and group size.

Findings of individual studies are summarized in Tables 3 to 6B, while summaries of mean outcomes in each category are listed in Table 7. Statistical significance is shown in Tables

7 and 8 for categories of programs and moderators. Statistical power for categories and moderators is greatly affected by the number of studies included. This means that every study in a given category could have a large and significant positive result, but the category might not be significant in the meta-regression because it has too few studies.

There was an overall positive impact on reading achievement across all qualifying studies (ES=+0.23, $p<.001$). However, these outcomes vary considerably, according to the programs implemented.

Whole-Class/Whole-School Approaches.

One potential solution to the reading problems of many struggling readers is to enhance the quality of instruction in their regular classrooms. Slavin, Lake, Chambers, Cheung, & Davis (2009) found that classroom instructional programs emphasizing professional development in cooperative learning, phonics, and other classroom strategies were the most effective approaches for children in general. Slavin et al. (2011) reported that these types of programs were also found to be very effective for struggling readers in grades K to 5. As noted earlier, there were two categories of programs of this type: Multi-tier whole-class/whole-school approaches, and whole-class Tier 1 instruction.

Multi-Tier Whole-Class/Whole-School Approaches

As shown in Table 3, the overall mean effect size for the four studies of multi-tier whole class/whole school approaches was +0.27 ($p<.10$). This was one of the largest category effect sizes in this review, but it was not statistically significant in the meta-regression because of the small number of studies.

Whole-Class Tier 1 Approaches

Table 4 shows a mean effect size of +0.31 (p=.11) for five studies of whole class Tier I approaches. This was the largest category effect size, except for one-to-one tutoring, but it was not statistically significant in the meta-regression due to too few studies.

Technology-Supported Adaptive Instruction (Tier 2)

Over the past 30 years, one of the most common solutions provided to children who are struggling to learn to read is CAI software. Modern CAI programs adapt to children's specific needs and give them activities with graphics and motivational elements that can supplement whole-class teaching with individualized instruction. However, previous reviews of research on elementary CAI applications in reading have reported only modest positive effects (Cheung & Slavin, 2012; Campuzano et al., 2009; Dynarski et al., 2007).

Table 5 summarizes research on outcomes of interventions using technology to support adaptive instruction for students who are at the lowest performance levels of their classes. Across eight studies of seven programs, the mean effect size was +0.09 (n.s.).

Small Group Tutoring (Tier 2)

Twenty-three studies of one-to-small group tutoring, summarized in Table 6A, had a weighted mean effect size of +0.24 (p<.001). These analyses were broken down according to who provided the tutoring and by group size (one-to-one or small group) in the following sections.

Small group tutoring by teachers. A common form of intervention for struggling readers is additional teaching in small groups (2-6 students), typically 30-45 minutes daily. Small-group tutorials are potentially more cost-effective than one-to-one tutoring from teachers, because several children are taught at the same time. The mean effect size for nineteen studies of small group tutoring by teachers was +0.21. See Table 6A, upper panel.

Small group tutoring by teaching assistants. Small group tutorials for struggling readers can also be provided by teaching assistants. The mean effect size for four studies of small group tutoring by teaching assistants was +0.27, similar to that for teachers as tutors. See Table 6A, lower panel.

One-to-One Tutoring (Tier 3)

Forty-six studies of one-to-one tutoring, summarized in Table 6B, had a mean effect size of +0.41 (p<.001). These outcomes are broken down below according to who provided the tutoring.

One-to-one tutoring by teachers. One-to-one instruction provided by certified teachers is one of the most popular choices for struggling readers. While this approach can be resource-intensive, it may be cost-effective in the long-term if these students see a substantial improvement in their reading proficiency that leads to fewer referrals to special education, less remediation, and fewer students retained. The mean effect size for 14 studies of one-to-one tutoring by teachers was +0.38. See Table 6B, top panel.

One-to-one tutoring by teaching assistants. One-to-one tutoring by certified teachers is expensive, so many schools have long used teaching assistants as tutors, usually with materials specifically designed for this purpose. The mean effect size for four studies of one-to-one tutoring by teaching assistants was +0.44, essentially identical to outcomes for programs using teachers as tutors. Note that in most cases teaching assistants were relatively well-qualified, usually with bachelor's degrees. See Table 6B, second panel.

One-to-one tutoring by paid volunteers. Another common way to provide tutoring to students in need of additional support is through the use of paid volunteers, such as AmeriCorps members (defined previously) and volunteers whose time is donated by their employers. The

mean effect size for three studies of one-to-one tutoring by paid volunteers was +0.46, similar to effects for teaching assistants. See Table 6B, third panel.

One-to-one tutoring by unpaid volunteers. Unpaid volunteers are also often used in tutoring approaches for struggling readers. The mean effect size for four studies of one-to-one tutoring by unpaid volunteers was +0.14, much lower than the effect sizes for tutoring by teachers, teaching assistants, or paid volunteers. See Table 6B, fourth panel.

Substantive and Methodological Moderators

Several important demographic and methodological moderators were identified and explored statistically (see Table 8). Age/grade level was strongly associated with intervention category (older students were much more likely to participate in educational technology interventions, younger students were much more likely to participate in tutoring), and for this reason, grade level could not be used as a moderator. All remaining moderators are reported below.

Research design. Differences in effect sizes between studies that used randomized designs (*k*=54, ES=+0.24) and studies that used quasi-experimental designs incorporating matching (*k*=11, ES=+0.31) were tested. This difference was not statistically significant. Studies using cluster-level assignment (*k*=20, ES=+0.25) had the same average effect size as studies using student-level assignment (*k*=45, ES=+0.25).

Race/ethnicity. Average effect sizes for studies with mainly White populations (k=14, ES=+0.12) were significantly lower than the average for mixed populations (k=38, ES=+0.28). African American (k=8, ES=+0.31) and Hispanic (k=5, ES=+0.30) students had effect sizes similar to those of studies with mixed populations.

Poverty. Student poverty was not a significant moderator, with studies of mainly high poverty students (k=31, ES=+0.20) and studies of mainly low poverty students (k=34, ES=+0.29) having similar outcomes.

Outcome type. Differences in effect sizes across outcome types were also statistically examined. The mean effect size across studies with general reading outcomes was +0.14. This contrasted with mean effect sizes across alphabetics outcomes (ES=+0.28) and fluency outcomes (ES=+0.27). These differences were statistically significant ($p<.05$). Outcomes for comprehension (ES=+0.21) were higher but not significantly different from those for general reading outcomes.

Tutoring-Specific Moderators

Studies were included in random effects models to explore three crosscutting features that varied across sub-categories within tutoring studies (see Tables 2 and 7).

Tutoring: One-to-one vs one-to-small-group. Tutoring can be delivered one-to-one or to small groups of two to six students at a time. When comparing the 25 studies of one-to-one tutoring with the 23 studies of small-group tutoring, we found that one-to-one tutoring had a mean effect size of +0.41 ($p<.001$), while one-to-small group tutoring had an effect size of +0.24 ($p<.001$). The difference between one-to-one and one-to-small group outcomes was statistically significant ($p<.05$).

Tutoring: Provider. Tutoring can be provided by certified teachers, teaching assistants, paid volunteers, or unpaid volunteers. Studies using teachers had a mean effect size of +0.34, teaching assistants had an average effect size of +0.29, paid volunteers averaged +0.36, while unpaid volunteers had an average effect size of 0.04. While the difference between teachers and

unpaid volunteers was significant ($p<.01$), the differences between teachers and teaching assistants, or paid volunteers as tutors were not significant.

Tutoring: Extra time. Tutoring can be provided to students either as a supplement to regular classroom instruction or as a replacement for classroom instruction. All tutoring studies were coded according to whether the intervention was provided outside the regular reading instructional time, so that students received extra time on reading, or whether the intervention was provided during regular reading instruction, where treatment and control students received the same amount of reading instruction. Across the 33 studies that supplemented classroom instruction with additional time for tutoring, the mean effect size was +0.28 ($p<.001$). There were somewhat larger impacts in the fourteen studies in which tutoring replaced regular reading instruction (ES=+0.29, $p<.001$). The difference between the two was not statistically significant.

