Effective Reading Programs for Secondary Students

Ariane Baye
University of Liege (Belgium)

Cynthia Lake
Amanda Inns
Robert E. Slavin
Johns Hopkins University

Updated August, 2017

Abstract

Recent initiatives in the U.S. and U.K. have added greatly to the amount and quality of research on the effectiveness of secondary reading programs, especially programs for struggling readers. This review of the experimental research on secondary reading programs focuses on 73 studies that used random assignment (n=66) or high-quality quasi-experiments (n=7) to evaluate outcomes of 55 programs on widely accepted measures of reading. Specific programs using one-to-one and small-group tutoring, cooperative learning, and social-emotional approaches showed positive outcomes, as did a small number of programs emphasizing technology or teaching of metacognitive strategies. Benchmark assessments did not affect reading outcomes. Leaving aside tutoring and benchmarks, programs that provide additional time (usually, a daily extra period) were no more effective than programs that did not provide instructional time. The findings suggest that secondary readers benefit more from engaging and personalized instruction than from additional time on supplemental courses.

Key words: secondary reading programs, effectiveness studies, best-evidence synthesis, middle school reading, high school reading

Effective Reading Programs for Secondary Students

The reading performance of students in America's middle and high schools is one of the most important problems in education. In 2015, the National Assessment of Educational Progress (NAEP; NCES, 2016) reported that only 34% of eighth graders scored at or above proficient. This is up somewhat from 1992, when 29% of eighth graders scored proficient or advanced. At the twelfth grade level, 37% of students scored at or above proficient, a slight decrease from 40% in 1992. Despite a substantial focus on reading at all levels and massive federal, state, and local investments, secondary reading is advancing very slowly. According to the OECD Program for International Student Assessment (PISA), the mean performance of American 15-year-old students did not change between 2000 and 2009, and the proportion of struggling readers also did not change (OECD, 2010). Further, the OECD (2013) survey of adult competencies shows that the average reading level of American young adults (16-24 year olds) is below the international average for developed countries.

Significant gaps continue to exist between groups. While 44% of White eighth graders scored at or above proficient on NAEP, only 16% of African American students, 21% of Hispanic students, and 22% of American Indian/Alaska Natives did so (NCES, 2016). Among students qualifying for free lunch, only 20% scored at or above proficient. At the twelfth grade level, 46% of White twelfth graders scored at proficient or better, but the proportion was 17% for African Americans, 25% for Hispanics, and 28% for American Indian/Alaska Natives. In all of these groups, scores have been essentially static since 1992, the first year NAEP currently reports.

The lack of progress in twelfth grade reading, combined with recent increases in high school graduation rates (from 73% to 82% from 2006 to 2013; NCES, 2016), suggests that the

number of students graduating from high school with very low reading levels must be increasing

substantially. Students who do not read well in high school are likely to drop out or to graduate

without the skills to obtain anything more than menial work (Joftus & Maddox-Dolan, 2003).

Poor readers who do graduate are likely to experience serious difficulties entering college. Even

if they do go to college, they may be required to take non-credit, remedial English courses

(American Diploma Project, 2004; Au, 2000), which many students do not pass (ACT, 2006).

Ideally, reading success would be assured for all in elementary school. Yet even if

improved practices in elementary schools could reduce the numbers of below-level secondary

readers, the numbers are so large that it would be a very long time, if ever, before improved

reading programs in elementary schools would entirely solve the secondary reading problem.

Secondary schools will always need strategies to continue to build the reading skills of their

students.

The Need for a New Synthesis of Research on Secondary Reading Programs

Several reviews have examined the evidence base for various programs designed to

improve secondary reading. In particular, Deshler, Palincsar, Biancarosa, & Nair (2007)

summarized evidence on widely-used programs for adolescent readers, and Slavin, Cheung,

Groff, & Lake (2008) reviewed the findings of 33 studies of secondary reading programs that

met high methodological standards. Herrera, Truckenmiller, & Foorman (2016), using the

methods of the What Works Clearinghouse (2015), found 33 qualifying studies of secondary

reading programs, of which 12 found positive effects on reading outcomes. Wanzek et al. (2013)

found ten experimental-control studies that met criteria for their meta-analysis on extensive

reading intervention for students in grades 4 to 12. Edmonds et al. (2009) identified 17 studies of

interventions for struggling readers in grades 6-12 that compared experimental and control

groups, but in seven of these, treatments were delivered by researchers rather than teachers, and the studies involved very small sample sizes, averaging n=65.

Dietrichson, Bøg, Filges, & Jørgensen (2017) published a recent review of research on academic interventions for students with low socioeconomic status. This review overlapped the present review, but it focused on math as well as reading, and grades K to 8 instead of 6 to 12. Flynn, Zheng, & Swanson (2017) recently reviewed research on programs for struggling readers in upper elementary and middle schools, reporting on 10 qualifying studies. Scammacca, Roberts, Vaughn, & Stuebing (2015) carried out a meta-analysis of studies of interventions for struggling readers in grades 4-12 published in the period 1980 to 2011, primarily to describe trends over time in interventions, methods, and outcomes. The review included 10 studies, focusing on one to one or small group interventions, mostly for students with learning disabilities.

Slavin, Cheung, Groff, & Lake (2008) concluded that secondary reading programs that incorporated cooperative learning and other innovations in classroom teaching practices had the strongest effects on reading achievement in grades 6-12. Herrera et al. (2016) also reported that cooperative learning approaches and other methods providing extensive professional development were particularly likely to have positive outcomes. Dietrichson et al. (2017) found the strongest support for tutoring, feedback/progress monitoring, and cooperative learning.

Since earlier reviews were published, the number of high-quality, mostly large, randomized evaluations of secondary reading programs has increased substantially. One factor in this is Striving Readers (Boulay, Goodson, Frye, Blocklin, & Price, 2015), a series of large-scale, rigorous evaluations undertaken with substantial funding from the U.S. Department of Education, mostly to state departments of education. Further, in the U.S., the Institute of

Education Sciences (IES) has been funding randomized evaluations for some time, and these

have included high-quality secondary reading studies. Starting in 2015, several studies of

secondary reading programs funded by Investing in Innovation (i3) began to appear. Investing in

Innovation is a substantial initiative that provides development, validation, or scale-up funding to

programs based on their initial levels of evidence. Also in 2015, the Education Endowment

Foundation (EEF) in England began to report on randomized evaluations of many K-12

programs, and these have included several secondary reading approaches, especially small-group

and one-to-one tutoring. Finally, given a climate of emphasis on rigorous research, some

educational publishers have begun to fund their own third-party evaluations of secondary reading

programs.

For all of these reasons, the landscape of research on secondary reading programs has

substantially changed in recent years. There are now many more programs evaluated in more

rigorous experiments than existed previously.

The purpose of the current review is to review the research on secondary reading

programs using more stringent standards than would have been possible in the earlier reviews,

and assembling data from a much larger pool of programs and studies. The larger set of rigorous

studies, more than twice the number included in any previous review, enabled us to search for

patterns in the findings that would not have been possible in the past.

Focus of the Review

The present review synthesizes research on reading outcomes of programs designed for

middle and high school students. It uses best evidence synthesis (Slavin, 1986), a method

adapted from meta-analysis (see Cooper, 1998; Lipsey & Wilson, 2001) that includes narrative

as well as numeric summaries of the methods and findings of all studies meeting a common set

of inclusion criteria. The review describes methods and outcomes of individual studies and programs, and places studies in well-justified categories to find patterns that may have broader applicability and may suggest where additional development and research may be most fruitful. This review considers the strength of evidence supporting particular programs, but it also seeks to determine which *categories* of programs work best, and why they do so. This review is part of a series that has carried out more than a dozen syntheses of programs in elementary and secondary reading, mathematics, and science (see www.bestevidence.org), using similar inclusion standards and methods (see Cheung & Slavin, 2016).

Method

Criteria for Inclusion

The review focused on a set of studies that met rigorous inclusion criteria. The criteria were designed to minimize bias and maximize potential replicability in ordinary schools not involved in research. These were as follows.

1. Studies had to have evaluated reading programs for middle and high schools.

2. Studies had to have involved middle and/or high school students, grades 7-12. Sixth graders in middle schools (but not those in elementary schools) were also included. Students who were struggling readers or who qualified for special education services but attended mainstream English or reading classes were included.

3. Studies had to have compared students in a given reading program to those taught in an alternative or "business-as-usual" control group.

4. Studies could have taken place in any country, but the report had to be available in English. In practice, all included programs took place in the U.S. or the U.K.

5. Studies had to have used random assignment to experimental and control conditions or quasi-experimental methods in which treatment assignments were known in advance. Post-hoc quasi-experiments, in which "matched" control groups were identified after outcomes were known, were not included.

6. Studies had to provide pretest data. Those with experimental-control differences equivalent to an effect size of +0.25 or more on pretests were excluded. Pretest equivalence had to be established based on pretests for the final sample, after attrition.

7. Treatments had to be delivered by teachers, not by researchers. Studies in which researchers or graduate students were helping students in the classroom were excluded (e.g., Vaughn et al., 2017), on the basis that treatments delivered by researchers inflate study outcomes (deBoer et al., 2014).

8. Studies' dependent measures had to be quantitative measures of reading performance. When standardized tests were used, "total reading" or "total comprehension" were accepted. If comprehension and vocabulary, decoding, or fluency measures were presented separately, a "total reading" score was computed weighting comprehension at twice the value of measures other than comprehension, to reflect the critical role of comprehension in secondary reading.

9. Assessments made by developers or researchers were excluded, as such measures have been found to greatly overstate program impacts (Cheung & Slavin, 2016; de Boer, Donker, & van der Werf, 2014; Edmonds et al., 2009; Kulik & Fletcher, 2016; Lipsey et al., 2012).

10. Studies had to have a minimum duration of 12 weeks, to make it more likely that effective programs could be replicated over extended periods.

11. Studies had to have at least two teachers and 30 students in each treatment group.

12. In general, studies had to be carried out after 1990, but for technology approaches we used a start date of 2000, due to the significant advances in technology since that date.

The inclusion criteria were similar to those used by Slavin, Cheung, Groff, & Lake (2008), with a few key exceptions: exclusion of post-hoc studies, use of a criterion for pretest equivalence of ES <0.25 (instead of ES<0.50), and exclusion of measures made by developers or researchers. These all had the effect of making standards of inclusion more stringent. Stringent standards result in fewer accepted studies and may diminish reported effect sizes (Cheung & Slavin, 2016; de Boer et al., 2014; Lipsey et al., 2012), but studies meeting these requirements are more likely to replicate in research and practice. In practice, these standards had a modest impact on the number of studies included, however, as most studies excluded based on these criteria were also ineligible on other criteria. Most excluded studies lacked control groups, used experimenter-made measures, or were very brief.

Literature Search Procedures

A broad literature search was carried out in an attempt to locate every study that could possibly meet the inclusion requirements. Electronic searches were made of educational databases (JSTOR, ERIC, EBSCO, Psych INFO, Dissertation Abstracts) using different combinations of key words (for example, "secondary students," "reading," "achievement") and the years 1990-2017. Results were then narrowed by subject area (for example, "reading intervention," "educational software," "academic achievement," "instructional strategies"). In addition to looking for studies by key terms and subject area, we conducted searches by program name. Web-based repositories and education publishers' websites were also examined. These efforts were made to identify unpublished studies because of the known difference in effect sizes between published and unpublished studies (Polanin, Tanner-Smith, & Hennessy, 2016). We

searched for studies reviewed by the What Works Clearinghouse (2016) and ones reported online

by i3, IES, EEF, and other funders and researchers. We contacted producers and developers of

reading programs to check whether they knew of studies that we had missed. Citations from

previous reviews of secondary reading programs or potentially related topics such as technology

(Chambers, 2003; Murphy et al., 2002) were further investigated. We also conducted searches of

recent tables of contents of key journals from 2003 to 2017: *American Educational Research

Journal, Reading Research Quarterly, Journal of Educational Research, Journal of Adolescent

& Adult Literacy, Journal of Educational Psychology,* and *Reading and Writing Quarterly.*

Citations of studies appearing in the studies found in the first wave were also followed up.

Effect Sizes

In general, effect sizes were computed as the difference between experimental and

control individual student posttests after adjustment for pretests and other covariates, divided by

the unadjusted posttest control group standard deviation. If the control group SD was not

available, a pooled SD was used. Procedures described by Lipsey & Wilson (2001) were used to

estimate effect sizes when unadjusted standard deviations were not available, as when the only

standard deviation presented was already adjusted for covariates or when only gain score SD's

were available. If pretest and posttest means and SD's were presented but adjusted means were

not, effect sizes for pretests were subtracted from effect sizes for posttests. Weighted mean effect

sizes and meta-analytic tests such as Q statistics were calculated in R (R Core Team, 2016) using

the *metafor* package (Viechtbauer, 2010).

Statistical Significance

Statistical significance is reported for all studies. The criteria for statistical significance

are generally those of the What Works Clearinghouse (2015) and the Every Student Succeeds

Act (ESSA) evidence standards. When studies used random assignment or matched assignment at the individual level, they usually compared experimental and control groups using analysis of covariance (ANCOVA) controlling for pretests and, in some studies, demographic variables (e.g., race, free lunch) or other inputs.

When studies randomly assigned classes or schools to treatments or when they compared matched classes or schools, they should have used multilevel modeling such as Hierarchical Linear Modeling (HLM; Raudenbush & Bryk, 2002) to analyze the data. If they did, this review reports the findings as they were in the article. However, if a clustered design used a student-level analysis, such as ANCOVA, the review recalculated the analysis to estimate the results that would have been obtained in HLM, using a formula provided by the What Works Clearinghouse (2015). Following both WWC and ESSA evidence standards, a program is considered effective if it has at least one statistically significant positive effect, and no significant negative effects.

No criterion was used for the practical value of effect sizes. Cheung & Slavin (2016) reported that large randomized studies (n ≥250) using measures not made by researchers or developers had an average effect size of +0.12, considerably smaller than effect sizes for quasi-experiments or smaller studies. Since most studies in this review were large and randomized, effect sizes above +0.12 may at least be considered "above average" for this type of study.

