Effective Reading Programs for

Secondary Students

Ariane Baye
University of Liege (Belgium)

Cynthia Lake
Amanda Inns
Robert E. Slavin
Johns Hopkins University

December, 2016

Abstract

In recent years, major 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 targeted programs for struggling readers. This review of the experimental research on secondary reading programs focuses on 65 studies that used random assignment (n=56) or high-quality quasi-experiments (n=9) to evaluate outcomes of 49 programs on widely accepted measures of reading. Programs using one-to-one and small-group tutoring (ES=+0.23) and cooperative learning programs (ES=+0.16), showed positive outcomes, on average. Among technology programs, metacognitive approaches, mixed-model programs, and programs for English learners there were individual examples of promising approaches. Except for tutoring, targeted extra-time programs were no more effective than programs provided to entire classes and schools without adding instructional time. Outcomes for middle school students were non-significantly higher than those for high school students. The findings suggest that secondary readers benefit more from engaging and personalized instruction than from remedial services.

Effective Reading Programs for Secondary Students

The reading performance of students in America's middle and high schools presents one of the most frustrating problems in all of 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, but despite a substantial focus on reading at all levels and massive investments, secondary reading is advancing very slowly. This is supported by two recent international reading surveys. According to the Program for International Student Assessment (PISA), the mean performance of American 15-year-old students has not changed between 2000 and 2009, and the proportion of struggling readers also did not change (OECD, 2010). Another survey from the Organisation for Economic Cooperation and Development (OECD) of adult competencies shows that the average level in reading of American young adults (16-24 year olds) is below the international average for developed countries (OECD, 2013).

Even more distressingly, 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. Twelfth grade scores present an even more depressing picture. Overall, 37% of twelfth graders scored at or above proficient, a decrease from 40% in 1992. Among White students, 46% of 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. In fact, 28% of twelfth graders scored below basic on NAEP, up from 20% in 1992. The percent reading below basic was 48% for African Americans, 37% for Hispanics, and 35% for American Indians/Alaska Natives.

Students who read below basic 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). In other words, reading deficits in secondary school are a key dynamic in the inequalities and dysfunctions that undermine the economic and social health of our nation.

Ideally, reading success would be assured for all in elementary school. By middle school, reading is often a remedial course. Students who are poor readers in secondary school are likely to see reading instruction as demeaning, babyish, and demotivating (Biancarosa & Snow, 2006). Yet middle and high schools must nevertheless find ways to teach below-level readers, because even if improved practices in elementary schools could reduce the numbers of below-level readers, the numbers are so large that it would be a very long time, if ever, before even dramatically improved reading programs in elementary schools would entirely solve the secondary reading problem. At a minimum, secondary schools will always need strategies to continue to build the reading skills of their students, even those who have adequate skills entering sixth grade.

Secondary school is the last chance for millions of students to improve their reading skills, thereby increasing their chances of graduation, college attendance, and employment (Biancarosa & Snow, 2006; Joftus & Maddox-Dolan, 2003). Poor readers who do graduate are

likely to experience serious difficulties entering college. If they do go to college, they may be required to take non-credit, remedial English courses (Au, 2000; ACT, 2006; American Diploma Project, 2004).

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

The importance of reading in middle and high schools has long been recognized, and several reviews have examined the evidence base for various programs designed to improve secondary reading. In particular, Deshler et al. (2007) summarized evidence on 48 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, though in four of these treatments were delivered by researchers rather than teachers. In an earlier review, 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. 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.

Programs emphasizing curricular changes, or use of technology without a strong link to teaching, were associated with modest impacts on reading achievement.