Discussion

The large number of studies meeting high methodological standards, evaluating a broad range of elementary interventions for struggling readers, allows for a more informed investigation of effective programs for this population than has ever been possible. The 65 qualifying studies of 51 programs are of very high quality, with 54 using randomized experiments and 11 high-quality quasi-experiments. The rigorous inclusion standards make the findings both statistically reliable and relevant to practice and policy. The 23 programs meeting ESSA evidence standards for “strong” and “moderate” levels of evidence indicate that educators have available many practical solutions to the problems of reading failure in elementary schools.

The findings of this review lead to several conclusions that confirm those of earlier reviews, but also several that challenge earlier conclusions. These conclusions relate to important advances in knowledge about “what works” in tutoring, about interventions for whole classes or

schools, and about interventions that integrate whole-class interventions with tutoring in a response to intervention framework. The meanings of these findings for theory, practice, and policy, are discussed in the following sections.

Studies of Tutoring

The overall weighted mean effect size for all forms of tutoring was +0.26, smaller but in line with previous reviews (e.g., Slavin et al., 2011; Gersten et al., 2020; Wanzek et al., 2016). However, the present review had available sufficient high-quality tutoring studies to enable comparisons among important categories of tutoring. The methodological quality of these tutoring studies is also very high, with 44 randomized (92%) and 4 quasi-experimental (8%) studies. The number, diversity, and quality of these studies allows for some important observations about what works in tutoring for struggling readers.

The most practically important finding is that teachers were no more effective as tutors than were teaching assistants. In fact, in both one-to-one tutoring and one-to-small group tutoring, tutors who were teaching assistants obtained slightly higher effect sizes than teacher tutors did (+0.44 to +0.38 for one-to-one tutoring, and +0.27 to +0.21 for one-to-small group). This contradicts the conclusions of a similar review by Slavin et al. (2011), who reported larger effects for teachers than for teaching assistants as tutors, but agrees with findings of reviews by Gersten et al. (2020) and Wanzek et al. (2007, 2013a, 2013b, 2016).

How can teaching assistants achieve equal tutoring outcomes to those achieved by tutors who are certified teachers? It is important to recall that in studies using teaching assistants, most of them had bachelor's degrees or more. Also, both teacher tutors and teaching assistant tutors usually used well-structured tutoring models, with extensive training.

It is interesting to note that in a recent review of research on elementary mathematics programs, tutors who were teaching assistants also obtained non-significantly better outcomes than did teachers (Pellegrini, Neitzel, Lake, & Slavin, 2020). The same was also true of U.K. studies evaluating reading tutoring programs for struggling young adolescents (Baye, Inns, Lake, & Slavin, 2019).

Another finding relating to tutoring involved the comparison of one-to-one and one-to-small group tutoring. In this case, effect sizes for both were statistically significant, with one-to-one at ES=+0.41 and one-to-small group at ES=+0.24. Gersten et al. (2020) found similar differences, but they were not statistically significant. Wanzek et al. (2016) found no trend toward higher effects for one-to-one. Schwartz, Schmitt, & Lose (2012) randomly assigned Reading Recovery students to receive 1-1, 1-2, 1-3, or 1-5 tutoring. Reading outcomes diminished as group size increased. However, there was no non-RR control group, so it is not possible to tell whether or not students in the largest group (1-5) were still scoring better than controls.

Making Widespread Tutoring Practical

The pattern of effects of different forms of tutoring have important implications for the ability of schools to serve many students. The largest cost of tutoring by far is the cost of the tutors. If positive tutoring effects can be obtained with teaching assistants, at half the cost of certified teachers, and with small groups rather than one-to-one, schools may be able to serve many more students at the same cost.

The outcome comparisons reported above could imply that teaching assistants providing, for example, one-to-four tutoring, would require about one-eighth of the personnel cost of one-to-one tutoring by teachers [1/2 (cost of TA vs teacher) x ¼ (cost per child of 1-1 vs 1-4

tutoring]=1/8). As a practical matter, this could mean that schools could afford several teaching assistants to work with many students for the cost of a single 1-1 certified teacher tutor. Using multiple teaching assistants with small groups could enable schools to do more tutoring in grades other than first, or to provide "booster sessions" to maintain students' skills over time. Further, in areas in which there are teacher shortages (such as inner-city and rural schools), it may not be possible to find sufficient certified teacher tutors, so knowing that well-trained teaching assistants can tutor as well as certified teachers could help make tutoring for many struggling readers a practical possibility in these locations.

Even though one-to-one tutoring was found in the present study to be superior in effect size to one-to-small group tutoring, this higher impact would only apply to the students who receive the tutoring. One-to-one tutoring is very expensive, so few students, presumably those most at risk, are likely to receive it. This means that students not quite so far behind but still at risk may not receive tutoring. As an illustration of this issue (not a statistical analysis), consider effects of tutoring on groups of four, either with one of the four receiving one-to-one tutoring, or all four receiving one-to-four tutoring. Thought of this way, students tutored in a group of four might each receive an effect size of +0.27, the mean for one-to-small group tutoring. In contrast, if only one of the four students received one-to-one tutoring, which had a mean effect size of +0.44, the whole group of four would receive effect sizes of +0.44/0/0/0, an average of +0.11. One-to-one tutoring might still be justified for students with serious deficits, such as those who, in the absence of tutoring, might be assigned to special education, but one might argue that reaching more students may be more important in many cases than a higher effect size for many fewer students.

Outcomes of volunteer programs depended on whether or not volunteers were "paid." When they were, outcomes were very good, non-significantly higher than those of teachers or teaching assistants working as tutors. When volunteers were not "paid," however, outcomes were much lower. "Paying" volunteer tutors presumably leads them to attend more regularly over a longer period, allowing for more continuity and closer relationships with students (see Jacob et al., 2015 for a description of difficulties with attendance and abrupt resignations of unpaid tutors).

Tutoring and Tutor-Student Relationships

The pattern of findings relating to tutoring, and to tutoring alternatives, offer some interesting insights into how and why tutoring works for struggling readers. Several alternatives to tutoring have sought to maintain the individualization inherent to tutoring, but diminished the role of personal relationships with valued adults. The most obvious example is computer-assisted instruction. Any modern CAI program provides all of the adaptation to individual learning levels and rates provided by one-to-one tutoring. Students are assessed, placed at their current level, and then given assignments tailored to their needs. Their progress is carefully monitored, and their progress is celebrated. Yet in elementary reading, no CAI approach that met the standards of this review had significant positive effects for struggling readers. The average effect size for technology-supported adaptive instruction was only +0.09 (n.s.), a substantial contrast to the effect size of +0.34 for tutoring programs provided by teaching assistants.

Moving from tutoring by teachers to tutoring by teaching assistants, however, did not entail any loss of efficacy. While one might be concerned that teaching assistants would be less skilled than teachers as instructors, there is no reason to expect teaching assistants to be less effective than teachers in encouraging, nurturing, and praising young struggling readers. The

observation that teaching assistants obtained reading outcomes as positive as those obtained by teachers may suggest that forming relationships is a key part of the explanation for the effectiveness of tutoring.

Classroom- and School-Level Programs

Non-technology classroom- and school-level programs had mean outcomes slightly higher than the mean for all tutoring programs (the mean effect size for all tutoring was +0.26). What this means, for example, is that whole-class models such as cooperative learning might have the same effect on the struggling readers in diverse classes as would tutoring those same struggling readers. No study made a direct comparison of such alternative approaches, however, so this is speculation based on mean effect sizes. The mean effects of multi-tier whole-school/whole-class approaches and of whole-class Tier 1 approaches were not statistically significant in the meta-regression analyses due to the small numbers of studies, but at ES=+0.27 and +0.31, respectively, they show promise. Classroom programs included two cooperative learning programs, Cooperative Integrated Reading and Composition (CIRC) and PALS. There was one multi-tier whole-school approach, Success for All (SFA), and one multi-tier classroom approach for grade 1, Enhanced Core Reading Instruction (ECRI).

Explanations for the effectiveness of whole-class and whole-school methods are unique to each. Such programs may affect everything teachers do in teaching reading. Note that three of the four strategies (CIRC, PALS, and SFA) involve cooperative learning (Slavin, 2017).

Because whole-class and whole-school approaches serve all students in particular grades or in whole schools, many more struggling readers are likely to benefit from them than is likely to be possible in interventions only consisting of tutoring. Classroom and school interventions

are likely to be more cost-effective on a per-student basis than any form of tutoring, as it would be very expensive to provide tutoring to all students.