Limitations

It is important to note several limitations of the current review. First, the review focuses on quantitative measures of reading. There is much to be learned from qualitative and correlational research that can add depth and insight to understanding the effects of secondary reading programs. Second, the review focuses on replicable programs used in realistic school settings over periods of at least 12 weeks, and excluding studies in which researchers directly

delivered treatments. This emphasis is consistent with the review's purpose in providing educators with useful information about the strength of evidence supporting various practical programs, but it does not attend to shorter, more theoretically-driven studies that may also provide useful information, especially to researchers. Finally, the review focuses on traditional measures of reading performance, primarily standardized tests. These are useful in assessing the practical outcomes of various programs and are fair to control as well as experimental schools or teachers, who are equally likely to be trying to help their students do well on these assessments. The review does not report on measures made by developers or researchers, because such measures tend to report greatly inflated effects (Cheung & Slavin, 2016). However, results on such measures may be of importance to researchers or educators.

Categories of Research Design

Included studies used either randomized or quasi-experimental designs, and within these, they used either clustered or student-level assessment. *Randomized experiments* were those in which students, classes, or schools were randomly assigned to treatments, and data analyses were at the level of random assignment. Student randomized experiments had students randomly assigned to treatments within classes or schools, while cluster randomized experiments were those in which schools or classes were the unit of random assignment and treatment. *Quasi-experimental* studies were ones in which experimental and control groups were matched on key variables at pretest, before posttests were known. These studies could also be clustered (matched classrooms or schools) or within-cluster (matched students within schools or classes). Studies using fully randomized designs are less subject to bias than quasi-experimental studies (Cheung & Slavin, 2016).

Program Categories

All studies that met the inclusion criteria were accepted, regardless of the experimental program they implemented. The studies were sorted by principal theories of action for program impacts, using information in the study reports and, if available, on the programs' websites. The categories, and their principal theories of action, were as follows.

1. **Tutoring**. Tutoring programs were ones in which struggling readers were given one-to-one or small group tutoring, in groups of one to four (the one exception was Butterfly Phonics, which had one teacher and one paraprofessional working with groups of 6-8). Tutors could be teachers, paraprofessionals, volunteers, or older students. Tutoring sessions were typically given either on some proportion of days (as few as once a week) or daily for a few months. One-to-one and small group tutoring approaches have been very effective in the elementary grades (Slavin, Lake, Davis, & Madden, 2011; Wanzek & Vaughn, 2007), so it seemed likely that they would also be effective in secondary school.

   The theory of action behind tutoring emphasizes personalization to the needs and learning strengths of individual students, opportunities to vary the level and pace of instruction for students, and forming personal, caring relationships between tutors and students (Slavin, Lake, Davis, & Madden, 2011; Elbaum, Vaughn, Hughes, & Moody, 2000; Wanzek & Vaughn, 2007). In this review, all tutoring studies took place in England, under funding from the Education Endowment Foundation. Otherwise, the tutoring models varied widely.

2. **Programs emphasizing social support for learning.** Two types of programs involved structuring social support for learning among students and teachers. One was **cooperative**

learning, where students worked in small groups to help each other learn. The other was **social-emotional** programs, which had only one example with qualifying studies: Building Assets Reducing Risk (BARR).

  Cooperative learning programs. Cooperative learning programs involved students working daily in small mixed-ability groups. Usually, cooperative learning groups had 4-5 members, and the students were encouraged to help each other learn academic content. In secondary reading, cooperative learning programs focused primarily on students helping each other to learn and apply metacognitive comprehension strategies, such as clarification, summarization, graphic organizers, and prediction.

  The theory of action behind cooperative learning emphasizes motivation through engagement with peers and encouragement from them, learning by explaining to peers and receiving explanations from them, and personalization through individualized feedback from peers and teachers (Slavin, 2015; Roseth, Johnson, & Johnson, 2008; Webb & Mastergeorge, 2003).

  Social-Emotional Programs. The one example in this category, Building Assets Reducing Risks (BARR) was based on the idea that ninth graders (the program's focus) are in need of strong personalized social connections and social emotional skills. Students and teachers are organized in interdisciplinary blocks in which all teachers and students are responsible for each others' success.

3. **Programs incorporating technology.** A number of widely used secondary reading programs incorporated **technology**. The theory of action behind such programs emphasizes adaptation of the level and pace of instruction to the individual needs of each student; ongoing formative assessment with immediate feedback to students and to

teachers; and the motivational value to students that computers can bring, not only intrinsically but also with recognition and celebration built into the software to encourage students' progress (Cheung & Slavin, 2013; Borgman et al., 2008; Hannafin & Land, 1997; MacArthur, Ferretti, Okolo, & Cavalier, 2001). Technology programs could be used within regular class time, or they could be used during additional class periods as supplemental instruction, replacing an elective such as art, music, or study hall.

4. **Metacognitive strategy approaches**. The largest number of included studies evaluated teaching of **metacognitive strategies**. In these programs, teachers taught students to apply metacognitive strategies to improve their comprehension, either during regular class time or during extra remedial class time.

  The theory of action behind metacognitive strategy approaches without extra time emphasizes the idea that students who are struggling in reading need to learn specific, well-defined strategies to become effective comprehenders (McCormick, 2003; Pressley, 2003). For example, they need to learn to get the gist, or meaning, of the texts they are reading, to learn and apply useful strategies when they run into unknown words, to learn story structure using predictions and identifying story grammar, and to learn to comprehend factual texts using outlining or graphic organizers. The core idea is that poor reading in secondary school, beyond any remaining problems with phonics, fluency, or vocabulary, can best be addressed by teaching students clear methods for understanding what they are reading.

  We divided metacognitive strategy programs into three subcategories. One category was **metacognitive strategies without extra time.** Another was **metacognitive strategy programs with additional time,** in which struggling readers received a daily

class period of reading emphasizing metacognitive skills in addition to their regular

English or reading class, replacing an elective. Extra-time metacognitive strategy

approaches had an additional theory of action, emphasizing the need low-performing

readers have for extra instruction to catch up to their peers in this critical skill.

Metacognitive strategies designed to serve English learners** emphasize teaching of

English vocabulary, grammar, and modes of expression in reading and writing.

5. **Benchmark assessments.** A popular practice in schools at all levels is the use of interim

or benchmark assessments, which test students a few times each year to see if they are on

track toward succeeding on state assessments. The theory of action is that with

information on students' current status, as well as specific areas of strength and

weakness, teachers and school leaders can find out early what students need and provide

additional resources to help students before it is too late.

Results

A total of 73 studies evaluating 55 different programs met the criteria of this review

(note: when two distinct programs were compared to control groups and reported in the same

article, they counted as two "studies"). As a group, the studies were of very high methodological

quality. 66 (90%) used random assignment, and only 7 (10%) used matched, quasi-experimental

designs. In 37 studies (51%), the unit of analysis was the school or classroom, and in each case

analyses were appropriate to the level of clustering (or corrected to be so).

As noted earlier, programs were grouped and shown in tables according to their main

features and theories of action: Tutoring, cooperative learning, technology without extra time,

technology with extra time, metacognitive strategies without extra time, metacognitive strategies

with extra time, metacognitive strategies for English learners, and benchmark assessments.

Tutoring Interventions for Struggling Readers

One-to-one or small-group tutoring is a widely used and effective intervention for struggling readers in elementary schools (Slavin, Lake, Davis, & Madden, 2011; Elbaum, Vaughn, Hughes, & Moody, 2000), but is rarely used in secondary schools. All seven of the qualifying tutoring studies (Table 1) were done in England, as a result of a funding initiative focused on helping struggling students enter secondary school with adequate reading skills. These interventions were applied in Year 6 (the last year of primary school) and/or in the early years of secondary school.

One-to-one tutoring.

Catch Up® Literacy (Bentley & Reid, 1995) is a structured one-to-one tutoring intervention. Paraprofessionals provide 15-minute sessions to struggling readers twice a week over the course of a school year. Each session includes prepared reading, reading out loud, discussing the text, and linked writing. A study of Catch Up® Literacy in Year 6 (Rutt, Kettlewell, & Bernardenelli, 2015) found a marginally significant difference favoring the tutored students (ES=+0.16, p=.08).

Perry Beeches provides struggling readers with one hour of one-to-one tutoring every two weeks. Coaches tailor activities according to students' needs. A study of Perry Beeches by Lord, Bradshaw, Stevens, & Styles (2015) found a large positive effect (ES=+0.36, p<.01).

REACH Tutoring provides struggling readers with one-to-one tutoring in 35-minute sessions, once a week for 20 weeks. The tutors are specially trained paraprofessionals. Two forms of the program were evaluated by Sibieta (2016). In ordinary REACH, children read aloud from books at their level while tutors keep a "running record," which they then used as a basis for remedial teaching. In a variation, REACH-LC, a language comprehension element was

added. In comparison to control groups, the effect size for REACH was +0.33 (p<.001), and for REACH+ LC, it was +0.51 (p<.001), for a mean of +0.42.

Small-group tutoring.

Butterfly Phonics. Butterfly Phonics uses formal phonics instruction, understanding the global aspects of a text, and class discussion of text meaning to improve reading comprehension. The program is delivered to groups of 6-8 students by a trained practitioner and an assistant. Merrell & Kasim (2015) evaluated Butterfly Phonics, and found an effect size of +0.30 (p<.001).

Rapid Phonics combined with Sound Discovery. Rapid Phonics and Sound Discovery are small-group tutoring programs designed to improve decoding skills and reading fluency using structured instruction in letter/sound correspondence. The intervention was delivered to struggling readers in disadvantaged secondary schools and their feeder primary schools three times a week for 30 minutes over two 6-week periods, before the summer break (Year 6) and during the fall semester (Year 7). Students in the experimental group were taught by specialists in groups of four or less, taken out of their regular classes while control students continued their schooling as usual. In the evaluation (King & Kassim, 2015), the effect size was not significant (ES= -0.05, n.s.).

Taken together, the overall weighted effect size for tutoring programs provided by paid adults was +0.23. Effect sizes were +0.26 for three one-to-one programs and +0.17 for two small group programs.

Volunteer tutoring.

TextNow utilizes volunteer coaches to provide daily 20-minute sessions focusing on encouraging students to read for pleasure. It also includes an online component that further

increases students' engagement with literacy. A randomized trial of TextNow in England

(Maxwell et al., 2014) found no significant effects on reading comprehension (ES = -0.06, n.s.).

Cross-age tutoring.

Paired Reading is a cross-age tutoring approach with the goal of improving general

literacy. Year 9 and Year 7 students work together in a structured process to choose, read, and

discuss a text, with the older partner providing support to the younger one. Paired Reading

showed no significant effects (ES = -0.02, n.s.) in a randomized experiment in secondary schools

in England (Lloyd et al., 2015).

Programs Emphasizing Social Support for Learning

Table 2 summarizes findings of studies emphasizing social support for learning:

Cooperative learning and BARR.

Cooperative learning programs.

Cooperative learning methods, in which students work in small groups to help each other

grow in reading skills, are widely used in elementary reading and in many other subjects in

elementary and secondary grades. Table 2 lists seven studies of three programs.

The Reading Edge, adapted from a program called Student Team Reading, is a

cooperative learning program for middle schools in which students work in four- or five-member

teams to help one another build reading skills. Students engage in partner reading, story retelling,

story related writing, word mastery, and story-structure activities to prepare themselves and their

teammates for individual assessments that form the basis for team scores. Instruction focuses on

explicit teaching of metacognitive strategies. Across three studies of The Reading Edge, the

weighted effect size was +0.29 (Slavin, Chamberlain, Daniels, & Madden, 2009; Stevens &

Durkin, 1992a, b). The outcomes in the Slavin, Chamberlain, Daniels, & Madden (2009) study were significantly positive (ES=+0.15, p<.05).

Talent Development High School (TDHS) is a whole-school reform model. Within TDHS, Strategic Reading and Student Team Literature is the reading component. Strategic Reading is used during the first half of ninth grade in high-poverty schools to help students make significant progress at this crucial point in their progression toward graduation.

In TDHS, ninth graders receive a "double dose" of reading and math, amounting to 90 minutes a day for each subject. Students take double-dose English 1 in the second semester, but Strategic Reading in the first. In Strategic Reading, students work in small, interdependent cooperative learning groups doing paired reading for fluency and comprehension, practicing new vocabulary related to novels and plays, and helping each other identify characters, plots, and informational content, using structured partner discussion guides that provide background and vocabulary, and comprehension questions provide mini-lessons on specific comprehension strategies. Students also are given time for self-selected reading and writing activities.

Balfanz, Legters, & Jordan (2004) carried out an evaluation of TDHS in non-selective high schools in Baltimore, comparing to control schools that also provided double-dose reading. There was an effect size of +0.32, not significant at the cluster level.

Collaborative Strategic Reading (CSR) teaches reading comprehension strategies to students working in small cooperative learning groups. During the first 4-6 weeks of the intervention, teachers model reading strategies such as activating prior knowledge, predicting what will be learned from an expository passage, identifying breakdowns in understanding, finding the main idea, and generating questions after reading. During the remaining 12-14 weeks, students are assigned to cooperative learning groups to allow them to master each strategy. The

intervention is implemented 50 minutes a day, two days a week, during regular English Language Arts lessons.

The weighted mean effect size across three CSR studies was +0.04 (Denver Public Schools, 2016; Vaughn et al., 2011, 2013). Adding the findings of the CSR studies to those of the three Reading Edge studies and the TDHS study, the weighted mean effect size for all cooperative learning studies was +0.16.

Social-emotional learning programs.

BARR (Building Assets Reducing Risks) is a whole-school reform approach focused on developmental, academic, and structural challenges during ninth grade. BARR is used in all major subjects to attempt to increase student achievement by improving students' social-emotional skills, building positive student-teacher relationships, and solving non-academic barriers to learning, such as truancy and behavior problems. The strategy focuses on building students' personal assets and reducing substance abuse, delinquency, and other problems. BARR staff closely monitor student achievement, including real-time analysis of student data. Students take English, math, and science or social studies in a block, to build connections among students and teachers. Teachers in each block meet regularly to review the progress of at-risk students. Extensive professional development and coaching are provided to teachers and school leaders. Corsello & Sharma (2015) found a positive effect of BARR on reading (ES=+0.14, p<.01), and also math, and a larger study by Borman et al. (2017) also found significant positive effects (ES = +0.08, p<.05). The weighted mean was +0.09 (p<.05).