The Deshler et al. (2007), Edmonds et al. (2009), Herrera et al. (2016), Slavin et al. (2008), and Wanzek et al. (2013) reviews have been overshadowed by dramatic changes in the strength of evidence for secondary reading approaches. One factor in this was 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, to support implementation and evaluation of secondary reading approaches. Striving Readers was well under way when the Deshler et al. (2007), Edmonds et al. (2009), and Slavin et al. (2008) reviews were written. The studies had been completed before the Wanzek et al. (2013) and Herrera et al. (2016) reports were issued, but the full reports on the Striving Readers studies were not yet available. The Striving Readers grants generally required random assignment of students within schools for targeted interventions focusing on struggling readers, usually students reading two or more years below grade level. Striving Readers also typically funded studies comparing whole schools using schoolwide approaches, often based on the targeted approaches, to control schools. These whole-school comparisons usually had too few schools to allow for analysis at the school level (using, for example, hierarchical linear modeling; Raudenbush & Bryk, 2002). Still, all of the Striving Readers studies, often combining data from several cohorts, added a great deal to the evidence base for secondary reading programs. Most state recipients of Striving Readers funds contracted with commercial companies or well-known researchers who already had secondary reading programs to provide their programs, but some states made up their own approaches.

Beyond Striving Readers, other recent developments have added significantly to the high-quality evidence base for secondary reading. In the U.S., the Institute of Education Sciences (IES) has been funding randomized evaluations for some time, and these have included several 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 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 tougher standards than would have been possible in the earlier reviews, and assembling data from a much larger pool of programs and studies.

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, but most importantly seeks to place 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 essentially the same inclusion standards and methods (see Cheung

& Slavin, 2016).

Review Methods

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. Studies of

   variables, such as ability grouping or single-sex classrooms, were not reviewed.

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, but students in self-

   contained special education settings were not.

3. Studies had to have compared students in a given reading program to those taught in an alternative or "business-as-usual" control group. However, comparisons of two experimental programs, lacking a control group representing ordinary practice, were not included. Studies lacking true control groups, such as pre-post studies or those comparing treatment groups to "expected gains," were excluded.

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. The researchers' roles could focus on professional development and coaching, but not direct service to students.

8. Studies' dependent measures had to be quantitative measures of reading performance. The ultimate goal for secondary readers is comprehension. Struggling readers may suffer from poor decoding, vocabulary, or fluency, but an effective intervention at this level should be primarily one yielding progress in reading comprehension. 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. Measures related to reading, such as writing, were not included.

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

10. Studies had to have a minimum duration of 12 weeks, to diminish the potential impact of Hawthorne effects and 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. However, all means were weighted by sample size (using an inverse variance weighting procedure), so small studies counted less than large ones in any means.

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 experimenter-made measures. These all had the effect of toughening standards of inclusion. Stringent standards of research design, large sample sizes, and exclusion of researcher-made measures all significantly diminish inflation of effect sizes (Cheung & Slavin, 2016; De Boer et al., 2014; Lipsey et al., 2012) so

these procedures are likely to reduce effect sizes reported in any research review. However, studies meeting stringent requirements are more likely to replicate in research and practice.

An online appendix (URL) lists studies of secondary reading programs that failed to meet the present inclusion criteria, with at least one of the key reasons for exclusion. 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 1990-2016. Results were then narrowed by subject area (for example, "reading intervention," "educational software," "academic achievement," "instructional strategies"). In addition to looking for studies by key terms and subject area, we conducted searches by program name. Web-based repositories and education publishers' websites were also examined. We searched for studies reviewed by the What Works Clearinghouse (2016). We contacted producers and developers of reading programs to check whether they knew of studies that we had missed. Citations from other reviews of secondary reading programs (e.g., Deshler et al., 2007; Edmonds et al., 2009; Herrera, Truckenmiller, & Foorman, 2016; Slavin, Cheung, Groff, & Lake, 2008; Slavin, Lake, Davis, & Madden, 2011; Wanzek et al., 2013) 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 2016: *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.

Statistical Significance

Statistical significance is reported for all studies. 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. This review reports such individual-level statistical tests along with effect sizes.