As noted previously, whole-class models may be seen as Tier 1 in a response to intervention framework, and tutoring may be seen as Tier 2 or Tier 3. The tiered whole school interventions (Success for All for grades K-5 and Enhanced Core Reading Instruction for first grade) already incorporate integrated instructional interventions (Tier 1) and tutoring (Tiers 2 and 3) as part of a coordinated RTI approach in which each tier is composed of proven instructional models.

Instructional Technology

As in previous reviews (e.g., Cheung & Slavin, 2012), there were minimal positive effects of instructional technology in reading (mean ES=+0.08). The same was true in the Baye et al. (2019) review of secondary reading programs, and it was true in a review of elementary mathematics approaches (Pellegrini et al., 2020). It may be that in elementary reading, the individualization provided by technology is not necessary for many students and not sufficient for the lowest achievers, who may instead need forms of tutoring.

Ensuring Long-Term Benefits for Struggling Readers

There is not enough evidence from long-term follow-up studies to permit definitive conclusions, but the limited evidence that does exist raises some important questions. May et al. (2016) found no lasting impacts of Reading Recovery on reading after first grade. While Pinnell et al. (1994) found positive effects of Reading Recovery on independent measures in the first half of first grade, by fall of second grade effects remained only on developer-made measures. Blachman et al. (2014) reported a ten-year follow-up of Intensive Reading Remediation, finding no lasting impact on reading comprehension. In contrast, a three-year follow-up of Success for

All, a whole-school multi-tier approach, found lasting impacts of an intervention ending in 5th grade on standardized test scores through eighth grade, as well as substantial reductions in special education placements and retentions (Borman & Hewes, 2002).

Educational programming for struggling readers needs to focus on each child's entire trajectory from preschool to secondary school. Some students may only need a one-time boost from tutoring, but it is not realistic to assume this will be enough for all students. Creating an ecology of core reading instruction with appropriate supports ready when needed, exactly the approach advocated (but rarely fully implemented) in response to intervention (Fuchs & Fuchs, 2006), seems more likely to succeed with a larger number of struggling readers.

Conclusion

The research reviewed in this article provides strong support for approaches to addressing the needs of struggling readers:

1. The largest number of studies evaluated forms of adult-to-child tutoring. Our findings of large positive effects of tutoring correspond to those of previous reviewers (e.g., Gersten et al., 2020; Wanzek et al., 2007, 2013a, 2013b, 2016, 2018, 2019). In agreement with these prior reviews, we also found that effects of tutoring by teachers, teaching assistants, and paid volunteers, were comparable to each other in student outcomes, although effects for unpaid volunteers were much smaller. While one-to-one tutoring was superior in reading outcomes to one-to-small group tutoring (ES=+0.41), one-to-small group tutoring still demonstrated meaningful positive outcomes (ES=+0.24). These findings have great practical importance, as tutoring by teaching assistants and in small groups are far less resource intensive than one-to-one

tutoring by certified teachers, which has dominated tutoring for struggling readers in the past.

2. Effect sizes for three forms of whole-class Tier 1 instruction using cooperative learning and an emphasis on phonics teaching, as well as two multi-tier programs that provide both Tier 1 whole-school instruction emphasizing cooperative learning and phonics, and aligned Tiers 2 and 3 tutoring for students who need them, had important positive outcomes for struggling readers, similar to the average impacts of tutoring. This finding agrees with an earlier review (Slavin et al., 2011), but other reviews have not assessed Tier 1 outcomes for struggling readers in comparison to Tiers 2 and 3 interventions.
3. Technology approaches focused on adapting instruction to meet the needs of struggling readers had small, non-significant impacts on these students' reading achievement.

Taking these findings together, the data from 65 rigorous evaluations of programs for struggling readers support the expanded use of all forms of tutoring, as well as whole-class approaches emphasizing cooperative learning and phonics. Approaches to response to intervention that coordinate Tier 1, Tier 2, and Tier 3 instruction around proven whole-school, whole-class, and tutoring approaches, show particular promise.

From a pragmatic standpoint, the positive effects of tutoring are beyond dispute, but there is a need to even further enhance the effectiveness of tutoring and to reduce the cost, the main barrier to widespread tutoring. Can effective tutoring be devised for groups of one to six or one to eight? Also, can effective Tier 1 instruction solve the problems of large numbers of struggling

readers, so that a smaller number of students may need expensive one-to-small group or one-to-one tutoring?

Another area in need of study relates to long-term maintenance of effects of tutoring and other proven interventions. Long-term effects have only inconsistently been documented. It may be that following initial success in tutoring, students may be more likely to build on gains if they are given brief "booster shot" tutoring each year, to keep them up to grade level, or it may be that using multi-tier, coordinated strategies combining whole-class instruction with tutoring as needed may produce long-term, sustained impacts.

In the recent Covid-19 crisis, millions of students had to receive instruction online due to school closures. The need for development and evaluation of distance tutoring to individuals and small groups became apparent. Even when schools are fully open, there may be benefits of online tutoring, which can be provided to students in school or at home, after school, or during the summer.

Perhaps the most important finding of this review is the demonstration, consistent with all other reviews, that struggling readers can make substantial gains if provided with proven interventions. This review adds evidence on effects of specific programs, and includes evidence on effective Tier 1 approaches, as well as investigating less costly means of ensuring the success of students at risk. But the consistent and unchallenged findings of substantial impacts of tutoring and other interventions indicate that struggling readers can be doing much better than they are doing today. To the extent that control groups in all of these studies represent the current state of practice, we know schools can do much better with struggling readers, and this should have important implications for policy and practice.

References

**Studies included in the meta-analysis.*

Allington, R. L., & McGill-Frazen A. (Eds.). (2018) *Summer reading: Closing the rich/poor reading achievement gap*. New York, NY: Teachers College.

*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.

*Amendum, S. J., Vernon-Feagans, L., & Ginsberg, M. C. (2011). The effectiveness of a technologically facilitated classroom-based early reading intervention: The Targeted Reading Intervention. *The Elementary School Journal*, *112*(1), 107–131.

Austin, C. R., Vaughn, S., & McClelland, A. M. (2017). Intensive reading interventions for inadequate responders in grades K–3: A synthesis. *Learning Disability Quarterly*, *40*(4), 191–210. https://doi.org/10.1177/0731948717714446

*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), 494–519. https://doi.org/10.1598/RRQ.35.4.3

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., Gersten, R., … MDRC. (2015). *Evaluation of response to intervention practices for elementary school reading. NCEE 2016-4000*.

Baye, A., Inns, A., Lake, C., & Slavin, R. E. (2019). A synthesis of quantitative research on reading programs for secondary students. *Reading Research Quarterly, 54* (2), 133-166. https://doi.org/10.1002/rrq.229

*Beam, M., Faddis, B. J., & Hahn, K. (2011). *Evaluation of System 44 (California), Final report*. Portland, OR: RMC Research Corporation.

*Beam, M., Faddis, B., & Hahn, K. (2012). *Evaluation of System 44 (Saginaw, MI), Final report*. Portland, OR: RMC Research Corporation.

*Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of Intensive Reading Remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology*, *96*(3), 444–461. (2004-18154-004).

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive Reading Remediation in grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology*, *106*(1), 46–57.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd.

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

Borman, G. D., & Hewes, G. M. (2002). The long-term effects and cost-effectiveness of Success for All. *Educational Evaluation and Policy Analysis*, *24*(4), 243–266. https://doi.org/10.3102/01623737024004243

Brown-Chidsey R. & Bickford, R. (2015). *Practical handbook of multi-tiered systems of support.* New York: Guilford Press.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts* (No. NCEE 2009-4041). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Case, L. P., Speece, D. L., Silverman, R., Ritchey, K. D., Schatschneider, C., Cooper, D. H., Montanaro, E., & Jacobs, D. (2010). Validation of a supplemental reading intervention for first-grade children. *Journal of Learning Disabilities*, *43*(5), 402–417. https://doi.org/10.1177/0022219409355475

Chappell, S., Nunnery, J., Pribesh, S., & Hager, J. (2011). A meta-analysis of supplemental education services provider effects on student achievement. *Journal of Education for Students Placed at Risk, 16*(1), 1–23.

Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review, 7*(3), 198–215. https://doi.org/10.1016/j.edurev.2012.05.002

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

Cheung, A., Xie, C., Zhang, T., Neitzel, A., & Slavin, R. E. (in press). Success for All: A quantitative synthesis of evaluations. *Journal of Research on Educational Effectiveness.*

*Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M., … Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness, 6*(1), 1–23.

de Boer, H., Donker, A. S., & van der Werf, M. P. C.. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, *84*(4), 509–545.

*Denton, C. A., Fletcher, J. M., Taylor, W. P., Barth, A. E., & Vaughn, S. (2014). An experimental evaluation of guided reading and explicit interventions for primary-grade students at-risk for reading difficulties. *Journal of Research on Educational Effectiveness*, *7*(3), 268–293.

*Denton, C. A., Nimon, K., Mathes, P. G., Swanson, E. A., Kethley, C., Kurz, T. B., & Shih, M. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children*, *76*(4), 394–416.

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

*Eddy, R. M., Ruitman, H. T., Hankel, N., Matelski, M. H., & Schmalstig, M. (2011). *Pearson Words Their Way: Word study in action: Intervention efficacy study final report*. La Verne, CA: Cobblestone Applied Research.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, *79*(1), 262–300. https://doi.org/10.3102/0034654308325998

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (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*(4), 605-619.

*Fives, A., Kearns, N., Devaney, C., Canavan, J., Russell, D., Lyons, R., … O'Brien, A. (2013). A one-to-one programme for at-risk readers delivered by older adult volunteers. *Review of Education, 1*(3), 254–280. https://doi.org/10.1002/rev3.3016

*Fogarty, M., Coyne, M. D., Simmons, L. E., Simmons, D. C., Henri, M., Kwok, O.-M., Ware, S. M., Dalton, K., Williams, K. A., & Wang, H. (2020). Effects of technology-mediated vocabulary intervention for third-grade students with reading difficulties. Journal of Research on Educational Effectiveness, 13(2), 271–297. https://doi.org/10.1080/19345747.2019.1698086

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*(1), 93–99.

Fuchs, D., & Fuchs, L. S. (2017). Critique of the National Evaluation of Response to Intervention: A case for simpler frameworks. *Exceptional Children*, *83*(3), 255–268. https://doi.org/10.1177/0014402917693580

Fuchs, D., Fuchs, L. S., Thompson, A., Otaiba, S. A., Yen, L., Yang, N. J., Braun, M., & O'Connor, R. E. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*(2), 251–267.

*Fuchs, D., Fuchs, L. S., Thompson, A., Otaiba, S. A., Yen, L., Yang, N. J., … O'Connor, R. E. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*(2), 251–267.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, *18*(3), 157-171.

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE, 9*(2), e89900.

*Gatti, G. (2013). *Pearson SuccessMaker response to intervention study: Final report*. Pittsburgh, PA: Gatti Evaluation Inc.

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayantha, M. (2020). Meta-analysis of the impact of reading interventions for students in the primary grades. *Journal of Research on Educational Effectiveness, 13* (2), 401-427.

*Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *The Journal of Special Education, 34*(2), 90–103. https://doi.org/10.1177/002246690003400204

*Hamilton, J., Gray-Adams, K., Chen, E., Gorga, C. P., McKithen, C., & Glatz, A. von. (2016). iRead impact study: Final report. Westat.

*Hanselman, P., & Borman, G. D. (2013). The impacts of Success For All on reading achievement in grades 3–5: Does intervening during the later elementary grades produce

the same benefits as intervening early? *Educational Evaluation and Policy Analysis*, *35*(2), 237–251.

*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*(1), 41–57.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341–370. https://doi.org/10.3102/1076998606298043

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

*Huggins, R. (1999). *Longitudinal study of the Reading Recovery program, 1994 - 1998*. Detroit, MI: Detroit Public Schools.

Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading*, *30*(3), 227–248. https://doi.org/10.1111/j.1467-9817.2007.00338.x

*Jacob, R. T., Armstrong, C., & Willard, J. A. (2015). *Mobilizing volunteer tutors to improve student literacy: Implementation, impacts, and costs of the Reading Partners program*. New York: MDRC. http://files.eric.ed.gov/fulltext/ED558508.pdf

*Jones, C. J. (2015). *The results of a randomized control trial evaluation of the SPARK literacy program*. Milwaukee, WI: Socially Responsible Evaluation in Education.

Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research, 83*(3),386-431. Doi:10.3102/0034654313483906

*Lee, Y. S., Morrow-Howell, N., Jonson-Reid, M., & McCrary, S. (2012). The effect of the Experience Corps® program on student reading outcomes. *Education and Urban Society*, *44*(1), 97–118.

*Lesnick, J. (2006). *A mixed-method multi-level randomized evaluation of the implementation and impact of an audio-assisted reading program for struggling readers* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3211103)

Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology*, *107*(2), 558–579.

Lipsey, M. W. (2019). Identifying potentially interesting variables and analysis opportunities. In *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 141–151). Russell Sage Foundation.

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

*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*(1), 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*(5), 573–586.

*Mathes, P. G., Torgesen, J. K., & Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted

instruction in phonological awareness. *American Educational Research Journal*, *38*(2), 371.

*May, H., Sirinides, P. M., Gray, A., & Goldsworthy, H. (2016). *Reading Recovery: An evaluation of the four-year i3 scale-up*. Philadelphia: Consortium for Policy Research in Education.

*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.

Muñoz, M., Chang, F., & Ross, S. (2012). No Child Left Behind and tutoring in reading and mathematics: Impact of supplemental educational services on large-scale assessment. *Journal of Education for Students Placed at Risk, 17*(3), 186–200.

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

Neitzel, A. Lake, C., Pellegrini, M., & Slavin, R.E. (2020). Data archive for "A Synthesis of Quantitative Research on Programs For Struggling Readers in Elementary Schools." Towson, MD: Center for Research and Reform in Education (CRRE), Johns Hopkins University. Retrieved from https://github.com/aj-neitzel/A-Synthesis-of-Quantitative-Research-on-Programs-For-Struggling-Readers-in-Elementary-Schools

*Pappas, S., York, A., Wang, Y., & Richards, K. (2015). *Examining the efficacy of Burst:Reading literacy intervention*.

Pellegrini, M., Neitzel, A., Lake, C., & Slavin, R. (2020). *Effective programs in elementary mathematics: A best-evidence synthesis*. Available at www.bestevidence.com. Manuscript submitted for publication.

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research, 90* (1), 24-46.

*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), 9–39.

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, *86*(1), 207–236. https://doi.org/10.3102/0034654315582067

Pustejovsky, J. (2020). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections* (Version R package version 0.4.1) [Computer software]. https://CRAN.R-project.org/package=clubSandwich

*Quint, J., Zhu, P., Balu, R., Rappaport, S., & DeLaurentis, M. (2015). *Scaling up the Success for All model of school reform: Final report from the Investing in Innovation (i3) evaluation*. New York, NY: MDRC.

R Core Team. (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

*Ransford-Kaldon, C. R., Flynt, E. S., Ross, C. L., Franceschini, L., Zoblotsky, T., Huang, Y., & Gallagher, B. (2010). *Implementation of effective intervention: An empirical study to evaluate the efficacy of Fountas & Pinnell's Leveled Literacy Intervention System (LLI) (2009-2010)*. Retrieved from Center for Research in Educational Policy (CREP) website: http://eric.ed.gov/?id=ED544374

*Ransford-Kaldon, C. R., Ross, C. L., Lee, C. C., Sutton Flynt, E., Franceschini, L. A., & Zoblotsky, T. A. (2013). *Efficacy of the Leveled Literacy Intervention System for K–2*

urban students: An empirical evaluation of LLI in Denver Public Schools. Memphis, TN: Center for Research in Educational Policy.

*Ritter, G. W., & Maynard, R. A. (2008). Using the right design to get the "wrong" answer? Results of a random assignment evaluation of a volunteer tutoring programme. *Journal of Children's Services*, *3*(2), 4–16.

*Ross, S., & Casey, J. (1998). *Longitudinal study of student literacy achievement in different Title I school-wide programs in Fort Wayne community schools. Year 2: First grade results*. Memphis, TN: Center for Research in Educational Policy.