The weighted mean effect size for all nine studies evaluating programs focusing on social support for learning was +0.15. This is the largest category mean for any approach except tutoring by paid adults.

Programs Incorporating Technology: No Extra Time

Table 3A summarizes research on approaches incorporating technology within regular class periods. The ten programs of this kind vary widely, especially based on how frequently technology is used and how well integrated technology use and live teaching are.

ITSS (Intelligent Tutoring System for the Structure Strategy) is a web-based approach in which students are taught to comprehend nonfiction text by categorizing text structures using key elements in the text to find the main idea, activating prior knowledge, supporting cognitive monitoring, and using graphic organizers and flow charts to summarize texts. ITSS is used during regular language arts classes 30-45 minutes a week with the assistance of a paraprofessional. Animated "tutors" model and guide learners. Students practice, take regular assessments, and proceed at their own pace through self-instructional units. A study by Wijekumar, Meyer, & Lei (2017) found a significant positive effect for 7th graders (ES=+0.18, p<.05).

The Thinking Reader teaches reading comprehension skills to struggling readers. It provides students novels with a range of difficulty. Animated coaches and peers on the computer mimic reciprocal teaching, modeling comprehension strategies (such as summarizing, questioning, predicting, or visualizing) and prompting students to use them. Teachers use the program in their regular English Language Arts activities for 110-165 minutes a week. In a study by Drummond et al. (2011), the average effect size was +0.01 (n.s.).

SuccessMaker is an adaptive K-8 computer-based reading program. It provides individualized reading activities, game-like environments, interactive aids, and a reporting system to inform teachers on student progress. A study by Gatti (2011) found a nonsignificant effect size for seventh graders of +0.11 (n.s.).

Achieve3000 is an online literacy program that provides non-fiction reading content for students in grades 2 through 12. Metacognitive skills and reading strategies (such as summarizing, generating questions, setting a purpose for reading) are used to improve the comprehension of informational texts. The program offers diagnostic and assessment data to teachers and school administrators. In an evaluation by Shannon & Grant (2015), the effect size was +0.29 across grades 6 and 9 (p<.05).

eMINTS (Enhancing Missouri's Instructional Networked Teaching Strategies) is a comprehensive schoolwide program that provides extensive professional development to teachers to help them with technology integration, inquiry-based learning, and high-quality lesson design. In an evaluation by Meyers, Molefe, Brandt, Zhi, & Dhillon (2016), the mean reading effect size for eMINTS was -0.06 (n.s.).

READ 180 is an instructional model used 90 minutes each day with struggling readers. It combines 30 minutes of whole-group instruction, followed by one hour during which students rotate through three 20-minute blocks devoted to independent reading, small-group direct instruction with the teacher, and use of READ 180 adaptive software. Teachers receive video workshops on content area and literature reading strategies, vocabulary, writing, and grammar. READ 180 is always used in addition to ordinary English language arts, but in two studies the control group also received supplemental instruction, so there was no difference in time. Across the two qualifying studies of Read 180 without extra time (Lang et al., 2009; Meisch et al., 2011), the weighted effect size was +0.08.

System 44 is a version of READ 180 for adolescent readers who have not mastered basic phonics and decoding skills. The program focuses on decoding, fluency, and comprehension. In daily 60-minute lessons, the teacher gives 5-10 minutes of whole-class instruction, the students

spend 25-30 minutes working in small groups or individually, and they then receive 20-25 minutes of computer-delivered instruction. Each software lesson has a set of corresponding activities and materials such as paperback and audio books. Like READ 180, System 44 is always used during supplemental time, but in a study by Beam, Faddis, & Hahn (2011), the control group also had additional time, so there were no time differences. Outcomes were non-significantly negative (ES= -0.14, n.s.).

Prentice Hall Literature combines off-line textbooks with online components. Online material includes vocabulary games, audios, and videos. A study by Eddy, Ruitman, Hankel, & Sloper (2010) found non-significantly negative impacts of Prentice-Hall Literature (ES= -0.10).

Comprehension Circuit Training (CCT) uses content delivered on tablet computers to teach reading comprehension skills. Following video instruction, students work with a partner to practice lesson content. Students cycle through four major components, focusing on vocabulary skills, pre-reading, reading of iBooks, and comprehension quizzes. Across two studies (Fogarty et al., 2014, 2016), the mean effect size for CCT was +0.13. In the Fogarty et al. (2016) study, significant positive effects were found on the TOSREC (ES=+0.24, p<.05), but not other measures.

Texas Technology Immersion Pilot (eTxTip) was a three-year evaluation of a technology immersion intervention covering language arts, math, science, and social studies in grades 6-8. Contractors provided schools with a) wireless, mobile computing devices for every student and teacher, b) productivity, communication, and presentation software, c) online resources supporting state standards, d) online assessments linked to state standards, e) extensive professional development, and f) initial and ongoing technical support. A study by Shapley, Sheehan, Maloney, & Caranikas-Walker (2009) found a mean effect size of +0.06 (n.s.).

Averaging across all 12 studies of technology applications without extra time, the mean effect size was +0.06.

Programs Incorporating Technology: Extra Time

Table 3B summarizes studies of five technology programs that provide supplemental instruction during daily extra periods.

READ 180, described previously, was evaluated in three studies in which control groups did not receive the extra time received by READ 180 students (Schenck, Feighan, Coffey, & Rui., 2011; Sprague, Zaller, Kite, & Hussar, 2012; Swanlund et al., 2012). Two of the studies found significant positive effects. The weighted effect size across all three studies was +0.08. Across all five studies of READ 180 with and without extra time, the weighted mean effect size was also +0.08.

System 44 was described earlier. One study of this program, by Beam & Faddis (2012), found significant positive effects (ES=+0.20, p<.05), but the other, Beam et al. (2011) found significantly negative effects on the TOSREC (ES= -0.24) and non-significantly negative effects on CST (ES= -0.04). Across the two studies of System 44, with and without extra time, the mean effect size was +0.03.

Accelerated Reader is a widely used U.S. program, primarily used in elementary schools, but the only qualifying evaluation in secondary reading took place in England (Gorard, Siddiqui, & See, 2015). Accelerated Reader provides students with a wide range of books at their reading level, determined by an on-line test. On-line comprehension tests are provided for each book, and students can earn points based on completing many books at a high readability level. The Gorard et al. (2015) evaluation found a significant effect size of +0.24 (p<.05).

Passport Reading Journeys is a supplemental literacy curriculum that provides fifteen two-week sequences of lessons mixing whole-class and small group lessons as well as individualized computer-based practice. The curriculum focuses on reading comprehension strategies, vocabulary, word study, and writing, using mainly science and social studies topics. Across three studies (Dimitrov et al, 2012; Schenck et al, 2012; Vaden-Kiernan et al., 2012) the weighted mean effect size was +0.07. The Vaden-Kiernan et al. (2012) study found significant positive effects on the GRADE (ES=+0.27, p<.05), but not the LEAP (ES=-0.01), for a mean of +0.12.

iLit is a digital instruction approach for struggling readers. Students choose among more than 500 eBooks and work on vocabulary and comprehension strategies. Students work independently, keeping on-line journals, answering questions, and discussing books in groups. A two-year study of iLit by Gatti (2016) found an effect size of +0.09 (n.s.).

The weighted mean effect size across 9 studies of technology programs with extra time was +0.09. The mean for all technology studies (n=21) was +0.06. Six individual programs, READ 180, Accelerated Reader, ITSS, Achieve 3000, Passport Reading Journeys, and CCT, showed significant positive outcomes on some measures in some studies.

Metacognitive Strategy Approaches: No Extra Time

Table 4A lists 17 studies of programs using metacognitive strategies to enhance students' reading during regular class periods, without extra time. In these programs, students are taught specific strategies to help them comprehend narrative and factual texts, to study, and to write.

Teacher Effectiveness Enhancement Programme (TEEP) is a British professional development approach for secondary teachers involving all major subjects. Over the course of a school year, TEEP provides all teachers in a school three days of inservice training. The

strategies include formative assessment, thinking skills, cooperative learning, and effective use of technology. An evaluation by the Institute for Effective Education (2016) found no positive effects of TEEP (ES= -0.04, n.s.).

Word Generation is an approach to vocabulary building in which students are encouraged to discuss and read about topics containing target words believed to be important, but not already in students' speaking or reading vocabularies. Lawrence, Francis, Pare-Blagoev, & Snow (2016) evaluated reading effects of Word Generation and found no significant differences on reading (ES= +0.05, n.s.).

Chicago Striving Readers provides students technology tools (media and listening centers, handheld computers) to support their learning. All teachers were encouraged to assign partner reading in social studies, science, and math. A four-year evaluation found no significant effects (Simon et al., 2011; ES= -0.01, n.s.).

Project CRISS is a professional development approach designed to help teachers of all subjects use proven reading comprehension strategies. The project provides summer institutes to local facilitators, who then create local teacher-to-teacher study groups. An evaluation of CRISS (Kushman, Hanita, & Raphael, 2011) found no significant effects (ES=+0.05, n.s.)

Reading Intervention through Strategy Enhancement (RISE) relies on teachers' capacity to build effective curriculum for struggling readers. During RISE classes, students are given the opportunity to read independently, to work in small groups, and to receive whole group lessons. A study of RISE by Lang et al. (2009) found positive effects on reading for "moderate risk" students (ES= +0.27, p<.04), but not for "high risk" students (ES= -0.06, n.s.). The weighted mean effect size was +0.16 (p<.05).

REACH provides explicit, intensive instruction to struggling students in grades 6-12. It focuses on phonics, fluency, word knowledge, and reasoning skills, on narrative structure and writing skills, and spelling. Lang et al. (2009) evaluated REACH and found a mean effect size of -0.02 (n.s.).

Expert 21 provides student texts and supportive materials focused on building language arts, writing, and comprehension skills, including whole class and small group discussions, teaching of metacognitive skills such as graphic organizers, and collaborative projects. Sivin-Kachala & Bialo (2012) found non-significant positive effects of Expert 21 at the cluster level (ES= +0.15, n.s.).

Expository Reading and Writing Course (ERWC) is a program for 12th graders designed to prepare them to pass the California Early Placement Test (EPT), used in the California State University system to determine whether freshmen must take non-credit remedial English courses. ERWC provides curriculum materials, two days of professional development for teachers, professional learning communities, and at least four on-site coaching sessions for each teacher. The emphasis of the program is on discussion of text meaning, developing critical thinking skills, encouraging group discussions, developing oral language skills, and developing writing skills in multiple genres. Fong, Finkelstein, Jaeger, Diaz, & Broek (2015) found significant positive effects of ERWC (ES= +0.13, p<.05).

Schoolwide Enrichment Model – Reading (SEM-R) exposes students to a variety of books, spend time independently reading self-selected challenging books, and meet their teacher individually a few minutes every one to two weeks to discuss reading strategies and respond to higher-level questions. The program is implemented 40-45 minutes daily or three hours per week

in regular English Language Arts classes. A study by Little, McCoach, & Reis (2014) found non-significant positive effects (ES=+0.10, n.s.).

Vocabulary Enrichment Intervention Programme (VEIP) combines vocabulary teaching, phonics, and connected text. Teachers are given the flexibility to use the components as needed by their students. An evaluation of VEIP by Styles et al. (2014) in England found no significant outcomes (ES = +0.06, n.s.).

Strategic Instruction Model (SIM) (Deshler & Schumaker, 2005) is a family of programs all of which emphasize teaching students reading comprehension, decoding, and other reading objectives with step-by-step strategies. The strategies include word identification, visual imagery, self-questioning, paraphrasing, vocabulary learning, and writing. All SIM programs are primarily for struggling readers, but some are for all students in schools with many struggling readers, taught during regular class time (SIM programs delivered during additional class time are discussed in a later section).

SIM: Content Literacy Curriculum (CLC) is used with all students in high-poverty high schools. A large two-year study by Corrin et al. (2012) found non-significant positive effects on reading (ES=+0.09, n.s.).

SIM: Adolescent Literacy Model (ALM) was evaluated by Cantrell, Almasi, Carter, & Rintamaa (2011) in Kentucky middle and high schools. They reported a non-significant effect size of +0.10.

SIM: Xtreme Reading was evaluated by Faddis et al.(2011) in Portland (OR) middle and high schools. The middle schools did not provide extra time (the high schools did). For middle schools, the overall effect size was +0.20 (p < .05), the largest for all SIM variations.

Across the three studies of SIM without extra time, the weighted mean effect size was +0.11.

Reading Apprenticeship. Reading Apprenticeship is a family of programs designed to improve reading comprehension by integrating metacognitive strategy instruction into content areas, such as science and social studies. The model incorporates extensive reading, teaching of comprehension strategies, and collaborative sense-making. Teachers receive extensive professional development, including inquiry into current practices, analysis of videos of classroom teaching, and modeling of reasoning processes. The professional development takes place in ten day-long sessions over a period of up to two years. A study of Reading Apprenticeship by Greenleaf et al. (2011) found no significant outcomes (ES= +0.03, n.s.).

Reading Apprenticeship Improving Secondary Education (RAISE) is a whole-class version of Reading Apprenticeship. A study of RAISE by Fancsali et al. (2015) found positive effects (ES=+0.14, n.s.), but these were not significant at the cluster level.

iRAISE is a form of Reading Apprenticeship designed to make the program less expensive and easier to implement by providing biology teachers with professional development on line. Jaciw, Schellinger, Lin, Zacamy, & Toby (2016) found no positive effects of this approach (ES = 0.00, n.s.).

Across studies of the three variations of Reading Apprenticeship without extra time, the weighted mean effect size was +0.11.

Content Knowledge-Building and Student-Regulated Comprehension Practices is designed to improve students' reading comprehension. First, the teacher introduces the text with critical questions, activates students' prior knowledge, and models a "checkpoints" process which helps students to anticipate the logical places to stop and check comprehension. Then

students work in pairs to analyze the text and answer critical questions. In a study by Simmons et al. (2014), the effect size was -0.01 (n.s.).

Across 17 studies of programs emphasizing metacognitive strategies without providing additional instructional time, the weighted mean effect size was +0.06.

Metacognitive Strategy Approaches: Extra Time

Eleven studies evaluated metacognitive strategy programs that provided a daily extra period of reading. These studies are summarized in Table 4B and described in the following section.