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) Procedures and Standards Handbook Version 3.0. When multiple measures were used within studies, or when multiple studies evaluated a given treatment, mean effect sizes were tested for statistical significance using procedures derived from Lipsey & Wilson (2001).

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. However, the review does not report on experimenter-made measures, although results on such measures may also be of importance to researchers or educators. In the large majority of interventions that supplement regular reading courses, the control students do not receive any

supplemental reading instruction, and instead receive "business-as-usual" programs (e.g., an elective course or study hall). For these interventions, the effect of the intervention cannot be disentangled from the effect of more opportunities to learn, since the impact of the supplemental time cannot be distinguished from that of the program itself.

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, while cluster randomized experiments were those in which schools or classes were the unit of random assignment and treatment. As noted earlier, clustered experiments analyzed at the student level were corrected to account for clustering. A few studies claimed to use random assignment because students were routinely assigned to classes by a scheduling computer, but these were categorized as quasi-experimental, not random. *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 schools) or within-school (matched students within schools or classes). Studies using fully randomized designs are less subject to bias than quasi-experimental studies. In the text and in tables, randomized studies are generally listed before quasi-experiments. Within these categories, studies with larger sample sizes are generally listed first. Exceptions were made to put together studies of a given program or similar programs. Therefore, studies discussed earlier in each section should be given greater weight than those listed later, all other things being equal.

Program Categories

All studies that met the inclusion criteria were accepted, regardless of the experimental program they implemented. Then the studies were sorted by program, using information in the study reports and, if available, on the programs' web sites. All study authors had to agree on categorization; if there were disagreements, discussion continued until consensus was achieved. The categories 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 or paraprofessionals. Tutoring sessions were typically given either on some proportion of days (as few as once a week) or daily for a few months, but not all year. 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 seems 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. In this review, all tutoring studies took place in England, under funding from the Education Endowment Foundation. Otherwise, the tutoring models varied widely.

2. **Cooperative learning programs**. Cooperative learning programs involve students working daily in small mixed-ability groups. Usually, cooperative learning groups have 4-5 members, and the students are encouraged to help each other to learn academic content. In secondary reading, cooperative learning programs focus primarily on students

helping each other to learn and apply metacognitive comprehension strategies, such as clarification, summarization, graphic organizers, and prediction. Note that cooperative learning programs that met the standards of this review were always used in whole, non-remedial classes rather than in targeted, remedial reading classes.

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 feedback from peers and teachers (Slavin, 2015; Roseth, Johnson, & Johnson, 2008).

3. **Programs incorporating technology**. A number of widely used secondary reading programs incorporate technology. These include *computer-assisted tutoring programs*, in which students go to a computer lab or use individualized software in class to access reading content. Examples include SuccessMaker, Achieve3000, and the Thinking Reader. Another category is what we called *mixed-methods programs,* in which students rotate through teacher-led, non-technology activities alternating with technology activities. Examples include READ 180, Voyager Passport, and Prentice-Hall Literature.

The theory of action behind programs incorporating technology 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.

4. **Metacognitive strategy approaches**. Many of the instructional process approaches emphasized teaching metacognitive strategies. These approaches are similar in curricular focus to the cooperative learning programs, except that they lack formal learning teams or

groups. Instead, teachers teach students to apply metacognitive strategies to improve their comprehension, either in targeted remedial classes or whole groups.

The theory of action behind metacognitive strategy approaches emphasizes the idea that students who are struggling in reading need to learn specific, well-defined strategies and become effective comprehenders. For example, they need to learn to get the gist, or meaning, of the texts they are reading, to learn and apply useful strategies to use when they run into unknown words, to learn story structure using predictions and identifying story grammar, or to learn to comprehend factual texts using outlining or graphic organizers. Metacognitive strategies share with cooperative learning programs a theory that poor reading in secondary school, beyond any remaining problems with phonics, fluency, or vocabulary, can best be addressed by teaching students methods of making their reading efforts for specific problems they encounter more successful.