*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.

Samson, J. F., Hines, S. J., & Li, K. (2015). Effective use of paraprofessionals as early intervention reading tutors in grades K-3. *Mentoring & Tutoring: Partnership in Learning*, *23*(2), 164–177. https://doi.org/10.1080/13611267.2015.1049014

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, *48*(4), 369–390. https://doi.org/10.1177/0022219413504995

*Scanlon, D. M., Vellutino, F. R., Small, S. G., Fanuele, D. P., & Sweeney, J. M. (2005). Severe reading difficulties—Can they be prevented? A comparison of prevention and intervention approaches. *Exceptionality*, *13*(4), 209–227. https://doi.org/10.1207/s15327035ex1304_3

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. https://doi.org/10.1037/0022-0663.97.2.257

Schwartz, R. M., Schmitt, M. C., & Lose, M. K. (2012). Effects of teacher-student ratio in response to intervention approaches. *Elementary School Journal, 112* (4), 547-567.

Shaywitz, S. E., & Shaywitz, J. (2020). *Overcoming dyslexia (2nd ed.).* New York: Penguin Random House.

*Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L., Johnson, C., … Crevecoeur, Y. C. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergartener's response. *Exceptional Children, 77*(2), 207–228.

Slavin, R. E., Lake, C., Chambers, B., Cheung, A. C., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, *79*(4), 1391–1466.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*(1), 1–26.

Slavin, R. E. (2017). Instruction based on cooperative learning. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., pp. 388–404). New York, NY: Routledge.

*Smith, J. L. M., Nelson, N. J., Fien, H., Smolkowski, K., Kosty, D., & Baker, S. K. (2016). Examining the efficacy of a multitiered intervention for at-risk readers in grade 1. *The Elementary School Journal*, *116*(4), 549–573.

Snow, C. E., Burns, S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. https://doi.org/10.17226/6023

*Solari, E. J., Denton, C. A., Petscher, Y., & Haring, C. (2018). Examining the effects and feasibility of a teacher-implemented tier 1 and tier 2 intervention in word reading, fluency, and comprehension. *Journal of Research on Educational Effectiveness*, *11*(2), 163–191. https://doi.org/10.1080/19345747.2017.1375582

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative Integrated Reading and Composition: Two field experiments. *Reading Research Quarterly*, *22*, 433-454.

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

*Stevens, R. J., & Slavin, R. E. (1995b). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal*, *32*(2), 321–351.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019). A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods*, *10*(2), 161–179. https://doi.org/10.1002/jrsm.1338

Torgerson, C., Ainsworth, H., Buckley, H., Hampden-Thompson, G., Hewitt, C., Humphry, D., … Torgerson, D. (2016). *Affordable Online Maths Tuition: Evaluation report and executive summary*. London: Education Endowment Foundation.

*Torgesen, J. K., Schirm, A., Castner, L., Vartivarian, S., Mansfield, W., Myers, D., … Haan, C. (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.* (No. NCEE 2008-4013). Washington, D.C.: US Department of Education, Institute of Education Sciences.

*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*(3), 217–234.

*Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, *60*(1), 40–56.

*Torgesen, J. K., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., … Institute of Education Sciences (ED), W., DC. (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* (No. NCEE 2006-4002). Retrieved from U.S. Department of Education, Institute of Education Sciences website: http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED491144&site=ehost-live&scope=site

U.S. Congress. (2015). *Every Student Succeeds Act*. Retrieved from U.S. Congress website: http://edworkforce.house.gov/uploadedfiles/every_student_succeeds_act_-_conference_report.pdf

*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. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading*, *13*(5), 383–425.

*Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology*, *102*(4), 786–803. (2010-21221-001).

*Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, *15*(6), 471–497.

*Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006a). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. Journal of Educational Psychology, 98(3), 508–528.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006b). Paraeducator-supplemented instruction in structural analysis with text reading practice for second and third graders at risk for reading problems. *Remedial & Special Education*, *27*(6), 365–378. https://doi.org/10.1177/07419325060270060601

Valentine, J. C., Hedges, L. V., & Cooper, H. M. (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.

*Vanacore, K. P., & Hurwitz, L. B. (2020). *Impact of the Lexia Core5 reading program on students with reading and language impairments: Lexia research brief*. Lexia Learning.

*Vaughn, S., Roberts, G. J., Miciak, J., Taylor, P., & Fletcher, J. M. (2019). Efficacy of a word- and text-based intervention for students with significant reading difficulties. *Journal of Learning Disabilities, 52*(1), 31-44.

*Vernon-Feagans, L., Kainz, K., Hedrick, A., Ginsberg, M., & Amendum, S. (2013). Live webcam coaching to help early elementary classroom teachers provide effective literacy instruction for struggling readers: The Targeted Reading Intervention. *Journal of Educational Psychology*, *105*(4), 1175–1187.

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the **metafor** package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

*Wang, C., & Algozzine, B. (2008). Effects of targeted intervention on early literacy skills of at-risk students. *Journal of Research in Childhood Education, 22*(4), 425–439.

Wanzek, J., Petscher, Y., Al Otaiba, S., & Donegan, R. (2019). Retention of reading intervention effects from fourth to fifth grade for students with reading difficulties, *Reading & Writing Quarterly*, 35:3, 277-288, DOI: 10.1080/10573569.2018.1560379

*Wanzek, J., Petscher, Y., Al Otaiba, S. A., Rivas, B. K., Jones, F. G., Kent, S. C., … Mehta, P. (2017). Effects of a year long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology, 109*(8), 1103–1119. https://doi.org/10.1037/edu0000184

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities*, *51*(6), 612–624. https://doi.org/10.1177/0022219418775110

Wanzek, J. & Vaughn, S. (2007*).* Research-based implications from extensive early reading interventions. *School Psychology Review*. *36*. 541-561.

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review*, *28*(3), 551–576.

Wanzek, J., Vaughn, S., Scammacca, N. K., Metz, K., Murray, C. S., Roberts, G., & Danielson, L. (2013a). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research*, *83*(2), 163–195. https://doi.org/10.3102/0034654313477212

Wanzek, J., Vaughn, S., Scammacca, N. K., Metz, K., Murray, C. S., Roberts, G., & Danielson, L. (2013b). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research, 83*(2), 163–195. https://doi.org/10.3102/0034654313477212

Webb, N. M. (2008). Learning in small groups. In T. L. Good (Ed.), *21st Century Education: A Reference Handbook* (pp. 203-211). Los Angeles: Sage.

What Works Clearinghouse. (2014). *Review protocol for beginning reading interventions version 3.0*. Institute of Education Sciences, US Department of Education.

What Works Clearinghouse. (2020). *Procedures handbook (Version 4.1)*. Institute of Education Sciences, US Department of Education.

*Wilkerson, S. B. (2008). *A study of Pearson's My Sidewalks program: Final report*. Louisa, VA: Magnolia Consulting.

*Wolff, U. (2011). Effects of a randomised reading intervention study: An application of structural equation modelling. *Dyslexia*, *17*(4), 295–311.

Table 1. Description of included studies.

Category	Level	Overall N (%)
Total Studies (k)		65
Average Students per Study		422.71
Type	Comprehensive Tiered Approach	4 (6.2)
	Classroom Approaches	5 (7.7)
	One-to-One Tutoring	25 (38.5)
	Small Group Tutoring	23 (35.4)
	Tech-supported Adaptive Instruction	8 (12.3)
Design	Cluster quasi-experiment	8 (12.3)
	Cluster randomized	12 (18.5)
	Quasi-experiment	3 (4.6)
	Student randomized	42 (64.6)
Publication Status	Published	46 (70.8)
	Unpublished	19 (29.2)
Grades	3-5	14 (21.5)
	K-2	37 (56.9)
	Mix	14 (21.5)
Free/Reduced Lunch	0-65%	34 (52.3)
	66-100%	31 (47.7)
Race/Ethnicity	African American (66-100%)	8 (12.3)
	Hispanic (66-100%)	5 (7.7)
	White (66-100%)	14 (21.5)
	Mix (all groups < 66%)	38 (58.5)
Tutoring Extra Time	Extra time	33 (50.8)
	No extra time	15 (23.1)
	Not tutoring	17 (26.2)
Tutoring 1-1 vs. One-to-Small Group	1:1 Tutoring	25 (38.5)
	1:Small Group Tutoring	23 (35.4)
	Not Tutoring	17 (26.2)
Tutoring Provider	Not Tutoring	17 (26.2)
	Paid Volunteer	3 (4.6)
	Teaching Assistant	8 (12.3)
	Teacher	33 (50.8)
	Unpaid Volunteer	4 (6.2)
Total Effect Sizes (n)		270

Category	Level	Overall N (%)
Outcome Type	Achievement	19 (7.0)
	Alphabetics	146 (54.1)
	Comprehension	75 (27.8)
	Fluency	30 (11.1)

Table 2. Meta-regression results.