SIM: Xtreme Reading is the main version of the Strategic Instruction Model designed as supplemental instruction. Students reading two to five years below grade level are usually given an additional daily period of SIM instruction, replacing an elective such as band or art. Across three qualifying studies (Somers et al., 2010; Faddis et al., 2011; Sprague, Zaler, Kite, & Hussar, 2012), the mean effect size for Xtreme Reading with extra time was +0.06 (n.s.).

SIM: Learning Strategies Curriculum (LSC) was used in an extra daily period with students who were at least two years below grade level. A study by Cantrell, Almasi, Rintamaa, & Carter (2016) found significant positive effects on reading (ES= +0.10, p<.05).

SIM: Fusion Reading is another SIM variation for struggling readers, provided during an extra daily period. A study by Schiller at al. (2012) found non-significant positive effects (ES= +0.07, n.s.).

Across all SIM variations that provided extra time for instruction, the weighted mean effect size was +0.08. Including SIM variations with and without extra time, the mean effect size was +0.09.

Every Classroom, Every Day (ECED) is a structured literacy curriculum based on authentic expository texts. It supplements the regular English course, doubling the amount of English Language Arts exposure. Early et al. (2015) found no significant effects of ECED on reading (ES=+0.06, n.s.).

Reading Apprenticeship Academic Literacy (RAAL) is an extra-time program based on Reading Apprenticeship (described previously). A study of RAAL by Somers et al. (2010) found significant positive effects (ES= +0.10, p<.05). Combining RAAL with three other variations of Reading Apprenticeship produced a mean of +0.10.

REWARDS is a one-year supplemental intervention given by trained teachers five times a week to develop reading skills. An evaluation of REWARDS by Newman, Kundert, Spaulding, White, & Gifford (2012) found a mean effect size of +0.09 (n.s.).

Kentucky Cognitive Literacy Model (KCLM) is an intervention for struggling readers focused on teaching comprehension strategies, vocabulary, study skills, and writing. Teachers receive templates to create or adapt units of instruction, rather than specified instructional materials. A study of the model by Cantrell, Carter, & Rintamaa (2012) found no significant outcomes (ES= -0.06).

Read to Achieve provides lessons on content area and narrative texts, and incorporates small group collaboration and independent activities. A study by Deussen, Scott, Nelsestuen, Roccograndi, & Davis (2012) found non-significant positive effects on reading (ES= +0.10, n.s.).

Strategic Adolescent Reading Intervention (STARI) is a multi-component program developed for middle school students reading below grade level. The program uses core novels and other engaging texts within thematic units. STARI includes fluency passages, comprehension instruction, reciprocal teaching strategies, as well as student discussion and

debate. Kim et al (2016) conducted a randomized trial of STARI and found significant positive effects of the program (ES = +0.15, p<.05).

Averaging across all 11 metacognitive strategy studies of programs that provided an extra period of instruction each day, the weighted mean effect size was +0.07.

Metacognitive Strategy Approaches: Focus on English Language Learners

Four studies of two programs focused on improving reading outcomes for English language learners and other language minority students. Table 4C summarizes the findings of these studies.

Pathway is a professional development program used primarily with mainstreamed Latino English learners. It is primarily focused on writing rather than reading, but it also teaches reading comprehension. In its main evaluation, teachers were taught over 46 hours how to teach cognitive strategies, such as preparing students to read, make inferences, and interpret complex texts. They were also taught process writing to develop students' interpretative reading and analytical writing abilities. Experienced Pathway teachers helped their colleagues as on-site coaches. The weighted average across the two Pathway studies (Olson et al., 2012, 2016) was +0.08. The study by Olson et al (2012) found significant differences (ES=+0.07, p<.05).

Academic Language Instruction for All Students (ALIAS) is a vocabulary intervention designed to be used 45 minutes a day in regular English Language Arts classrooms including many language minority students. Each cycle of lessons is based on one informational text from which are extracted a small number of high-utility and abstract words. The intervention includes a variety of whole-group, small-group, and independent activities, and gives opportunities for listening, speaking, reading, and writing with the targeted words. The weighted mean effect size across the two ALIAS studies was +0.06 (Lessaux, Kieffer, Faller, & Kelly,

2010; Lesaux, Kieffer, Kelley, & Harris, 2014). The weighted mean for the four studies of programs for ELLs was +0.08.

Since none of the metacognitive strategy approaches for ELLs provided extra time, the studies in Tables 4A and 4C (n=21) can be combined to produce a weighted mean effect size of +0.06 for all metacognitive strategy studies without extra time. Combining studies in Tables 4A, 4B, and 4C, the weighted mean effect size for all metacognitive strategy approaches was also +0.06. A few individual programs emphasizing metacognitive strategy instruction found significant positive effects on student reading outcomes on at least some important measures in some studies. These included SIM, the Expository Reading and Writing Course (ERWC), Reading Apprenticeship Academic Literacy (RAAL), Reading Intervention through Strategy Enhancement (RISE), STARI, and Pathway.

Benchmark Assessments

Four studies of the provision of benchmark assessments are summarized in Table 5.

4Sight is a benchmark assessment system in which students take tests four times a year to predict their performance on state tests. A large evaluation in 59 districts in 7 states (Carlson, Borman, & Robinson) found small, non-significant impacts on reading performance (ES= +0.03, n.s.).

Personalized Assessment Reporting System (PARS) provides reports of students' strengths and weaknesses on benchmark tests given three times a year to 10th graders, based on Ohio Graduation Tests (OGT) given in prior years. An evaluation by May & Robinson (2007) found no significant effects on OGT reading performance (ES= -0.03, n.s.).

Achievement Network (ANet) provides quarterly interim assessments in English and math, data tools to report on students' progress, and networks of peer schools to share results and

engage in joint professional development. West, Morton, & Herlihy (2016) carried out an evaluation of ANet, and found an effect size of -0.03 (n.s.).

Acuity is a series of diagnostic or predictive benchmark assessments designed to forecast performance on state tests. Teachers are given materials that they may assign to students to improve in areas of need. A study by Konstantopoulos, Miller, van der Ploeg, & Li (2016) evaluated Acuity across Indiana and found a non-significant impact on middle school reading (ES=+0.03, n.s.).

Across the four benchmark studies, the weighted mean effect size was 0.00.

Differences by Design Factors

All studies were included in a fixed effects model, which indicated a significant amount of heterogeneity in effect sizes (Q=380.55, df=72, p<.001). In order to examine this variance, a random effects model was used to explore several design features (e.g., grade, research design). The tutoring studies (Table 1) and benchmark studies (Table 5) were excluded in these analyses because they differed so much from the other approaches.

Extra time vs no extra time. In many of the qualifying studies, the intervention was provided to target groups of low achievers, during an extra daily class period not received by control students. In these studies, the control group was typically participating in electives, such as art or band, so the intervention provided substantial additional teaching time to the experimental group over one or more years. In extra-time treatments, group sizes were typically small (usually 12 to 20).

An analysis comparing extra-time programs and no extra time studies found no significant differences (Q_m=0.01, n.s.). 20 studies of programs providing extra time had a mean weighted effect size of +0.07, while 42 studies of programs providing no extra time also had a weighted mean effect size of +0.07.

Middle vs. high school. We tested the difference in outcomes between programs used in the middle grades (6-8) and those used in high school (9-12). Weighted mean effect sizes were slightly higher (Q_m= +0.01, n.s.) for middle schools (n=39 studies, ES=+0.08) than for high schools (n=30 studies, ES=+0.07).

Differences by research design. We compared effect sizes between studies that used random assignment to conditions (n=55 studies, ES=+0.05) and quasi-experiments, which used matching (n=7 studies, ES=+0.14). This significant difference (Q_m=8.53, p<0.01) is similar to the differences between randomized and quasi-experimental studies reported by Cheung & Slavin (2016). Effect sizes were non-significantly (Q_m=0.52, n.s.) lower for studies using clustered designs (n=31, ES=+0.06) than for those using designs employing student-level analyses (n=31, ES=+0.10).

Programs Meeting ESSA Standards for Strong and Moderate. The Every Student Succeeds Act (ESSA) defines "Strong" and "Moderate" criteria for evidence supporting educational programs. "Strong" requires that at least one randomized study found significantly positive effects and no significantly negative effects, and "Moderate" requires that at least one quasi-experimental (matched) study found significantly positive effects and no significantly negative effects.

Table 6 summarizes the programs that meet these ESSA categories, along with the numbers of studies, weighted mean effect sizes, and ESSA ratings. What the table illustrates is that even though there are program categories particularly associated with positive outcomes, every category, with the sole exception of benchmark assessments, contains individual programs that have been found in at least one study to have significantly positive outcomes on secondary reading measures.

Discussion

This review of rigorous research on programs designed to enhance the reading of students in middle and high schools found that most studies meeting inclusion criteria had relatively small effects on student reading. However, three quite different categories contained programs with more positive impacts. One of these was tutoring programs, all done in England, in which adult tutors worked with groups of one to four students (or in one case, two adults to 6-8 students). Three of the five tutoring studies using paid adult tutors found significant positive effects. The weighted mean effect size across all five studies was +0.23. It is not surprising that tutoring would be effective, as it has also been effective in elementary reading (Slavin, Lake, Davis, & Madden, 2011; Wanzek et al., 2013). Tutors are able to build individual relationships with students, and to personalize instruction to their individual needs. It is important to note that one study using volunteer tutors and one using Year 9 students to tutor Year 7 students did not find positive outcomes, but well-trained, paid tutors working with one or a few students generally made an important difference in student outcomes.

Another category with particularly positive outcomes was cooperative learning, especially The Reading Edge, with one study with a significant positive outcome and a weighted mean effect size across three studies of +0.29. The related Talent Development High School's Strategic Reading Approach had an effect size of +0.32, though this was not significant at the cluster level. What makes cooperative learning distinctive is that it taps into the social motivations that drive most of adolescent behavior. By having students work in teams, with team recognition based on the achievement gains of all team members, teammates encourage each others' efforts, explain ideas to each other, and have opportunities to ask others for help.

However, three studies of another cooperative learning approach, Collaborative Strategic

Reading, found small impacts (weighted mean ES= +0.04). Combining across all seven studies of cooperative learning, the mean effect size was +0.16. Another approach emphasizing social support for learning, BARR, had small but significant positive effects across two studies (weighted mean effect size= +0.09).

The conclusion that tutoring and cooperative learning were the most effective approaches in secondary reading is consistent with the conclusions of previous reviews by Dietrichson et al. (2017), and reviews by Slavin, Cheung, Groff, & Lake (2008) and Herrara et al. (2016) also emphasized the positive effects of cooperative learning (studies of tutoring in secondary school were rare before the British studies appeared in recent years.)

Overall impacts of technology approaches were modest (mean ES=+0.06), but six of the 13 programs had at least one significant positive effect. The significant effect size in a study of Achieve3000 was +0.29. A program called ITSS focused on teaching comprehension strategies for nonfiction text had a significant effect size of +0.18. A British study of a form of Accelerated Reader using technology reported a significant effect size of +0.24. Two studies of Comprehensive Circuit Training had a mean effect size of +0.13, and one measure in one of the studies was significant. Of five studies of READ 180, two found significant impacts (ES=+0.18 and +0.14), but three reported small and non-significant impacts, for a mean effect size of +0.08. One study of Passport Reading Journeys found significant positive impacts, and the mean weighted effect size was +0.07.

Six of the 20 metacognitive strategy programs found at least one significant positive effect size. One family of approaches was the Strategic Instruction Model (SIM), which teaches struggling adolescent readers step-by-step strategies for phonics, comprehending, writing, note-

taking, and other skills. Seven studies evaluated SIM variations and three of these found significant positive effects. The outcomes of all seven studies averaged +0.09.

Another interesting family of programs was Reading Apprenticeship (Greenleaf et al., 2011), which focuses on comprehension and writing strategies. The only Reading Apprenticeship variation to have significant positive effects was Reading Apprenticeship Academic Literacy (RAAL; Somers et al., 2010), and the mean across the four studies was +0.10.

Other metacognitive strategies found significantly positive outcomes in single studies. One was RISE, a professional development strategy focused on differentiated instruction and cooperative learning. RISE was found to be very effective for students at moderate risk (ES=+0.27) but not high risk (ES=-0.06), for a weighted mean ES of +0.16. The Expository Reading and Writing Course (ERWC), a program designed to help twelfth graders prepare for the test they will take as freshmen if they attend California State Universities, reported a significant positive effect on this test, with an effect size of +0.13. Pathway, a professional development program for teaching mainstreamed English learners, also showed significant effects across two studies (ES = +0.08).

The findings of this review permit some general conclusions about effective programs for struggling adolescent readers. Perhaps the most surprising finding is the lack of positive effects of providing an additional class period for reading each day. Programs focused on improving teachers' practices during regular class produced virtually identical impacts as programs that also focused on improving teachers' practices but added a daily period of instruction. The finding matches the conclusions of reviews by Kidron & Lindsay (2014) and Zief, Lauver, & Maynard (2006), though these reviews focused on extra-time approaches such as before- and after-school

programs and summer school, not an additional period for reading instruction for a year (as in the current review).

Another surprising finding was the minimal impact in most studies of providing professional development in metacognitive strategies to teachers of struggling adolescents. Numerous theorists (e.g., McCormick, 2003; Pressley, 2003) have emphasized the importance of teaching metacognitive strategies. Because metacognitive strategy instruction is so widespread, it is likely that many control teachers were also using these strategies, reducing the apparent impact. Yet in the 31 studies evaluating metacognitive strategy instruction, teachers typically received far more professional development and coaching than usual, yet outcomes were generally modest, averaging an effect size of +0.05.

The lack of positive effects of benchmark assessments (mean effect size=0.00) is interesting in light of the widespread use of these strategies. Three of the four studies summarized in Table 5 all assessed impacts of benchmark assessments in elementary as well as middle school reading and also in elementary and middle school math, and the outcomes were similarly near zero. Benchmark assessments may be useful as a support for effective forms of professional development, for example, but they clearly do not increase achievement on their own.

Looking at the nature of the programs that found significant positive effects, a pattern emerges. The more successful programs tended to place a strong emphasis on relationships, not just teaching strategies. In tutoring and BARR, the relationships were particularly between teachers and small groups of students. In cooperative learning the emphasis was on student-student relationships.

In addition to relationships, these approaches give students opportunities to be active and social, and combine learning with fun. This may be particularly important for adolescents who have not been very successful in school.