5. **Mixed-model professional development**. Many approaches for secondary reading, especially those used in targeted classes for struggling readers, provide extensive professional development and often in-class coaching to help teachers do a better job of teaching. Unlike cooperative learning, metacognitive approaches, or direct instruction, mixed-model approaches are eclectic, emphasizing pedagogy, content, and learning strategies. Teachers are generally not given new student materials, teachers' manuals, or software, but are given quite a lot of professional development. These mixed-model professional development approaches focus on improving the quality of implementation of widely-accepted teaching approaches. The theory of action for mixed-model professional development is that the most direct way to achieve enhanced student

outcomes is to enhance the overall quality of teaching rather than implementing a

particular, systematic classroom approach.

6. **Programs for English language learners.** Interventions specifically designed to meet

the needs of English learners or language minority students have focused on building

language skills for students who are not native speakers of English. In secondary schools

these approaches use English as the medium of instruction, but seek to help English

learners build English vocabulary and facility in reading.

Targeted vs. Whole-Class/Whole-School Approaches

Because of the structure of the Striving Readers grants, versions of the same program

were often evaluated in two forms, *targeted* and *whole class/whole school*. The targeted form

was given to students with the greatest difficulties in reading. Evaluations of targeted versions of

programs invariably randomly assigned struggling readers to receive a relatively intensive form

of the model, to be compared to a business-as-usual control group in which students participated

in electives such as music, art, or study hall.

At the same time, a less intensive form of the same model was often provided to whole

classes or whole schools. In this case, experimental schools were compared to control schools,

usually in a quasi-experimental (matched) design. Some studies compared experimental and

control classes within the same schools, with either random assignment or matching within

schools.

When the whole-school comparisons involved the same schools in which targeted

students were randomly assigned to conditions, struggling students in the targeted group received

the experimental program both in their regular English or reading class classes and in an

intensive reading class, while their control group counterparts received the generalized form but not the intensive one. The whole-school versions of the programs were more diffuse, but did compare experimental schools to pure controls.

In the final year of Striving Readers all evaluations focused only on the targeted designs (Boulay et al., 2015).

Throughout this review, programs are identified as targeted (i.e., small group) or whole-class/whole-school, whether or not the particular evaluations were part of Striving Readers, and the differences in outcomes between the two categories are tested.

Findings

A total of 65 studies evaluating 49 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. 56 (86%) used random assignment, and only 9 (14%) used matched, quasi-experimental designs.

As noted earlier, programs were organized according to their main features and theories of action into six categories: Tutoring, cooperative learning, technology, metacognitive skills, mixed-model professional development and programs for English learners. The findings are summarized in Tables 1 to 6, and described in more detail in the following sections (note that abbreviations used throughout this article for measures, demographic groups, and categories of students are explained in a footnote below Table 1).

===================
TABLE 1 HERE
===================

19

Tutoring Interventions

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), but is rarely used in secondary schools. None of the Striving Readers studies used tutoring, and in fact, all of the qualifying tutoring studies were done in England, as a result of a funding initiative there focused on helping struggling students enter secondary school with adequate reading skills.

One-to-one tutoring

Catch Up® Literacy – Targeted. Catch Up Literacy[1] (Bentley & Reid, 1995) is a structured one-to-one tutoring intervention for reading, administered to struggling readers by paraprofessionals. It was originally created in England to work with students in primary (elementary) schools, but was extended to secondary schools (see www.catchup.org). Students receive 15-minute sessions twice a week over the course of a school year. Each session includes prepared reading, reading out loud, discussing the text, and linked writing.

In schools using Catch Up, a member of the school staff serves as a manager to assist at least one teaching assistant. Paraprofessional teaching assistants learn to select appropriate starting points and books for students, to set targets, and to monitor progress toward those targets.