Coefficient	Reference Group	beta	SE	*t*	*df*	*p*
Null Model						
Intercept		0.23	0.03	8.89	54.55	0.000
Meta-Regression						
Intercept		0.25	0.02	10.97	34.36	0.000
Whole-school Approaches/Multi-tier	Tutoring	0.01	0.14	0.04	5.43	0.969
Whole-class		0.05	0.17	0.28	5.92	0.786
Technology-supported adaptive instruction		-0.18	0.07	-2.44	12.05	0.031
QED	Randomized	0.07	0.13	0.53	10.06	0.607
Cluster-assignment	Student-level assignment	0.00	0.07	0.04	10.35	0.973
African-American population	Mix of races and ethnicities in population	0.04	0.11	0.34	7.73	0.739
Hispanic population		0.02	0.09	0.25	6.76	0.812
White population		-0.15	0.06	-2.57	14.55	0.022
High poverty	Low poverty	-0.09	0.07	-1.29	13.60	0.217
Tutoring Extra Time	No extra time	-0.01	0.07	-0.17	11.86	0.866
Tutoring Group Size: 1:1	Small Group	0.17	0.07	2.35	13.05	0.035
Tutor Provider: Teaching Assistant	Teacher	0.06	0.11	0.53	7.37	0.612
Tutor Provider: Paid Volunteer		0.07	0.10	0.71	4.91	0.513
Tutor Provider: Unpaid Volunteer		-0.25	0.09	-2.75	9.75	0.021
Outcome Type: Alphabetics	General Reading	0.14	0.04	3.35	12.29	0.006
Outcome Type: Comprehension		0.08	0.04	1.73	14.67	0.105
Outcome Type: Fluency		0.13	0.05	2.66	15.77	0.017

Table 3: Multi-Tier Whole Class/Whole School Approaches

Study	Design	Duration	N	Grade	Sample Description	ES by Domain				Study ES
						Gen Rdg	Alpha	Comp	Fluency	
									Category Mean:	**+0.27**
Success for All									**Program Mean:**	**+0.35**
Quint et al. (2015)	CR	3 years	37 schools 759 students	K-2	57% FRL, 12% W, 18%AA, 66% H, 24% ELL		+0.20	+0.14		+0.18
Madden et al. (1993)	CQE	3 years	10 schools 342 students	1-5	Students in Baltimore 100% AA		+0.95		+0.87	+0.92
Ross & Casey (1998)	CQE	2 years	8 schools 92 students	K-1	Students in Ft. Wayne, IN 75% FRL, 45% minority		+0.42	+0.23	+0.32	+0.35
Enhanced Core Reading Instruction										
Smith et al. (2016)	CR	1 year	44 schools 749 students	1	20%ELL, 19%H, 47%FRL	+0.21	+0.25		+0.12	+0.21

Table Notes for Tables 3 to 6B: AA = African American, CQE=Cluster Quasi-Experiment, CR=Cluster Randomized, ELL = English Language Learner, FRL = Free & Reduced Lunch, H = Hispanic, NR = Not Reported, QE = Quasi-Experiment, SPED = Special Education, SR = Student Randomized, W = White, *$p < .05$

Table 4: Whole-Class/Tier 1 Approaches

Study	Design	Duration	N	Grade	Sample Description	ES by Domain				Study ES
						Gen Rdg	Alpha	Comp	Fluency	
								Category Mean:		**+0.31**
Ladders to Literacy										
Fuchs et al. (2001)	CQE	20 weeks	22 teachers 115 students	K	13% AA, 25% W, 55% FRL		+0.34			+0.34
Cooperative Integrated Reading and Competition								**Program Mean:**		**+0.11**
Stevens & Slavin (1995a)	CQE	2 years	7 schools 137 students	2-6	Mainstreamed SPED students 9%FRL, 95%W			+0.35		+0.35
Stevens & Slavin (1995b)	CQE	2 years	5 schools 76 students	2-6	Mainstreamed SPED students 10%FRL, 92%W			+0.80		+0.80
Hanselman & Borman (2013)	CRE	1 year	35 schools 2860 students	3-4	80%FRL, 80% minority, 10% SPED			-0.01		-0.01
PALS										
Mathes et al. (2001)	CQE	16 weeks	24 classes 75 students	1	Students in Southeast 47%W, 51%AA		+0.55	+0.74		+0.58

Table 5: Technology-Supported Adaptive Instruction (Tier 2)

Study	Design	Duration	N	Grade	Sample Description	ES by Domain				Study ES
						Gen Rdg	Alpha	Comp	Fluency	
								Category Mean:		**+0.09**
System 44								**Program Mean:**		**+0.07**
Beam et al. (2011)	SR	1 year	197 students	4,5	Students in California		+0.07		-0.04	+0.05
Beam et al. (2012)	SR	1 year	172 students	4,5	Students in an urban district in Michigan		+0.17		-0.15	+0.09
Fast ForWord										
Rouse & Krueger (2004)	SR	1 year	454 students	3-6	100% FRL, 66% H, 27% AA, 61% ELL	+0.05				+0.05
Successmaker										
Gatti (2013)	SR	1 year	292 students	3	32%ELL, 60% FRL, 60%H, 7%AA			+0.04		+0.04
New Heights Reading Program										
Lesnick (2006)	CR	18 weeks	59 classes 233 students	3 and 5	Philadelphia and suburban PA		-0.02	+0.07	-0.04	0.00
Lexia Core5										
Vanacore & Hurwitz (2020)	CR	1 year	5 schools 115 students	K-5	Students receiving special education services for reading difficulties	+0.23				+0.23
Vocabulators										
Fogarty et al. (2020)	SR	1 year	184 students	3	Students identified as at risk on reading comprehension 27% AA, 45% H, 18% W, 15% SPED, 26% ELL			+0.11		+0.11

Study	Design	Duration	N	Grade	Sample Description	ES by Domain				Study ES
						Gen Rdg	Alpha	Comp	Fluency	
								Category Mean:		**+0.09**
iRead										
Hamilton et al. (2016)	CQE	1 year	16 schools 138 students	1-2	Students identified as low-performing at baseline		+0.13	+0.28		+0.23

Table 6A: One-to-Small Group Tutoring (Tier 2)

Study	Design	Duration	N	Grade	Sample Description	ES by Domain				Study ES
						Gen Rdg	Alpha	Comp	Fluency	
									Category Mean:	**+0.24**
					Subcategory: Small Group Tutoring by Teachers				**Subcategory Mean:**	**+0.21**
Leveled Literacy Intervention									**Program Mean:**	**+0.17**
Ransford-Kaldon et al. (2013)	SR	1 year (14-18 weeks)	285 students	K,1,2	69% H, 15% W, 8% AA, 34% ELL, > 70% FRL	+0.10				+0.10
Ransford-Kaldon et al. (2010)	SR	1 semester (14-18 weeks)	422 students	K,1,2	37% H, 14% ELL, 34% AA, 85% FRL		+0.21		+0.27	+0.23
My Sidewalks										
Wilkerson (2008)	SR	30 weeks	278 students	2,3	7%AA, 14%H, 54%FRL, 8%ELL, 5%SPED			+0.10	-0.04	+0.05
Words Their Way										
Eddy et al. (2011)	SR	1 year (18 weeks)	257 students	2 and 4	At risk for academic failure. 47%FRL, 11%AA, 24%H, 29%ELL		+0.12	+0.01		+0.06
Spell Read										
Torgesen et al. (2006, 2007)	SR	1 year (20 weeks)	196 students	3 and 5	Schools around Pittsburgh 44%FRL, 69%W, 31%AA	-0.15	+0.23	+0.08	+0.06	+0.12
Failure Free Reading + Verbal Master										
Torgesen et al. (2006, 2007)	SR	1 year (20 weeks)	219 students	3 and 5	Schools around Pittsburgh; 44%FRL, 80%W, 20%AA	-0.18	+0.03	+0.09	+0.02	+0.02
Wilson Reading										
Torgesen et al. (2006, 2007)	SR	1 year (20 weeks)	163 students	3 and 5	Schools around Pittsburgh, 48%FRL, 56%W, 44%AA	+0.05	+0.19	+0.15	+0.07	+0.15