Another common feature across the more successful programs was that they generally allowed for personalization. This is most obvious in the tutoring programs, which can substantially personalize teaching to meet students' learning, motivational, and social needs. However, personalization is also facilitated by cooperative learning, where teammates can adapt tasks and explain concepts to each other. Technology-assisted instruction, where the teacher and computer, along with the student, can readily determine what individual students need, also facilitates personalization.

These patterns are far from definitive, especially in light of the low effect sizes and inconsistent outcomes seen in many programs and program types. However, these findings may suggest means of building more effective methods for improving outcomes in secondary reading. Programs that focus on secondary students as social beings with strong needs for social connections and for teaching that is personalized to their unique strengths and needs may produce outstanding outcomes in reading. The research reviewed here outlines promising avenues toward more effective approaches, but much remains to be done to understand how to create replicable, cost-effective strategies that can reliably and meaningfully improve reading outcomes for middle and high school students.

References[*]

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

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

*Archer, A.L., Gleason, M. M., & Vachon, V. (2005a). *REWARDS intermediate: Multisyllabic word reading strategies.* Longmont, CO: Sopris West.

*Archer, A.L., Gleason, M. M., & Vachon, V. (2005b). *REWARDS plus: Reading strategies applied to social studies passages.* Longmont, CO: Sopris West.

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

Awbrey, A. (2008). *Where everyone gets it: CTL's adolescent literacy model*. Unpublished manuscript.

*Balfanz, R., Jordan, W., & Legters, N. (2004). *Catching Up: Impact of the Talent Development ninth grade instructional interventions in reading and mathematics in high-poverty high schools*. Center for Research on the Education of Students Placed at Risk, Johns Hopkins University.

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

---

[*] Studies included in the review

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

Bently, D., & Reid, D. (1995). *Supporting struggling readers.* Widnes, England: UK Reading Association.

Biancarosa, C., & Snow, C. E. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.

*Boardman, A., Klingner, J., Buckley, P., Annamma, S., & Lasser, C. (2015). The efficacy of Collaborative Strategic Reading in middle school science and social studies classes. *Reading and Writing, 28* (9), 1257-1283. doi:10.3102/0034654316652942.

Borgman, C., Abelson, H., Dirks, L., Johnson, R., Koedinger, K., Linn, M., Lynch, C., et al. (2008). *Fostering learning in the networked world: The cyberlearning opportunity and challenge.* Washington, DC: National Science Foundation.

*Borman, T., Bos, J., O'Brien, B., Park, S. J., Liu, F., Corsello, M., & Jerabek, A. (2017). *The BARR program: Impacting social emotional skills and academic achievement of 9th grade students in 6 high schools - results from a randomized controlled trial.* Washington, DC: Society for Research on Educational Effectiveness.

Boulay, B., Goodson, B., Frye, M., Blocklin, M., & Price, C. (2015). *Summary of research generated by Striving Readers on the effectiveness of interventions for struggling adolescent readers.* Washington, DC: USDOE.

*Cantrell, S. C., Almasi, J. F., Carter, J. C. & Rintamaa, M. (2011). *Striving Readers final evaluation report: Danville, KY.* Lexington, KY: Collaborative Center for Literacy Development, University of Kentucky.

*Cantrell, S. C., Almasi, J. F., Rintamaa, M., & Carter, J. C. (2016). Supplemental reading

strategy instruction for adolescents: A randomized trial and follow-up study. *The Journal

of Educational Research*, *109*(1), 7-26. doi:10.1080/00220671.2014.917258.

*Cantrell, S. C., Carter, J. C., & Rintamaa, M. (2012). *Striving Readers Cohort II Evaluation

Report: Kentucky*. Lexington, KY: Collaborative Center for Literacy Development,

University of Kentucky.

*Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster

randomized trial of the impact of data-driven reform on reading and mathematics

achievement. *Educational Evaluation and Policy Analysis, 33*(3), 378–398.

https://doi.org/10.3102/0162373711412765

Chambers, E.A. (2003). *Efficacy of educational technology in elementary and secondary

classrooms: A meta-analysis of the research literature from 1992–2002*. Unpublished

doctoral dissertation, Southern Illinois University at Carbondale.

Cheung, A., & Slavin, R. E. (2013). Effects of educational technology applications on reading

outcomes for struggling readers: A best-evidence synthesis. *Reading Research Quarterly, 48*

(3), 277-299.

Cheung, A., & Slavin, R. (2016). How methodological features affect effect sizes in education.

Educational Researcher, 45 (5), 283-292. DOI:

https://doi.org/10.3102/0013189X16656615.

*Clarke, P., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading

comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*,

1106-1116. doi:10.1177/0956797610375449

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

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

*Corrin, W., Lindsay, J. J., Somers, M-A., Myers, N. E., Myers, C. V., Condon, C. A., & Smith,
J. K. (2012). *Evaluation of the Content Literacy Continuum: Report on program impacts,
program fidelity, and contrast. (NCEE2013-4001).* Washington, DC: National Center for
Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S.
Department of Education.

*Corsello, M., & Sharma, A. (2015). *The Building Assets-Reducing Risks Program: Replication
and expansion of an effective strategy to turn around low-achieving schools*. Napa, CA:
Corsello Consulting.

de Boer, H., Donker, A. S., & van der Werf, M. P. (2014). Effects of the attributes of educational
interventions on students' academic performance: A meta-analysis. *Review of
Educational Research*, *84*(4), 509–545. doi:10.3102/0034654314540006

*Denver Public Schools (2016). *Final report to the U.S. Department of Education on
Collaborative Strategic Reading.* Denver, CO: Author.

*Deussen, T., Scott, C., Nelsestuen, K., Roccograndi, A., & Davis, A. (2012). *Washington
Striving Readers: Year 1 evaluation report*. Portland, OR: Education Northwest.

Deshler, D., & Schumaker, J.B. (2005). *Teaching adolescents with disabilities: Accessing the
general education curriculum.* Thousand Oaks, CA: Sage.

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

Dietrichson, J., Bøg, M., Filges, T., & Jørgensen, A-M. K. (2017). Academic interventions for
elementary and middle school students with low socioeconomic status: A systematic
review and meta-analysis. *Review of Educational Research, 87*(2), 243-282.

*Dimitrov, D., Jurich, S., Frye, M., Lammert, J., Sayko, S., & Taylor, L. (2012). *Year one evaluation report/impact study: Illinois Striving Readers*. Arlington, VA: RMC Research Corporation.

*Drummond, K., Chinen, M., Duncan, T., Miller, H., Fryer, L., Zmach, C., & ... National Center for Education Evaluation and Regional Assistance, (2011). *Impact of the Thinking Reader[R] Software Program on grade 6 reading vocabulary, comprehension, strategies, and motivation: Final report. NCEE 2010-4035.* Washington, DC: National Center For Education Evaluation And Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

*Early, D., Berg, J. K., Alicea, S., Si, Y., Aber, J. L., Ryan, R. M., & Deci, E. L. (2015). The impact of every classroom, every day on high school student achievement: Results from a school-randomized trial. *Journal of Research on Educational Effectiveness, 9* (1), 3-29.

*Eddy, Ruitman, Hankel, Sloper (2010). *The effects of Pearson Prentice Hall Literature on student performance: Efficacy study final report*. La Verne, CA: Cobblestone.

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

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.

*Faddis, B., Beam, M., Maxim, L., Vale Gandhi, E., Hahn, K., & Hale, R. (2011). *Portland*

  Public Schools' Striving Readers Program Year 5 evaluation report. Portland, OR: RMC

  Research Corporation.

*Fancsali, C., Abe, Y., Pyatigorsky, M., Ortiz, L., Hunt, A., Chan, V., Saltares, E., Toby, M.,

  Schellinger, A., & Jaciw, A. P. (2015). *The impact of the Reading Apprenticeship*

  Improving Secondary Education (RAISE) Project on academic literacy in high school: A

  report of a randomized experiment in Pennsylvania and California schools. (Empirical

  Education Rep. No. Empirical_ RAISE-7019-FR1-O.2). Palo Alto, CA: Empirical

  Education Inc.

Flynn, L. J., Zheng, X. & Swanson, H. L. (2012). Instructing struggling older readers: A

  selective meta-analysis of intervention research. *Learning Disabilities Research and*

  Practice, 27(1), 21-32.

*Fogarty, M., Clemens, N., Simmons, D., Anderson, L., Davis, J., Smith, A., ... & Oslund, E.

  (2016). Impact of a technology-mediated reading intervention on adolescents' reading

  comprehension. *Journal of Research on Educational Effectiveness.* Retrieved from

  http://www.tandfonline.com/doi/abs/10.1080/19345747.2016.1227412?journalCode=uree

  20.

*Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., ... & Roberts, G.

  (2014). Examining the effectiveness of a multicomponent reading comprehension

  intervention in middle schools: a focus on treatment fidelity. *Educational Psychology*

  *Review, 2*6(3), 425-449. doi:10.1007/s10648-014-9270-6

*Fong, A., Finkelstein, N., Jaeger, L., Diaz, R., & Broek, M. (2015). *Evaluation of the expository Reading and Writing Course: Findings from the Investing in Innovation development grant.* San Francisco, CA: WestEd.

*Gatti, G. & Miller, J. (2011) *Pearson SuccessMaker Reading efficacy study 2010-2011 final report.* Pittsburgh, PA: Gatti Evaluation, Inc.

*Gatti, G.G. (2016). *iLit 2013-2015: Longitudinal efficacy study final report*. Pittsburgh, PA, Gatti Evaluation.

*Gorard, S., N. Siddiqui S & Beng, H.S. (2015). *Accelerated reader: Evaluation report and executive summary*. London, England: EEF.

*Greenleaf, C. , Litman, C. , Hanson, T. , Rosen, R. , Boscardin, C. , et al. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal, 48*(3), 647-717. doi:10.3102/0002831210384839

*Hofstetter, C., Strick, B., Der-Martirosian, C., Ong-Dean, C., Long, P., & Lou, Y. (2011). *Striving Readers: Implementation and impact of the targeted and the whole-school interventions: Summary of year 4 (2009-2010). San Diego Unified School District, CA*. San Diego, CA: University of California, San Diego.

Hannafin, M. J., & Land, S. M. (1997). The foundations and assumptions of technology-enhanced learning environments. *Instructional Science, 25*, 167-202.

Herrera, S., Truckenmiller, A.J., & Foorman, B.R. (2016). *Summary of 20 years of research on the effectiveness of adolescent literacy programs and practices.* Washington, DC: U.S. Department of Education.

*IEE (2016). *Teacher Efectiveness Enhancement Programme: Evaluation report and executive summary.* York, England: Author.

*Jaciw, A. P., Schellinger, A. M., Lin, L., Zacamy, J., & Toby, M. (2016). *Effectiveness of internet-based reading apprenticeship Improving science education (iRAISE).* Palo Alto, CA: Empirical Education.

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

Kidron, Y., & Lindsay, J. (2014). *The effects of increased learning time on student academic and nonacademic outcomes: Findings from a meta-analytic review* (REL 2014-2015). Washington, DC: U. S. Department of Education, Institute of Education Science.

*Kim, J. S., Hemphill, L., Troyer, M., Thomson, J. M., Jones, S. M., LaRusso, M. D., & Donovan, S. (2017). Engaging struggling adolescent readers to improve reading skills. *Reading Research Quarterly, 52*(3), 357–382. https://doi.org/10.1002/rrq.171

*Kim, J.S., Olson, C.B., Scharcella, R., Kramer, J., Pearson, M., van Dyk, D., Collins, P., and Land, R.E. (2011). A randomized experiment of a cognitive strategies approach to text-based analytical writing for mainstreamed Latino English language learners in Grades 6 to 12. *Journal of Research on Educational Effectiveness, 4* (3), 231–263.

*King, B., & Kasim, A. (2015). *Rapid Phonics: Evaluation report and executive summary.* Millbank, England: Educational Endowment Foundation.

*Konstantopoulos, S., Miller, S. R., Ploeg, A. van der, & Li, W. (2016). Effects of interim assessments on student achievement: Evidence from a large-scale experiment. *Journal of Research on Educational Effectiveness, 9*(sup1), 188–208. https://doi.org/10.1080/19345747.2015.1116031

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A Meta-

    analytic review. *Review of Educational Research*, *86*(1), 42–78.

    doi:10.3102/0034654315581420

*Kushman, J., Hanita, M., & Raphael, J. (2011). *An experimental study of the Project CRISS*

    reading program on grade 9 reading achievement in rural high schools. Final Report

    NCEE 2011-4007. Washington, DC: National Center for Education Evaluation and

    Regional Assistance.

*Lang, L., Torgesen, J., Vogel, W., Chanter, C., Lefsky, E., & Petscher, Y. (2009). Exploring the

    relative effectiveness of reading interventions for high school students. *Journal of*

    Research on Educational Effectiveness, 2(2), 149–175.

    doi.org/10.1080/19345740802641535

*Lawrence, J., Francis, D., Pare-Blagoev, J., & Snow. C. (2016). The poor get richer:

    Heterogeneity in the efficacy of a school-level intervention for academic language.

    Journal of Research on Educational Effectiveness. Retrieved from

    www.tandfonline.com/doi/full/10.1080/19345747.**2016**.1237596

*Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of

    implementation of an academic vocabulary intervention for linguistically diverse students

    in urban middle schools. *Reading Research Quarterly, 45*(2), 196–228.

    doi:10.1598/RRQ.45.2.3

*Lesaux, N.K., Kieffer, M.J., Kelley, J.G., & Harris, J.R. (2014). Effects of academic vocabulary

    instruction for linguistically diverse adolescents: Evidence from a randomized field trial.

    American Educational Research Journal, 51(6), 1159–1194.

    doi:10.3102/0002831214532165

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

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K., & Busick, M. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSR 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncser.

*Little, C. A., McCoach, D. B., & Reis, S. M. (2014). Effects of differentiated reading instruction on student achievement in middle school. *Journal of Advanced Academics, 25*(4), 384–402. doi:10.1177/1932202X14549250

∗Lloyd, C., Edovald, T., Kiss, Z., Morris, S., Skipp, A., & Ahmed, H. (2015). Paired Reading: Evaluation report and exective summary. Education Endowment Foundation. Retrieved from http://socialwelfare.bl.uk/subject-areas/services-activity/education-skills/natcen/175224Paired_Reading.pdf

*Lord, P., Bradshaw, S., Stevens, E. and Styles, B. (2015). *Perry Beeches Coaching Programme: Evaluation Report and Executive Summary*. London, England: Education Endowment Foundation.