In a third-party evaluation of Catch Up (Rutt, 2015), struggling readers were identified in Year 6 (the final year of primary school). They were randomly assigned to Catch Up or to business-as-usual teaching. Students who were assigned to tutoring received some tutoring during Year 6, but the main intervention was in Year 7, the first year of secondary school in

---

[1] Catch Up® is a not-for-profit UK registered charity (1072425) and Catch Up® is a registered trademark

England. Fifteen secondary schools located throughout England had roughly equal numbers of experimental and control students (n=286 E, 271 C). Key Stage 2 tests, given to all English students at the end of Year 6, served as pretests. The posttests were New Group Reading Tests (NGRT), a test widely used in England. An intent-to-treat-design was used, meaning that all students randomly assigned to experimental or control groups were included no matter how many tutoring sessions they received. Analyses of covariance, controlling for Key Stage 2 scores, found a marginally significant difference favoring the tutored students (ES=+0.16, p=.08, n=268 E, 271 C).

The Perry Beeches Coaching Program – Targeted. The Perry Beeches Coaching Program provides struggling readers with one hour of one-to-one tutoring every two weeks. The program does not provide pre-defined content or supporting materials. Coaches tailor activities according to students' needs.

The program was evaluated in four secondary schools in Birmingham, England. One-to-one or small group tutoring sessions were offered to Year 7 students (i.e., the first year of secondary education) over one academic year. Each participating student received five one-hour sessions biweekly. Students were randomly placed in the tutoring intervention group instead of another class (usually not English), while control students continued with their regular school schedule.

A randomized independent evaluation of the intervention was carried on by Lord, Bradshaw, Stevens, & Styles (2015). The effect of one year of intervention in the Perry Beeches Coaching Program was large and significant (ES=+0.36, p<.001; n=149 E, 142 C) on the GL Assessment Progress in English test given at pre- and posttest.

REACH Tutoring – Targeted. REACH (no connection to a U.S. program of the same name) is a U.K. program that provides struggling readers in early secondary school with one-to-one tutoring in 35-minute sessions, once a week for 20 weeks. The tutors are specially trained paraprofessionals.

REACH was evaluated by Sibieta (2016) in 27 secondary schools in and around Leeds, England. Two forms of the program were evaluated. In ordinary REACH, children read aloud from books at their level while tutors keep a "running record," which they then use as a basis for remedial teaching. They then provide instruction in phonemic awareness and phonics. In a variation, REACH-LC, a language comprehension element was added. This component emphasizes metacognitive skills, reading comprehension, inference, and writing (Clarke, Snowling, Truelove, & Hulme, 2010).

In a study by Clarke, Snowling, Truelove, & Hulme (2010), students in Years 7 and 8 who scored poorly on the Single Word Reading Test were randomly assigned to REACH, REACH-LC, or waiting-list-control treatments for 20 weeks. They were pre- and posttested on the NGRT. At posttest, controlling for pretests, students who received REACH tutoring scored significantly higher than controls (ES=+0.33, $p < .001$, n=70 E, 63 C) and students who received REACH + LC also outperformed controls (ES = +0.51, $p<.001$, N=69 E, 63 C). Averaging across the two variations, the mean effect size was +0.42.

Small-group tutoring

Butterfly Phonics – Targeted. Butterfly Phonics uses formal phonics instruction, understanding the global aspects of a text, and class discussion of text meaning to improve

22

reading comprehension. The program is delivered to groups of 6-8 students by a trained practitioner and an assistant.

Merrell & Kasim (2015) evaluated a four-month intervention of Butterfly Phonics in six London schools. Three hundred ten seventh grade struggling readers were randomly assigned within schools either to withdraw from their regular English courses to receive two hours of Butterfly Phonics lessons per week (n=161 E) or to serve as the control group (n=149 C). Researchers found a significant effect size of +0.30 (p<.001) in favor of the experimental group on the New Group Reading Test (NGRT).