Corrective Reading										
Torgesen et al. (2006, 2007)	SR	1 year (20 weeks)	165 students	3 and 5	Schools around Pittsburgh 44% FRL, 81% W, 19% AA	-0.10	+0.16	+0.12	+0.18	+0.12
Early Reading Intervention							**Program Mean:**			**+0.12**
Coyne et al. (2013)	CR	1 year	48 teachers 162 students	K	60%W, 24%H, 9%AA, 15%ELL		-0.06			-0.06
Simmons et al. (2011)	CR	1 year (Sept/Oct - April/May)	57 teachers 206 students	K	19%AA, 42%H, 38%W, 12%SPED, 26%ELL		+0.34	+0.07		+0.31
Lindamood Phonemic Sequencing (LiPS)/ Auditory Discrimination in Depth (ADD)										
Torgesen et al. (2010)	SR	1 year (20 weeks)	74 students	1	Florida elementary schools. 33% minority, 35% FRL		+0.67	+0.45		+0.64
Read, Write, and Type-Small Group										
Torgesen et al. (2010)	SR	1 year (19)	73 students	1	33% minority, 35% FRL		+0.43	+0.32		+0.42
Responsive Reading Instruction										
Denton et al. (2010)	SR	1 year	422 students	1	41%W&Asian, 15%AA, 43%H		+0.41	+0.53	+0.45	+0.43
Guided Reading										
Denton et al. (2014)	SR	1 year	136 students	2	60%AA, 35%H, 95%FRL		+0.17	+0.06	+0.20	+0.14
Explicit Intervention (Sound Partners + Quick Reads + Comprehension)										
Denton et al. (2014)	SR	1 year	136 students	2	64%AA, 30%H, 88%FRL		+0.38	+0.38	+0.44	+0.39

Burst									
Pappas et al. (2015)	CR	1 year	57 clusters 4022 students	K-3	18%AA, 12%H, 49%W, 54%FRL, 25% ELL	+0.10			+0.10
Passport to Literacy									
Wanzek et al. (2017)	SR	1 year (25 weeks)	451 students	4	46% H, 13% ELL, 35% AA, 85% FRL		+0.03	+0.18	+0.12
RR Intervention									
Solari et al. (2018)	CR	17 weeks	21 clusters 98 students	1	34% AA, 76% H, 29%ELL, 94% FRL		+0.43	+0.57	+0.42	+0.46
Small Group Tutoring									
Vaughn et al. (2019)	SR	1 year (Oct - April)	252 students	4-5	34% AA, 76% H, 29%ELL, 94% FRL		+0.06	+0.06	+0.02	+0.05
					Small Group Tutoring by Teaching Assistants				**Subcategory Mean:**	**+0.27**
Quick Reads									**Program Mean:**	**+0.22**
Vadasy & Sanders (2008)	SR	15 weeks	162 students	2-3	30%W, 28%AA, 23%H, 23% ELL, 17% SPED, 75% Title I		+0.22	+0.20	+0.35	+0.27
Vadasy & Sanders (2009)	SR	15 weeks	202 students	2-3	21%AA, 28%H, 74%Title I, 29%ELL		+0.03	+0.14	+0.30	+0.18
Targeted Intervention									
Wang & Algozzine (2008)	CR	1 year	6 schools 139 students	1	80% FRL, 89% AA or H		+0.12	+0.13		+0.13
Schools and Homes in Partnership (SHIP)									
Gunn et al. (2000)	SR	2 years	195 students	K-3	Students below grade level on screening measure 62% H		+0.78	+0.35	+0.46	+0.55

Table 6B: One-to One Tutoring (Tier 3)

Study	Design	Duration	N	Grade	Sample Description	ES by Domain				Study ES
						Gen Rdg	Alpha	Comp	Fluency	
									Category Mean:	**+0.41**
					Subcategory: One-to-One Tutoring by Teachers				**Subcategory Mean:**	**+0.38**
Reading Recovery									**Program Mean:**	**+0.42**
May et al. (2016)	SR	12-20 weeks	6888 students	1	19% ELL, 13% AA, 20% H, 43% W		+0.43	+0.43		+0.43
Pinnell et al. (1994)	SR	5 months	194 students	1	74%W, 26% AA, 65% FRL	+0.50				+0.50
Huggins (1999)	QE	1 year	123 students	1	High-poverty students in Detroit			-0.06		-0.06
Targeted Reading Intervention									**Program Mean:**	**+0.50**
Vernon-Feagans et al., (2013)	CR	1 year	15 schools 272 students	k,1	Students in disadvantaged rural schools. 50% minority		+0.45	+0.48		+0.46
Amendum et al., (2011)	CR	1 year	7 schools 175 students	K, 1	Districts in the southwestern U.S.		+0.52	+0.72		+0.59
TEACH										
Mantzicopoulos et al. (1992)	SR	2 years	116 students	1-2	Middle-class children in suburban San Francisco 76%W, 8%AA, 5%H		+0.23	+0.10		+0.20

Phonetic Intervention Tutoring									
Mantzicopoulos et al. (1992)	SR	2 years	109 students	1-2	Middle-class children in suburban San Francisco 76%W, 8%AA, 5%H		+0.20	-0.11	+0.12
Intensive Reading Remediation									
Blachman et al. (2004)	SR	1 year	69 students	2,3	Students in upstate NY. 80%W, 14%AA		+0.92	+0.54	+0.80	+0.82
Early Steps/ Next Steps - Teachers									
Morris et al. (2000)	CQE	1 year	11 schools 86 students	1	High-poverty AA schools in urban TN.		+0.92	+0.80	+0.86
Variations on Tutoring									
Hatcher et al. (1994)	QE	7 months	124 students	Ages 6-7	Schools in rural Northern England.		+0.21	+0.45	+0.29
Lindamood Phonemic Sequencing (LiPS)/ Auditory Discrimination in Depth (ADD)									
Torgesen et al. (1997)	SR	2 1/2 years	65 students	K-2	50%W, 49%AA		+0.84	+0.39	+0.69
Intensive Tutorial Intervention									
Scanlon et al. (2005)	SR	1 year	93 students	1	Districts in Albany, NY area		+0.44	+0.67	+0.52
Reading and Fluency Training (RAFT)									
Wolff (2011)	SR	12 weeks	112 students	3	Rural and urban Sweden			+0.15	+0.15
Direct Instruction Skills Plan									
Pinnell et al. (1994)	SR	5 months	194 students	1	74%W, 26% AA, 65% FRL	+0.20			+0.20

One-to-One Tutoring by Teaching Assistants								**Subcategory Mean:**	**+0.44**
Sound Partners								**Program Mean:**	**+0.44**
Vadasy & Sanders (2011)	SR	1 year	187 students	1	47%ELL, 78%FRL, 20%Asian, 26%AA, 38%H		+0.21	+0.19	+0.20
Vadasy & Sanders (2010)	SR	18 weeks	148 students	K	85%FRL, 21% Asian, 24%AA, 42% H, 56% ELL		+0.63	+0.96	+0.74
Vadasy et al. (2006)	SR	18 weeks	67 students	K	Students scoring in the at-risk range on the DIBELS 87% Minority, 59% FRL, 25% ESL, 13% SPED		+0.38	+0.17	+0.48
Reading Rescue									
Ehri et al. (2007)	QE	6 months	126 students	1	Spanish-dominant students 95%FRL		+0.45	+0.22	+0.39
One-to-One Tutoring by Paid Volunteers								**Subcategory Mean:**	**+0.46**
Volunteer Tutoring									
Allor & McCathren (2004)	SR	6 months	243 students	1	94% FRL, 96%AA		+0.54	+0.10	+0.46
SPARK Literacy Program									
Jones (2015)	SR	2 years	194 students	K-2	76%AA, 15%H, 96%FRL	+0.36	+0.66		+0.51
SMART (Start Making a Reader Today)									
Baker et al. (2000)	SR	2 years	84 students	1-2	30%AA, 6%H, 47%W		+0.44	+0.38	+0.40