MacArthur, C. A., Ferretti, R. P., Okolo, C. M., & Cavalier, A. R. (2001). Technology applications for students with literacy problems: A critical review. *Elementary School Journal*, *101*(3), 273-301.

*Maxwell, B., Connolly, P., Demack, S., O'Hare, L., Stevens, A., & Clague, L. (2014). *TextNow Transition Programme: Evaluation report and executive summary*. London, England: Education Endowment Foundation. Retrieved from http://www.paulconnolly.net/publications/EEF_TextNow_2014.pdf

*May, H., & Robinson, M. (2007). *A randomized evaluation of Ohio's Personalized Assessment*

    Reporting System (PARS). Philadelphia, PA: University of Pennsylvania, CPRE Research

    Reports. Retrieved from http://repository.upenn.edu/cpre_researchreports/50

McCormick, C. B. (2003). Metacognition and learning. In W. M., Reynolds & G. E. Miller

    (Eds.), *Handbook of Psychology, Vol. 7, Educational Psychology* (pp. 79-102). Hoboken,

    NJ: Wiley.

McDonald, T., & Thornley, C. (2004). Literacy strategies for unlocking meaning in content area

    texts: Using student voices to inform professional development. *Thinking Classroom,*

    5(3), 7–14.

*Meisch, A., Hamilton, J., Chen, E, Quintanilla, P., Fong, P., Gray-Adams, K., Petta, I., &

    Thornton, N. (2011) *Striving Readers study: Targeted & whole-school interventions –*

    Year 5. Rockville, MD: Westat.

*Merrell, C., & Kasim, A. (2015). *Butterfly Phonics: Evaluation report and executive summary*.

    Millbank, England: Educational Endowment Foundation.

*Meyers, C. V., Molefe, A., Brandt, W. C., Zhu, B., & Dhillon, S. (2016). Impact results of the

    eMINTS professional development validation study. *Educational Evaluation and Policy*

    Analysis, 38 (3), 455-476. doi:10.3102/0162373716638446

Murphy, R.F., Penuel, W.R., Means, B., Korbak, C., Whaley, A., & Allen, J.E. (2002, April). *E-*

    DESK: A review of recent evidence on the effectiveness of discrete educational software

    (SRI Project No. 11063). Menlo Park, CA: SRI International.

National Center for Education Statistics (NCES) (2016). *National Assessment of Educational*

    Progress. Washington, DC: Author.

*Newman, D. L., Kundert, D. K., Spaulding, D. T., White, S. P., & Gifford, T. A., (2012). *Striving Readers Project: New York State Department of Education/New York City Department of Education: Striving Readers Local Evaluation*. Albany, NY: University at Albany, State University of New York.

OECD (2010). *PISA 2009 results: Learning trends: Changes in student performance since 2000 (Volume V).* Paris, France: Author. doi: 10.1787/9789264091580-en

OECD (2013). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris, France: Author. doi: 10.1787/9789264204256-en

*Olson, C., Kim, J., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., & ... Land, R. (2012). Enhancing the interpretive reading and analytical writing of mainstreamed English learners in secondary school: Results from a randomized field trial using a cognitive strategies approach. *American Educational Research Journal, 49*(2), 323–355. doi: 10.3102/0002831212439434

*Olson, C.B., Matuchniak, T., Chung, H.Q., Stumpf, R., & Farkas, G**.** (April 2016). Reducing achievement gaps in academic writing for Latinos and English learners in grades 7-12. *Journal of Educational Psychology*. doi:10.1037/edu0000095.

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.

Pressley, M. (2003). Psychology of literacy and literacy instruction. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of Psychology, Vol. 7* (pp. 333-356). Hoboken, NJ: Wiley.

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*, Second Edition. Thousand Oaks: Sage Publications, Inc.

Roseth, C., Johnson, D., & Johnson, R. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134*(2), 223–246. doi:10.1037/0033-2909.134.2.223

*Rutt, S., Kettlewell, K., & Bernardenelli, D. (2015). *Catch Up® Literacy: Evaluation report and executive summary.* Millbank, England: Educational Endowment Foundation.

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.

*Schenck, A., Feighan, K., Coffey, D., & Rui, N. (2011). *Memphis Striving Readers project: Evaluation report, Year 4.* New York, NY: Research for Better Schools & RMC Research Corporation.

*Schenck, A., Jurich, S., Frye, M., Lammert, J., Sayko, S, Najera, K. &…Willard, T. (2012) *Evaluation report/Impact study: Virginia Striving Readers Intervention Initiative*. Arlington, VA: RMC Research Corporation.

*Schiller, E., Wei, X., Thayer, S., Blackorby, J., Javitz, H., & Williamson, C. (2012). *A randomized controlled trial of the impact of the Fusion Reading Intervention on reading achievement and motivation for adolescent struggling readers*. Evanston, IL: Society For Research On Educational Effectiveness.

*Schoenbach, R., Greenleaf, C., Cziko, C., & Hurwitz, L. (1999).The Reading Apprenticeship framework. *In Reading for understanding: A guide to improving reading in middle and high school classrooms* (pp. 17–42). San Francisco: Jossey-Bass.

*Shannon, L., & Grant, B-J. (2015). *An evaluation of the Achieve 3000 programs.*
    Charlottesville, VA: Magnolia Consulting.

*Shapley, K., Sheehan, D., Maloney, C., & Caranikas-Walker, F. (2009). *Evaluation of the
    Texas Technology Immersion Pilot: Final outcomes for a four-year study (2004-05 to
    2007-08).* Austin, TX: Texas Center for Educational Research.

*Sibieta, L. (2016). *REACH: Evaluation report and executive summary*. London, England:
    Education Endowment Foundation.

*Simmons, D., Fogarty, M., Oslund, E., Simmons, L., Hairrell, A., Davis, J., et al. (2014).
    Integrating content knowledge-building and student-regulated comprehension practices in
    secondary English language arts classes. *Journal of Research on Educational
    Effectiveness, 7*, 309-330. doi:10.1080/19345747.2013.836766

*Simon, A., Tunik, J., Alemany, J., Zhu, J., Zacharia, J., Ramsay, L., & …Mendes, R. (2011).
    Chicago Public Schools Striving Readers Initiative: Year four evaluation report. New
    York, NY: Metis Associates.

*Sivin-Kachala, J., & Bialo, E. (2012). *Program evaluation research on Expert 21 in grades 6-8
    for Scholastic Education.* New York: IESD.

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

Slavin, R.E. (2015). Cooperative learning in schools. In J. Williams (Ed.), *International
    encyclopedia of the social and behavioral sciences (2nd ed.),* 881-886. Amsterdam:
    Elsevier.

* Slavin, R., Chamberlain, A., Daniels, C., & Madden, N. (2009). The Reading Edge: a

    randomized evaluation of a middle school cooperative reading program. *Effective*

    Education, 1(1), 13-26.

Slavin, R.E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle

    and high schools: A best evidence synthesis. *Reading Research Quarterly, 43* (3), 290-

    322*.*

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling

    readers: A best-evidence synthesis. *Educational Research Review, 6*, 1-26.

*Somers, M. A., Corrin, W., Sepanik, S., Salinger, T., Levin, J., Zmach, C., & Wong, E. (2010).

    The Enhanced Reading Opportunities Study final report: The impact of supplemental

    literacy courses for struggling ninth-grade readers. NCEE 2010-4021. Washington, DC:

    National Center for Education Evaluation and Regional Assistance.

*Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2012). *Springfield-Chicopee School District*

    Striving Readers (SR) Program: Final report years 1–5: Evaluation of implementation

    and impact. Providence, RI: Brown University: The Education Alliance.

*Stevens, R. J., & Durkin, S. (1992a, b). *Using student team reading and student team writing in*

    middle schools: Two evaluations. Baltimore, MD: Johns Hopkins University, Center for

    Research and Reform in Education.

Styles, B., Stevens, E., Bradshaw, S., & Clarkson, R. (2014). *Vocabulary Enrichment*

    Intervention Programme. Millbank, England: EEF.

*Swanlund, A., Dahlke, K., Tucker, N., Kleidon, B., Kregor, J., Davidson-Gibbs, D., &

    Hallberg, K. (2012). *Striving Readers: Impact study and project evaluation report:*

Wisconsin Department of Public Instruction (with Milwaukee Public Schools).

Naperville, IL: American Institutes for Research.

*Vaden-Kiernan, M., Caverly, S., Bell, N., Sullivan, K., Fong, C., Atwood, E... & Jones, D.

(2012). *Louisiana Striving Readers: Final evaluation report*. Austin, TX: SEDL.

*Vaughn, S., Klingner, J. K., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S. S.,

& Stillman-Spisak, S. J. (2011). Efficacy of Collaborative Strategic Reading with middle

school students. *American Educational Research Journal, 48*(4), 938-964.

doi:10.3102/0002831211410305

*Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A.-M. (2017).

Improving content knowledge and comprehension for English language learners:

Findings from a randomized control trial. *Journal of Educational Psychology, 109*(1),

22–34.

*Vaughn, S., Roberts, G., Klingner, J. K., Swanson, E. A., Boardman, A., Stillman-Spisak, S. J.,

... & Leroux, A. J. (2013). Collaborative strategic reading: Findings from experienced

implementers. *Journal of Research on Educational Effectiveness*, *6*(2), 137-163.

doi:10.1080/19345747.2012.741661

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of

Statistical Software,36*(3), 1-48. URL: http://www.jstatsoft.org/v36/i03/

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. K., Metz, K., Murray, C. S., & Roberts, G. (2013).

Extensive reading interventions for students with reading difficulties after grade 3.

Review of Educational Research, *83*(2), 163–195. doi: 10.3102/0034654313477212

Webb, N. M., & Mastergeorge, A. M. (2003). The development of students' helping behavior and learning in peer-directed small groups. *Cognition and Instruction, 21,* 361-428.

*West, M. R., Morton, B. A., & Herlihy, C. M. (2016). *Achievement Network's Investing in Innovation expansion: Impacts on educator practice and student achievement.* Cambridge, MA: Harvard.

What Works Clearinghouse (2015). *Procedures and standards handbook.* Retrieved from http://ies.ed.gov/ncee/wwc/Protocols.

What Works Clearinghouse. (2016). *Adolescent reading.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

*Wijekumar, K., Meyer, B. J. F., & Lei, P. (2017). Web-based text structure strategy instruction improves seventh graders' content area reading comprehension. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000168

Zeif, G., Lauver, S., & Maynard, R. (2006). *The impacts of afterschool programs on student outcomes: A systematic review for the Campbell Collaboration.* Retrieved August 3, 2017 from https://www.campbellcollaboration.org/library/impacts-of-after-school-programmes-on-student-outcomes.html.

Table 1
Tutoring Interventions for Struggling Readers

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
				One-to-One Tutoring				
Catch Up® Literacy								
Rutt et al. (2015)	Student randomized	30 weeks	Students: 557 (286 E, 271 C)	Years 6-7	Students reading at least one year below grade level from 15 schools mainly in urban areas across England. 21% FRL.	NGRT	+0.16	+0.16
The Perry Beeches Coaching Programme								
Lord et al. (2015)	Student randomized	1 year	Students: 291 (149 E, 142 C)	Year 7	Students from 4 secondary schools in Birmingham, England who were reading at least one year below grade level. 55% W, 20% ELL, 58% FRL.	GL Assessment - Progress in English	+0.36*	+0.36*
REACH (tutoring)								
Sibieta (2016)	Student randomized	20 weeks	Students: 202 (70 REACH, 69 REACH + LC, 63 C)	Years 7-8	Lowest readers in 27 disadvantaged secondary schools in or near Leeds, England. 68% W, 32% non-white, 63% SPED, 24% ELL, 31% FRL.	NGRT		+0.42*
						REACH	+0.33*	
						REACH + Language Comprehension	+0.51*	
				Small-Group Tutoring				
Butterfly Phonics								
Merrell & Kasim (2015)	Student randomized	4 months	Students: 310 (161 E, 149 C)	Year 7	Students from 6 secondary schools in London, England who were reading at least one year below grade level. 78% W, 16% AA, 35% SPED, 64% ELL, 51% FRL.	NGRT	+0.30*	+0.30*
Rapid Phonics combined with Sound Discovery								
King & Kasim (2015)	Student randomized	12 weeks	Students: 178 (86 E, 92 C)	Years 6-7	Students from 22 primary and 13 secondary schools in Norfolk, England who were reading at least one year below grade level. 50% W, 50% SPED, 50% ELL, 50% FRL.	NGRT	-0.05	-0.05

Volunteer Tutoring								
TextNow Transition Program								
Maxwell et al. (2014)	Student randomized	15 weeks	Students: 391 (199 E, 192 C)	Years 6-7	Students reading below grade level from 53 primary schools and 29 secondary schools across England. 16% ELL, 25% FRL.	NGRT	-0.06	-0.06
Cross-Age Tutoring								
Paired Reading								
Lloyd et al. (2015)	Cluster randomized	16 weeks	Classes: 120 (58 E, 62 C) Students: 1306 (625 E, 681 C)	Year 7	10 schools in North of England. 96% W, 20% FRL	NGRT	-0.02	-0.02

Notes for Tables 1-5

CAHSEE: California High School Exit Examination, CAT: California Achievement Test, CEM: Centre for Evaluation & Monitoring, CST-ELA: California Standards Test – English Language Arts, CTBS: Comprehensive Test of Basic Skills, DRP: Degree of Reading Power, EAL: English as a second language, ELA: English Language Arts, EPT: Early Placement Test (California), ETS: Educational Testing Service, FCAT: Florida Comprehensive Assessment Test, GORT: Gray Oral Reading Test, GSRT: Gray Silent Reading Test, GRADE: Group Reading Assessment and Diagnostic Evaluation, GMRT: Gates-MacGinitie Reading Tests, iLEAP, Louisiana State Reading Assessment, ISAT: Illinois Student Achievement Test, ISTEP+: Indiana State Test of Educational Proficiency, ITBS: Iowa Test of Basic Skills, KCCT: Kentucky Core Content Test, MAP : Measure of Academic Progress, MCAS: Massachusetts Comprehensive Assessment System, MEAP: Michigan Educational Assessment Program, MSP: Measurements of Student Progress state reading assessment, NGRT : New Group Reading Test (U.K.), NJASK: New Jersey State Test; NYS-ELA: New York State English Language Arts, NWEA: Northwest Evaluation Association, OAKS: Oregon Assessment of Knowledge and Skills, ORF: Oral Reading Fluency, RISE: Reading Inventory and Scholastic Evaluation, SAT 10: Stanford Achievement Test 10, SDRT-4: Stanford Diagnostic Reading Test 4, STAAR: State of Texas Assessments of Academic Readiness, SOL: Virginia Standards of Learning English/Reading, SWE: Sight Word Efficiency, TAKS: Texas Assessment of Knowledge and Skills, TCAP: Transitional Colorado Assessment Program, TOSREC, Test Of Silent Reading Efficiency and Comprehension, WJ III: Woodcock-Johnson III

A=Asian, AA=African-American, H=Hispanic, W=White, FRL=Free/Reduced Lunch, ELL=English Language Learner, LD=Learning Disabilities, LEP=Limited English-proficient, SPED=Special Education.