Rapid Phonics combined with Sound Discovery – Targeted. 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. They are typically used with beginning readers, but a study in England adapted them as catch-up programs for older struggling readers at the transition between primary (Year 6) and the first year of secondary school (Year 7) (King & Kasim, 2015). 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 22 participating primary schools, struggling readers in need of decoding improvement were randomly assigned to the treatment or control condition (n=86 E, 92 C), and the students were then followed into Year 7. Students using Rapid Phonics combined with Sound Discovery did

not perform better than control students on the New Group Reading Test (ES=-0.05, *n.s.*), controlling for pretests.

Taken together, the overall sample size-weighted effect size for tutoring programs was +0.23. It was +0.26 for three one-to-one models, and +0.17 for two small group models. However, it is important to note that because of the cost of tutoring, the number of students served was very low compared to numbers in whole-class and whole-school strategies, and this must be taken into account in evaluating tutoring as a practical intervention for secondary reading.

=================

TABLE 2 HERE

=================

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, but are less often used in secondary reading. Table 2 lists the seven studies that met the inclusion standards. These studies evaluated a total of three programs.

The Reading Edge/Student Team Reading – Whole Class/School. The Reading Edge (Slavin, Chamberlain, Daniels, & Madden, 2009), adapted from a program called Student Team Reading (Stevens & Durkin, 1992), is a cooperative learning program for middle schools in which students work in four- or five-member teams to help one another build reading skills. Students engage in partner reading, story retelling, story related writing, word mastery, and story-structure activities to prepare them and their teammates for individual assessments that

form the basis for team scores. Instruction focuses on explicit teaching of metacognitive strategies. The initial professional development is followed by in-class coaching about once a month.

Slavin, Chamberlain, Daniels, & Madden (2009) evaluated the effectiveness of The Reading Edge on two successive cohorts of grade 6 students randomly assigned to the program or to a control group in two high-poverty rural schools in West Virginia and Florida. Teachers were also randomly assigned to the intervention. On the Gates-MacGinitie Reading Comprehension test, students in The Reading Edge scored significantly higher than students enrolled in control classes (n =405 E, 383 C). The effect size was +0.15 (p<.01).

Stevens and Durkin (1992) carried out two large matched studies of the earlier form of The Reading Edge, Student Team Reading, in Baltimore City middle schools. The larger study involved grade 6 to 8 students from two experimental middle schools matched with three control schools. On CAT reading comprehension posttests, students involved in the program (n=1798 E, 2188 C) scored substantially better than control students, with an effect size of +0.38, though this was not significant at the cluster level.

In a smaller study, 20 experimental classes from three schools were matched with 34 classes in three control schools. At the end of the school year, grade 6 students (n=455 E, 768 C) involved in the experimental schools scored significantly better than control students on the Reading Comprehension scale of the California Achievement Test (CAT) (ES=+0.13, not significant at the cluster level).

Across the three studies of The Reading Edge, the weighted effect size was +0.29.

Talent Development High School – Reading (Strategic Reading and Student Team Literature) – Whole Class/School. The Talent Development High School (TDHS) is a whole-school reform model. Within TDHS, Strategic Reading and Student Team Literature (Strategic Reading for short) constitute 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, similar to The Reading Edge, 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, vocabulary, and comprehension questions.

Teachers model the comprehension process through "read-aloud/think-aloud" demonstrations, and 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 three high-poverty non-selective high schools in Baltimore. Three well-matched control schools were selected in advance. Eight teachers used Strategic Reading with 20 classes. Control schools also provided double dose classes in English and math, but followed the district curriculum, focused on test preparation. There were a total of 457 students in the study (257 E, 200 C).

District-administered CTBS scores from the fall of eighth grade were used as pretests, and CTBS Terra Nova scores from spring of ninth grade were used as posttests. Posttests,

controlling for pretests and demographic factors, showed an effect size of +0.32. This is

significant at the student level but not the cluster level.

Collaborative Strategic Reading (CSR) – Whole Class/School

Collaborative Strategic Reading (CSR) is an intervention developed by Vaughn et al.

(2011) that teaches reading comprehension strategies to students working in small cooperative

learning groups.