One-to-One Tutoring by Unpaid Volunteers									**Subcategory Mean:**	**+0.14**
Reading Partners										
Jacob et al. (2015).	SR	1 year	1166 students	2-5	65% H, 19% AA, 91% FRL, 55% ELL	+0.06	+0.11	+0.10	+0.09	+0.09
Experience Corps										
Lee et al. (2012)	SR	1 year	881 students	1-3	94%FRL, 58%AA, 36%H, 24%ELL		+0.10	+0.13		+0.11
West Philadelphia Tutoring Project										
Ritter (2000)	SR	1 year	328 students	2-5	87% FRL, 96% AA	-0.11				-0.11
Wizards of Words										
Fives et al. (2013)	SR	1 year	227 students	1-2	Schools in Dublin and Limerick		-0.09	0.00	-0.06	-0.05

Table 7. Mean Effect Sizes of Program Categories

Category	k	ES	SE	t	df	p
Multi-tier whole-class/whole-school approaches	4	+0.27	0.13	2.13	4.14	0.098
Whole-class Tier 1 Approaches	5	+0.31	0.16	1.98	4.78	0.108
Tech-supported adaptive instruction	8	+0.09	0.06	1.54	10.20	0.153
Tutoring	48	+0.26	0.03	8.36	16.37	0.000
One-to-Small Group Tutoring (Tier 2)	23	+0.24	0.04	6.09	21.78	0.000
Teachers	19	+0.21				
Teaching Assistants	4	+0.27				
One-to-One Tutoring (Tier 3)	25	+0.41	0.06	7.06	13.80	0.000
Teachers	14	+0.38				
Teaching Assistants	4	+0.44				
Paid Volunteers	3	+0.46				
Unpaid Volunteers	4	+0.14				

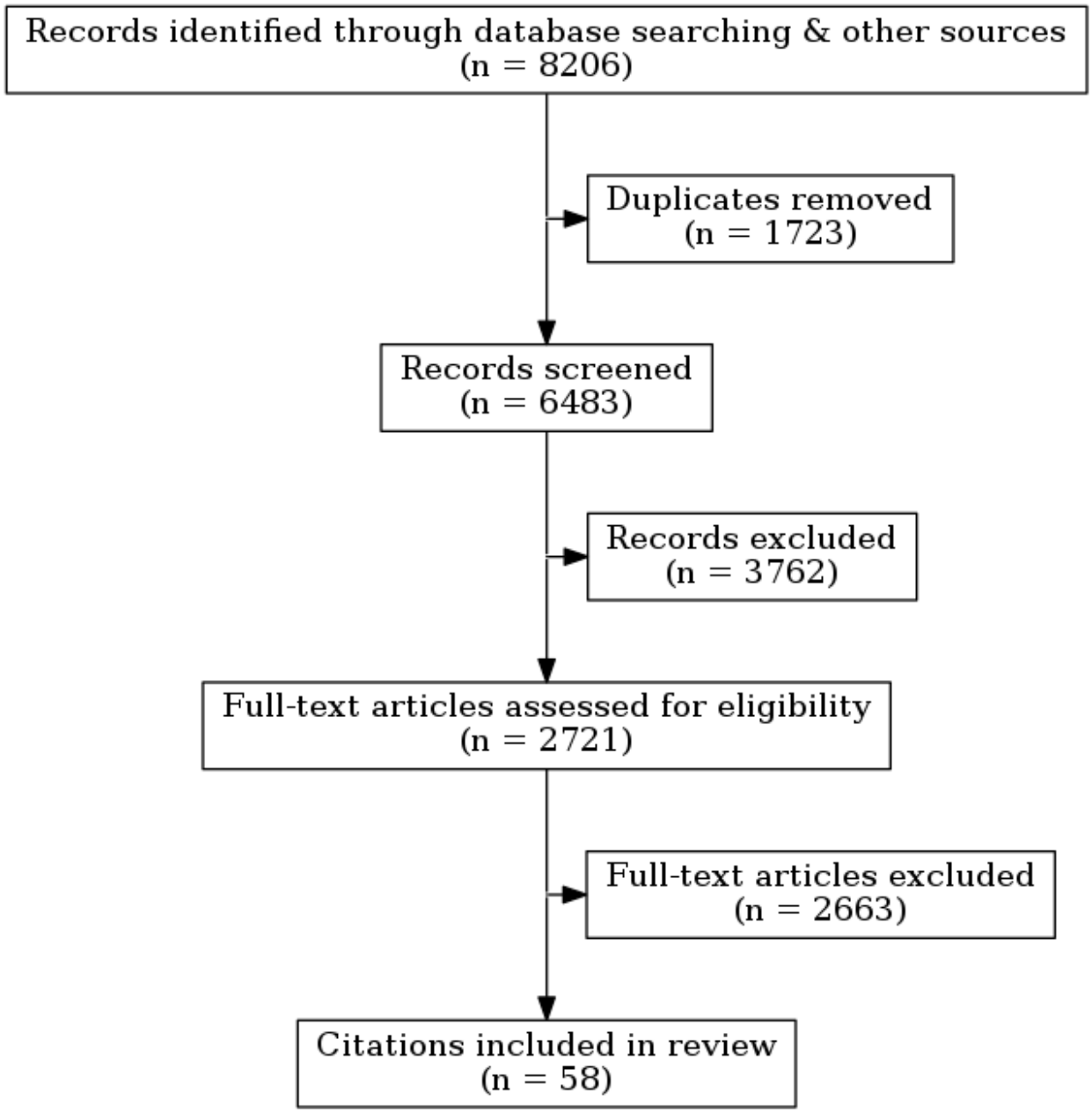
Table note: Mean effect sizes for different types of tutoring were calculated by combining the intercept (where not tutoring served as the reference category), provider of tutoring, and group size coefficients from an uncentered model including the covariates shown in Table 2.

Table 8. Substantive and methodological moderators

Moderator	Level	ES	SE	t	df	p
Experimental Design	Randomized	+0.24	0.02	9.84	28.62	0.000
	Quasi-Experimental	+0.31	0.12	2.60	9.40	0.028
Level of Assignment	Cluster-assignment	+0.25	0.06	4.18	8.68	0.003
	Student-assignment	+0.25	0.03	8.84	30.14	0.000
Race & Ethnicity	African American	+0.31	0.11	2.79	7.59	0.025
	Hispanic	+0.30	0.09	3.37	5.99	0.015
	Mix	+0.28	0.03	10.53	23.49	0.000
	White	+0.12	0.05	2.47	16.32	0.025
Poverty	0-65%	+0.29	0.04	7.12	15.12	0.000
	66-100%	+0.20	0.04	4.40	19.41	0.000
Outcome Type	General Reading	+0.14	0.04	3.32	12.62	0.006
	Alphabetics	+0.28	0.03	10.69	36.52	0.000
	Comprehension	+0.21	0.03	7.50	39.88	0.000
	Fluency	+0.27	0.04	7.42	28.03	0.000
Tutoring Specific Moderators						
Extra Time	Extra time	+0.28	0.04	6.40	13.02	0.000
	No extra time	+0.29	0.05	5.98	13.93	0.000
Group Size	1:1 Tutoring	+0.41	0.06	7.06	13.80	0.000
	1:Small Group Tutoring	+0.24	0.04	6.09	21.78	0.000
Tutoring Provider	Paid Volunteer	+0.36	0.09	3.97	4.63	0.012
	Teaching Assistant	+0.34	0.09	3.72	6.61	0.008
	Teacher	+0.29	0.04	7.19	18.03	0.000
	Unpaid Volunteer	+0.04	0.08	0.47	9.16	0.650

Table note: Mean effect sizes for each moderator category were calculated by estimated a model including the same covariates as those in Table 2 without an intercept, with the moderator as a categorical variable.

Figure 1. PRISMA Flow Diagram of Study Search and Review Process.



*A total of 58 unique citations were included in the review. Of those citations, 5 reported on more than one intervention, so they are included as having multiple studies, bringing the total number of included studies to 65.