*p<.05 at the appropriate level of analysis (cluster or individual).

Table 2

Programs Emphasizing Social Support for Learning (No Extra Time)

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
				Cooperative Learning				
The Reading Edge/Student Team Reading								
Slavin, et al., (2009)	Student randomized	1 year	Students: 788 (405 E, 383 C) (2 cohorts)	6	2 Title I rural, mostly White middle schools in West Virginia and Florida. 90% W, 61% FRL, 15% SPED.	GMRT	+0.15[*]	+0.15*
Stevens & Durkin (1992a)	Cluster quasi-experiment	1 year	Schools: 5 (2 E, 3 C) Students: 3986 (1798 E, 2188 C)	6-8	High poverty, majority AA middle schools in Baltimore, Maryland.	CAT Comprehension	+0.34	+0.40
						CAT Vocabulary	+0.46	
Stevens & Durkin (1992b)	Cluster quasi-experiment	1 year	Schools: 6 (3 E, 3 C) Classes: 59 (20 E, 34 C) Students: 1223 (455 E, 768 C)	6	Middle schools in Baltimore, Maryland. 75% AA, 58% FRL.	CAT Comprehension	+0.13	+0.06
						CAT Vocabulary	-0.02	
Talent Development High School (Strategic Reading and Student Team Writing)								
Balfanz et al., 2004	Cluster quasi-experiment	1 year	Schools: 6 (3 E, 3 C) Teachers: 20 E Students: 457 (257 E, 200 C)	9	High-poverty high schools in Baltimore, MD. 89% AA, 10% W, >90% FRL.	CTBS Terra Nova	+0.32	+0.32
Collaborative Strategic Reading (CSR)								
Denver Public Schools (2016)	Cluster randomized	1 year	Schools : 16 Students : 5660 (3101 E, 2559 C) 3 cohorts	6-8	16 middle schools in Denver, CO. 62% H, 19% W, 11% AA, 30% ELL, 11% SPED, 76% FRL.	GMRT	+0.03	+0.03
						State Reading Test (TCAP)	+0.02	

Study	Design	Duration	Sample	Grades	Setting	Measure	ES	
Vaughn et al. (2011)	Cluster randomized	18 weeks	Classes: 61 (34 E, 27 C) Students: 782 (400 E, 382 C)	7, 8	6 middle schools from 3 school districts in Colorado and Texas. 43% W, 51% H, 52% FRL.	GMRT Comprehension	+0.12	+0.04
						AIMSweb maze	-0.08	
						TOSREC	+0.07	
Vaughn et al. (2013)	Cluster randomized	20 weeks	Classes: 48 (26 E, 22 C) Students: 472	7, 8	Same teachers and schools as in Vaughn et al. (2011). 51%W, 42%H, 6% LEP, 7% SPED.	GMRT	+0.10	+0.10
						TOSREC	+0.11	
Social-Emotional Programs								
BARR (Building Assets Reducing Risk)								
Borman et al. (2017)	Student randomized	1 year	Students: 2172 (981 E, 1191 C)	9	6 schools in California (3), Maine (2), Minnesota (1). 71% minority, 21% ELL, 70% FRL.	NWEA	+0.08*	+0.08*
Corsello & Sharma (2015)	Student randomized	1 year	Students: 495 (261 E, 234 C)	9	1 school in Southern California 52% W, 37% H, 11% AA, 17% ELL, 68% FRL.	NWEA	+0.14*	+0.14*

Table 3A

Programs Incorporating Technology: No Extra Time

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
ITSS (Intelligent Tutoring System for the Structure Strategy)								
Wijekuma, et al. (2017)	Cluster randomized	6-7 months	Classrooms: 108 (59 E, 49 C) Students: 2489	7	25 rural and suburban schools in the Northeast. 92% W, 8% minority, 42% FRL.	GSRT	+0.18*	+0.18*
The Thinking Reader								
Drummond et al. (2011)	Cluster randomized	1 year	Teachers: 90 (48 E, 42 C) Students: 2149 (1154 E, 986 C)	6	32 high-poverty schools from 16 districts in Connecticut, Massachusetts, and Rhode Island. 37% W, 28% H, 11% SPED, 10% ELL, 71% FRL.	GMRT	+0.01	+0.01
						Comprehension	+0.03	
						Vocabulary	-0.04	
SuccessMaker								
Gatti (2011)	Cluster randomized	1 year	Classes: 22 (11 E, 11 C) Students: 453 (254 E, 199 C)	7	Students reading at least 1 year below grade level from 5 schools in 8 urban and suburban school districts in 4 states (Arizona, Kansas, Michigan, Missouri). 51% W, 53% FRL.	GRADE	+0.11	+0.11
						Comprehension	+0.10	
						Vocabulary	+0.12	
Achieve3000								
Shannon & Grant (2015)	Cluster randomized	1 year	Teachers: 33 (16 E, 17 C) Students: Grade 6: 494 (263 E, 231 C) Grade 9: 248 (122 E, 126 C)	6, 9	12 schools in 4 suburban and city districts across the US. 37% H, 67% W, 12% SPED, 12% ELL, 62% FRL.	GMRT		+0.29*
						Grade 6	+0.22	
						Grade 9	+0.44	
eMINTS (Enhancing Missouri's Instructional Networked Teaching Strategies)								
Meyers et al. (2015)	Cluster randomized	3 years	Schools: 59 (20 E, 20 E+, 19 C) Students: 3295 (1208 E, 1216 E+, 871 C)	6-8	Rural middle schools across Missouri. 93%W.	MAP		-0.06
						eMINTS	-0.04	
						eMINTS + Intel	-0.08	

READ 180 (No extra time)

Meisch et al. (2011)	Cluster randomized	3 years	Schools: 19 (10 E, 9C) Students 1023 (552 E, 471 C) (4 cohorts)	6-8	Students from 19 Title I middle schools across Newark, NJ who scored below proficient on state reading tests. 55% AA, 42% H, 13% ELL, 44% SPED, 62% FRL.	SAT 10	
						Comprehension +0.06	+0.06
						Vocabulary +0.05	
Lang et al. (2009)	Student randomized	1 year	High Risk Students: 190 (100 E, 90 C) Moderate Risk Students: 409 (207 E, 202 C)	9	Students from 7 comprehensive high schools in a large district in Florida who were reading below 4th grade levels (high risk) or between 4th and 6th grade levels (moderate risk). 19% H, 19% AA, 41% FRL.	FCAT Reading	
						High Risk -0.27*	+0.12
						Moderate Risk +0.30*	

System 44 (No extra time)

Beam et al. (2011)	Student randomized	1 year	Students: 147 (75 E, 72 C)	6-8	4 schools from one large suburban school district in southern California. 7% SPED, 63% W, 33% FRL.	TOSREC -0.24*	-0.14
						CST -0.04	

Prentice Hall Literature

Eddy et al. (2010)	Cluster randomized	1 year	Teachers: 29 (16 E, 13 C) Students: 1518 (744 E, 774 C)	7-10	8 schools from California, Oregon, Arizona, Ohio. 6 suburban and 2 rural areas. 55% H, 15% AA.	GMRT -0.10	-0.10

Comprehensive Circuit Training (CCT)

Fogarty et al. (2014)	Cluster randomized	1 semester	Classes: 61 (30 E, 31 C) Students: 859 (411 E, 448 C)	6-8	3 middle schools from 3 districts in the Southwest. 43% H, 35% W, 22% AA, 9% ELL, 6% SPED, 67% FRL.	GMRT +0.12	+0.12
Fogarty et al. (2016)	Cluster randomized	1 semester	Classes: 16 (9 E, 7 C) Students: 228 (112 E, 116 C)	6-8	3 middle schools in Texas. 30% AA, 27% W, 26% H, 9% SPED, 67% FRL.	STAAR +0.09 GMRT +0.12 GRADE +0.11 GORT +0.18	+0.15

						TOSREC	+0.24*	
Texas Technology Immersion Pilot (eTxTip)								
Shapley et al. (2009)	Cluster quasi-experiment	2-3 years	Schools: 42 (21 E, 21 C) Students: 10,234 (4,767 E, 5,467 C) 3 cohorts	6-9	Rural, suburban, and urban middle schools across Texas. 70% H, 22% W, 7% AA, 15% LEP, 70% FRL.	TAKS	+0.06	+0.06

Table 3B

Programs Incorporating Technology: Extra Time

READ 180 (Extra time)

Schenck et al. (2011)	Student randomized	2 years	Students: 1295 (556 E, 739 C) (3 cohorts)	6-8	Students from 8 Title I middle schools in Memphis City, TN who tested in the bottom quartile of the reading/ELA portion of the state test. 93% AA, 6% ELL, 92% FRL.	ITBS Total Reading	+0.02	+0.02
						Comprehension	-0.01	
						Vocabulary	+0.06	
Swanlund et al. (2012)	Student randomized	1 year	Students: 619 (335 E, 284 C)	6-9	Students from 5 Title I schools in Milwaukee who performed below proficient on standardized reading tests. 70% AA, 19% H, 36% SPED, 8% ELL, 88% FRL.	MAP Reading	+0.14*	+0.14*
Sprague et al. (2012)	Student randomized	1 year	Students:456 (231 E, 225 C) (5 cohorts)	9	Students from 5 Title I eligible high schools in western Massachusetts who tested between a 4th and 6th grade reading level 73% minority, 19% SPED, 72% FRL.	SDRT-4	+0.18*	+0.18*

System 44 (Extra time)

Beam & Faddis (2012)	Student randomized	1 year	Students: 145 (70 E, 75 C)	6-8	4 middle schools from one urban district in Michigan. 78% AA, 53% SPED, 96% FRL.	TOSREC	+0.20*	+0. 20*

Accelerated Reader

Gorard et al. (2015)	Student randomized	22 weeks	Students: 349 (166 E, 183 C)	Year 7	Low-achieving students in four schools in England. 88% W, 23% SPED, 35% FRL.	NGRT	+0.24*	+0.24*

Passport Reading Journeys

Study	Design	Duration	Sample	Grades	Population	Measure	Effect	Overall
Vaden-Kiernan et al. (2012)	Student randomized	1 year	iLEAP Students: 1102 (548 E, 554 C) GRADE Students: 983 (485 E, 498 C)	6,7	Students from 10 Title I middle schools across Louisiana who scored below proficient on state standardized reading assessments. 76% minority, 15% SPED, 88% FRL.	iLEAP Reading	-0.01	+0.12*
						GRADE Overall	+0.27*	
						Vocabulary	+0.13*	
						Comprehension	+0.30*	
Schenck et al. (2012)	Student randomized	1 year	SOL Students: 701 (343 E, 358 C) GMRT Students: 568 (279 E, 289 C)	7,8	Students from 9 middle schools in urban, high-poverty settings across Virginia who scored at least two years below grade level on reading tests. 68% AA, 24% SPED, 8% ELL, 88% FRL.	SOL Reading	+0.06	+0.06
						GMRT Overall	+0.06	
						Comprehension	+0.05	
						Vocabulary	+0.07	
Dimitrov et al. (2012)	Student randomized	1 year	Students: 460 (238 E, 222 C)	9	Students from 6 Title I high schools across Illinois who performed in the bottom two quartiles on the EXPLORE reading assessment. 58% AA, 5% H, 30% W, 18% SPED, 85% FRL.	GMRT	+0.02	-0.03
						EXPLORE	-0.09	

iLit

Study	Design	Duration	Sample	Grades	Population	Measure	Effect	Overall
Gatti (2016)	Student randomized	2 years	Students: 213 (114 E, 99 C)	7-8	Students reading below grade level from 6 middle schools in AZ, CA, CO, MI, NJ & NY. 53% H, 22% W, 17% AA, 26% LEP, 13% SPED, 80% FRL.	GRADE Total	+0.09	+0.09
						Comprehension	+0.12	
						Vocabulary	+0.01	

Table 4A

Metacognitive Strategy Approaches: No Extra Time

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
Teacher Effectiveness Enhancement Programme (TEEP)								
Institute for Effective Education (2016)	Cluster Randomized	1.5-2 years	Schools: 45 (23 E, 22 C) Students: 10,385 (5327 E, 8058 C) 12 cohorts	Year 9	Low-performing secondary schools across England. 30% FRL, 16% ELL.	CEM Insight-English	-0.04	-0.04
Word Generation								
Lawrence et al. (2016)	Cluster randomized	1 year	Schools: 44 (25 E, 19 C) Students: 8466 (4796 E, 3670 C)	6-8	Schools in 2 Northeast, 1 Western urban districts. 81% FRL.	GMRT		+0.05
						Reading Comprehension	+0.07	
						Vocabulary	0.00	
Chicago Striving Readers								
Simon et al. (2011)	Cluster randomized	1 to 3 years	Schools: 59 (29 E, 30 C) Students: 8127 (4074 E, 4053 C) (2 cohorts)	6-8	Middle schools across Illinois. 58% AA, 35% H, 9% SPED, 96% FRL.	ISAT Reading	-0.01	-0.01
Project CRISS								
Kushman et al. (2011)	Cluster randomized	1 year	Schools: 49 (23 E, 26 C) Students: 4959 (2460 E, 2499 C)	9	Schools in rural and urban fringe towns across 6 Northwest states. 79% W, 15% FRL.	SDRT	+0.05	+0.05