Denver Public Schools (2016; also see Boardman, Klingner, Buckley, Annamma, &

Lasser, 2015) carried out an evaluation of Collaborative Strategic Reading in science, social

studies, and language arts classes in 16 Denver middle schools. Within schools, social studies

and science teachers of grades 6-8 were usually paired, and the pairs were randomly assigned to

CSR or control conditions for one year. Control students received no CSR, and experienced

ordinary teaching. Sections were randomly assigned to conditions, so this is a cluster randomized

trial with clustering at the teacher level. Outcome variables were GMRT and the TCAP state

standardized reading test, and GMRT pretests were used for both outcomes. Combining across

three cohorts and grades 6-8, there were no significant differences between experimental and

control students. Effect sizes were +0.04 for GMRT and +0.02 for TCAP, for a mean of +0.03

(n.s.).

Vaughn et al. (2011) conducted a large cluster randomized evaluation of the CSR

program with seventh and eighth grade students in six middle schools from Texas and Colorado.

During the first 4-6 weeks of the intervention, teachers modeled 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 were assigned to cooperative learning groups to allow them to master each strategy. The intervention was implemented 50 minutes a day, two days a week, during regular English Language Arts lessons. Teachers were provided three days of professional development and materials such as sample lessons. Students (n=400 E, 382 C) were randomly assigned to classes, and classes (n classes=34 E, 27 C) to experimental or control conditions. In order to control for teacher effects, the 17 teachers involved in the study taught both treatment and control classes. They were required to cover the same curriculum over the same period with all the students. The research team provided in-class support to CSR teachers one or two times a month at the beginning of the intervention. On three reading comprehension measures included at pre- and posttest, one showed a small, marginally significant positive effect (ES=+0.12, p<.10) at the student level on the Gates-MacGinitie Reading Comprehension (GMRT) test. The effects were -0.08 (*n.s.*) on the AIMSweb Reading Curriculum-Based Measure and +0.07 (*n.s.*) on the Test of Silent Reading Efficiency and Comprehension, for an average effect size of +0.04.

A study by Vaughn et al. (2013) followed up the Vaughn et al. (2011) study of CSR. Twelve teachers in the same six middle schools participated. Schools non-randomly assigned students to one of 48 English or reading classes, and these classes were then randomly assigned within teacher to CSR (n=26) or control (n=22) treatments. All 12 teachers had participated in the previous study, and had received 18 hours of CSR training. During the second year, they received another six hours of training, plus three half-hour reinforcement sessions during the school year, and in-class support and coaching once or twice a month.

Students were individually tested at pre- and posttest by trained testers unaware of treatment assignments. On Gates posttests, controlling for pretests, the effect size was +0.10 (n.s.). On the TOSREC, the effect size was +0.11, for a mean across the two measures of +0.10.

The weighted mean effect size across three CSR studies was +0.04. Adding the findings of the CSR studies to those of the three Reading Edge studies and the TDHS study, the weighted mean ES for all cooperative learning studies was +0.16. This is significantly different from zero.

=================

TABLE 3 HERE

=================

Programs Incorporating Technology

Table 3 summarizes research on approaches incorporating technology. The approaches vary widely, especially based on how frequently technology is used and how well integrated technology use and live teaching are. Within Table 3, programs are divided into two categories: *Computer-assisted instruction (CAI)* and *mixed methods.* Computer-assisted instruction provides self-instructional material at students' performance level, frequent feedback, and celebration when students make progress. Students may work on CAI software in a computer lab or in their classroom, but there is little link to teachers' instruction. In contrast, mixed-methods instruction has a more important role for the teacher and cycles students through off-line and on-line activities, with an attempt to link the different types of instruction.

Computer-assisted instruction (CAI).