Reading Intervention through Strategy Enhancement (RISE)

Lang et al. (2009)	Student randomized	1 year	High Risk Students: 194 (104 E, 90 C)	9	Students from 7 comprehensive high schools in a large district in Florida who were reading below 4th grade levels (high risk) or between 4th and 6th grade levels (moderate risk). 19% H, 19% AA, 43% FRL.	FCAT Reading		+0.16*
						High risk	-0.06	
			Moderate Risk Students: 406 (204 E, 202 C)			Moderate risk	+0.27*	

REACH

Lang et al. (2009)	Student randomized	1 year	High Risk Students: 181 (91 E, 90 C)	9	Students from 7 comprehensive high schools in a large district in Florida who were reading below 4th grade levels (high risk) or between 4th and 6th grade levels (moderate risk). 20% H, 20% AA, 43% FRL.	FCAT Reading		-0.02
						High risk	-0.19	
			Moderate Risk Students: 401 (199 E, 202 C)			Moderate risk	+0.06	

Expert 21

Sivin-Kachala & Bialo (2012)	Cluster randomized	1 year	Teachers: 6 (3 E, 3 C) Students: 276 (137 E, 139 C)	6-8	1 middle school in urban New Jersey. 71% H, 27% AA, 100% FRL.	NJASK	+0.20	+0.15
						Language & Literature	+0.22	
						Reading Comp.	+0.18	
						GMRT Comprehension	+0.10	

Expository Reading and Writing Course (ERWC)

Fong et al. (2015)	Within-school quasi-experimental	1 year	Students: 6618 (3309 E, 3309 C)	12	24 schools across California (15 urban, 3 rural, and 6 suburban). 45% H, 27% A, 24% W.	EPT	+0.13*	+0.13*

Schoolwide Enrichment Model-Reading (SEM-R)

Little et al. (2014)	Cluster randomized	1 year	Teachers: 47 (27 E, 20 C) Students: 2028 (1198 E, 830 C)	6-8	4 high-poverty, low-achieving middle schools. 48% AA, 32% H, 67% FRL.	GMRT Comprehension	+0.10	+0.10

Vocabulary Enrichment Intervention Programme								
Styles et al. (2014)	Student randomized	6 months	Schools: 11 Students: 570 (282 E, 288 C)	7	Students reading below grade level from schools in England, 28% FRL.	NGRT	+0.06	+0.06
SIM: Content Literacy Curriculum (CLC)								
Corrin et al. (2012)	Cluster randomized	1, 2 years	Schools: 28 (15 E, 13 C) Students: 1 year: 5011 (2975 E, 2036 C) 2 years: 4546 (2908 E, 1638 C)	9-10	High-poverty, low-achieving urban high schools across the Midwest. 47% AA, 9% ELL, 11% SPED, 65% FRL.	GRADE 1 year Comprehension +0.06 Vocabulary +0.09 2 years Comprehension +0.10 Vocabulary +0.10		+0.09
SIM: Adolescent Literacy Model (ALM)								
Cantrell et al. (2011)	Cluster quasi-experimental	4 years	Middle schools: 18 (9 E, 9 C) High schools : 20 (10 E, 10 C)	6-12	Schools in 7 rural school districts in Kentucky. >90% W. Middle schools: 50-56% FRL. High schools 34-44% FRL.	KCCT Reading Middle schools +0.08 High schools +0.12		+0.10
SIM: Xtreme Reading (No extra time)								
Faddis et al. (2011)	Student randomized	1 year	Students GRADE: 822 (401 E, 421 C) 4 Cohorts Students OAKS: 954 (472 E, 482 C)	7-8	6 Title 1 middle schools in Portland OR who were reading at least 2 years below grade level. 34% H, 23% AA, 27% SPED, 34% ELL.	GRADE +0.29* Comprehension +0.32* Vocabulary +0.20* OAKS +0.12		+0.20*
Reading Apprenticeship								
Greenleaf et al. (2011)	Cluster randomized	3 years	Schools: 78 (39 E, 39 C) Teachers: 111 (56 E, 55 C)	9-11	Biology teachers in 78 California schools. 48% H, 31% W, 19% ELL, 41% FRL.	DRP -0.04 CST ELA +0.10 Reading Comprehension +0.13		+0.03

Reading Apprenticeship Improving Secondary Education (RAISE)								
Fancsali et al. (2015)	Cluster quasi-experimental	1 to 2 years	Schools: 42 (22 E, 20 C) Students: 10173 (5531 E, 4642 C) (2 cohorts)	9-12	High schools in California and Pennsylvania. 49% AA, 33% H, 10% ELL, 40% FRL.	ETS Literacy Achievement Assessment	+0.14	+0.14
iRAISE								
Jaciw et al. (2016)	Cluster randomized	1 year	Schools: 26 Teachers: 69 (35 E, 34 C) Students: 1468 (751 E, 717 C)	9-12	High schools in Michigan and California. 73% W, 16% AA, 52% FRL.	ETS Literacy Assessment	0.00	0.00
Content Knowledge-Building and Student-Regulated Comprehension Practices								
Simmons et al. (2014)	Cluster randomized	1 semester	Classes: 65 (36 E, 29 C) Students: 786 (413 E, 373 C)	7-10	6 Title I schools (3 middle, 3 high) from 3 districts in one state in the Southwest. 36% H, 31% AA, 71% FRL.	GMRT Comprehension	-0.01	-0.01

Table 4B

Metacognitive Strategy Approaches: Extra Time

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
SIM: Xtreme Reading (Extra time)								
Somers et al. (2010)	Student randomized	1 year	GRADE Students: 2329 (1341 E, 988 C) (2 cohorts)	9	Students from 17 high schools across multiple districts who were reading 2-5 years below grade level. 31% H, 47% AA, 67% FRL.	GRADE		+0.06
						Comprehension	+0.05	
						Vocabulary	+0.03	
			State Test Students: 1191 (2 cohorts)			State Tests ELA	+0.08	
Faddis et al. (2011)	Student randomized	1 year	GRADE: Students: 757 (355 E, 402 C) (4 cohorts)	9-10	Students from 4 Title I high schools in Portland, OR who were reading at least 2 years below grade level. 26% H, 35% AA, 24% SPED, 20% ELL.	GRADE	+0.12*	+0.08
						Comprehension	+0.15*	
						Vocabulary	+0.07	
			OAKS Students: 514 (260 E, 254 C) (4 cohorts)			OAKS (10th grade only)	+0.02	
Sprague et al. (2012)	Student randomized	1 year	Students: 448 (223 E, 225 C) (5 cohorts)	9	Students from 5 Title I high schools in western Massachusetts who were reading between a 4th and 6th grade level. 75% minority, 22% SPED, 75% FRL.	SDRT-4	0.00	0.00
SIM: Learning Strategies Curriculum (LSC)								
Cantrell et al. (2016)	Student randomized	1 year	Grade 6 Students: 1135 (605 E, 530 C) (4 cohorts)	6, 9	Students from 21 middle and high schools across multiple rural districts in Kentucky who were reading at least 2 years below grade level. 88% W, 26% SPED, 62% FRL.	GRADE		+0.10*
						Grade 6	+0.08	
			Grade 9 Students: 1128, (593 E, 535 C) (4 cohorts)			Grade 9	+0.12*	

SIM: Fusion Reading

Schiller et al. (2012)	Student randomized	1 year	GRADE: Students: 581 (285 E, 296C)	6-10	Students from 7 schools (4 middle, 3 high) across 3 school districts in Michigan who scored between the 5th and 35th percentile on a reading screening test. 81% AA, 13% SPED.	GRADE	+0.05	+0.07
			MEAP: Students: 256 (118 E, 138 C)	6-7		Comprehension	+0.08	
						Vocabulary	0.00	
						MEAP Reading	+0.11	

Every Classroom, Every Day (ECED) /Literacy Matters

| Early et al. (2016) | Cluster randomized | 2 years | Schools : 20 (10 E, 10 C) Students: 8250 (3935 E, 4315 C) | 9-10 | 20 high schools from 5 districts, 4 states (Arizona, Tennessee New York California). 51% H, 24% AA 22% ELL, 76% FRL. | State test ELA | +0.06 | +0.06 |

Reading Apprenticeship Academic Literacy (RAAL)

Somers et al. (2010)	Student randomized	1 year	GRADE: Students: 2255 (1331 E, 924 C) (2 cohorts)	9	Students from 17 high schools across multiple districts who were reading 2-5 years below grade level. 31% H, 47% AA, 67% FRL.	GRADE Overall	+0.08	+0.10*
			State Test: Students: 1053 (2 cohorts)			Comprehension	+0.12*	
						Vocabulary	0.00	
						State Tests ELA	+0.15*	

Reading Excellence: Word Attack and Rate Development Strategies (REWARDS)

Newman & Kundert (2012)	Student randomized	1 year	NYS: Students: 517 (253 E, 264 C)	7	Students from 11 Title I middle schools across 4 boroughs of New York City who performed below proficient on state standardized reading test. 64% H, 22% AA, 12% A, 95% FRL.	NYS ELA	+0.15	+0.09
			GMRT: Students: 469 (232 E, 237 C) (10 schools)			GMRT Total	+0.02	
						Comprehension	-0.01	
						Vocabulary	+0.08	

Kentucky Cognitive Literacy Model (KCLM)

| Cantrell et al. (2012) | Student randomized | 1 year | Students: 485 (232 E, 253 C) | 9 | Students from 9 high schools in 9 districts who were reading at least two years below grade level. 88% W, 16% SPED, 62% FRL. | GRADE | -0.06 | -0.06 |

Read to Achieve

Deussen et al. (2012)	Student randomized	1 year	**Read to Achieve + PhonicsBlitz** GMRT Students: 63 (32 E, 31 C) MSP Students: 76 (37 E, 39 C) **Read to Achieve** GMRT Students: 295 (144 E, 151 C) MSP Students: 325 (192 E, 191 C)	6, 8	Students from 5 Title I middle schools & 1 junior high school from 3 districts in Western Washington who were reading at least two years below grade level. 43% W, 23% ELL, 58% FRL.	**Read to Achieve + PhonicsBlitz** GMRT +0.13 MSP +0.11 **Read to Achieve** GMRT +0.02 MSP +0.16	+0.10

Strategic Adolescent Reading Intervention (STARI)

Kim et al. (2016)	Student randomized	1 year	Schools: 8 Students: 398 (170 E, 228 C)	6-8	Schools in 2 urban, 2 rural districts in Massachusetts. Students were below 30th percentile on prior year MCAS. 49% W, 26% H, 19% AA, 13% ELL, 69% FRL.	**RISE** Reading Comp +0.08 Basic Reading +0.21* Vocabulary +0.16	+0.15

Table 4C

Metacognitive Strategy Approaches: Focus on English Language Learners (No Extra Time)

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
Pathway								
Kim et al. (2011); Olson et al. (2012)	Cluster randomized	1 year	Teachers: 161 (79 E, 82 C) Students: 4459 (2200 E, 2259C) (2 cohorts)	6-11	15 schools (9 middle, 6 high) from a large school district in California. Eligible students: mainstreamed Latino ELLs. 95% H, 88% ELL, 79% FRL.	CST ELA	+0.07*	+0.07*
Olson et al. (2016)	Cluster randomized	1 year	Teachers: 16 (9 E, 7 C) Students: 575 (313 E, 262 C)	10	Schools in Anaheim, CA. 68% H, 18% A, 12% W, 20% ELL, 71% FRL.	CAHSEE	+0.19	+0.19
Academic Language Instruction for All Students (ALIAS)								
Lesaux et al. (2010)	Cluster randomized	18 weeks	Classes: 21 (13 E, 8 C) Students: 476 (296 E, 180 C)	6	7 middle schools in an urban Southwestern district. 49% H, 73% ELL.	GMRT Comprehension	+0.15	+0.15
Lesaux et al. (2014)	Cluster randomized	20 weeks	Teachers: 50 (25 E with their 37 classes, 25 C with their 39 classes) Students: 2082 (971 E, 1111 C)	6	14 urban middle schools in a large urban school district, California. 71% ELL, mainly Spanish speaking.	GMRT Comprehension Vocabulary	+0.04 -0.04 +0.17*	+0.04

Table 5
Benchmark Assessments

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
4Sight								
Carlson et al. (2011)	Cluster randomized	1 year	Districts: 59 Schools: 514	3-8	Districts in 7 states.	State tests	+0.03	+0.03
Ohio Personalized Assessment Reporting System (PARS)								
May & Robinson (2007)	Cluster randomized	1 year	Schools: 100 (50 E, 50 C) Students: 51,580	10	Schools across Ohio.	Ohio Graduation Test	-0.03	-0.03
ANet								
West et al. (2016)	Cluster randomized	2 years	Schools: 89 (45 E, 44 C) Students: 8070	6-8	Schools in MA, LA, Chicago. 13% W, 87% minority, 85% FRL, 18% ELL, 17% SPED.	State tests	-0.03	-0.03
Acuity								
Konstantopoulos et al. (2016)	Cluster randomized	1 year	Schools: 55 (28 E, 27 C)	6-8	Schools across Indiana.	ISTEP+	+0.03	+0.03

Table 6
Programs Meeting ESSA Evidence Standards for Strong and Moderate Ratings

	Number of Studies	Average Effect Sizes	ESSA Rating
Tutoring			
Perry Beeches	1	+0.36	Strong
REACH Tutoring	1	+0.42	Strong
Butterfly Phonics	1	+0.30	Strong
Cooperative Learning			
The Reading Edge	3	+0.29	Strong
Social-Emotional Programs			
BARR	2	+0.09	Strong
Programs Incorporating Technology			
ITSS	1	+0.18	Strong
Achieve3000	1	+0.29	Strong
READ180	5	+0.08	Strong
Accelerated Reader	1	+0.24	Strong
CCT	2	+0.13	Strong
Passport Reading Journeys	3	+0.07	Strong
Metacognitive Strategy Approaches			
RISE	1	+0.16	Strong
ERWC	1	+0.13	Moderate
SIM	7	+0.09	Strong
Reading Apprenticeship	4	+0.10	Strong
STARI	1	+0.15	Strong
Pathway	2	+0.08	Strong