The Thinking Reader - Whole Class/School. The Thinking Reader is a software program for students in grades 5 to 8 that teaches reading comprehension skills to struggling readers. It

29

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 were asked to incorporate the program in their regular English Language Arts activities for 110-165 minutes a week. Teachers were encouraged to develop introductory and after-reading activities offline, but only about half of them did so. Control classes received hard copies of the novels, and teachers taught them as usual.

The impact of the program was evaluated in a large cluster randomized study involving 48 intervention and 42 control teachers and their sixth graders (n students=1154 E, 986 C) (Drummond et al., 2011). Thirty-two high-poverty schools from three states participated. The effect on GMRT at pre- and posttest was close to zero (ES=+0.01, ns).

SuccessMaker - Whole class/school.

SuccessMaker is an adaptive K-8 computer-based reading program from Pearson Education Inc. It provides individualized reading activities, game-like environments, interactive aids, and a reporting system to inform teachers on student progress. A large cluster randomized quasi-experimental study was carried out for Pearson by Gatti Evaluation Inc. (2011) on the effectiveness of the program for grades 3, 5 and 7 students. The data relating to grade 7 are relevant here. Twenty-two grade 7 classrooms from five urban and suburban schools from different states were assigned at random to the program or to the control condition. Most of the schools involved were at or well above their state reading standards. The intervention was usually implemented in two or three sessions a week within the regular English Language Arts scheduled block. On the GRADE given at posttest, the intervention students (n =254 E, 199 C)

performed significantly better than the comparison students, controlling for pretests. The article incorrectly used the standard deviation of gain scores in computing effect size. The corrected effect size, using the unadjusted posttest standard deviation, was +0.11 (n.s).

Achieve3000 - Whole class/school. Achieve3000 is an online literacy program that provides reading content for students grades 2 through 12. The program is focused on non-fiction reading and writing. Metacognitive skills and reading strategies (such as summarizing, generating questions, setting the purpose) are used to improve the comprehension of informational texts. The program offers diagnostic and assessment data to teachers and school administrators.

Magnolia Consulting (Shannon & Grant 2015) was asked to perform an independent evaluation of Achieve3000. Twelve schools from four urban and suburban districts were identified across the U.S. to evaluate the effectiveness of the program for grades 6 and 9 students. Within schools, teachers were randomly assigned to intervention or control conditions. Experimental teachers were asked to use Achieve3000 at least 90 minutes each week during English Language Art lessons. Control teachers were asked to continue with their usual literacy program. On GMRT tests administrated by the teachers at the beginning and at the end of the school year, Grade 6 students (263 E, 231 C) scored non-significantly better than control students (ES=+0.22). The effect size for Grade 9 students (122 E, 126 C) was also large but not significant at the cluster level (ES=+0.44). The weighted average effect size across grades 6 and 9 was +0.29.

Mixed-model approaches

eMINTS - Whole class/school. eMINTS (Enhancing Missouri's Instructional Networked Teaching Strategies) is a comprehensive schoolwide program for rural middle schools that provides extensive professional development to teachers to help them with technology integration, inquiry-based learning, high-quality lesson design, and communities of learners. eMINTS is not limited to reading, but has an equal focus on math. eMINTS was an Investing in Innovation project, and in its i3 evaluation it provided schools more than 240 hours of professional development over two years, monthly classroom-level coaching visits, and in a third year, provided some eMINTS schools with access to the Intel Teach program, web-based tools designed to facilitate inquiry learning.

In an evaluation by Meyers, Molefe, Brandt, Zhu, & Dhillon (2016), 60 high-poverty rural middle schools across Missouri were randomly assigned to receive eMINTS (n=40) or to continue with ordinary practices (n=20) over a two-year period. In a third year, the eMINTS schools were randomly divided into a group that also received the Intel Teach program and a group that continued with ordinary eMINTS. Controlling for scores on the Missouri state test (MAP) Communication Arts scale from spring, 2011, before treatments began, Communication Arts scores after 3 years were nearly identical in all three conditions. The effect size comparing eMINTS to control was -0.04 (n.s.), and for eMINTS + Intel vs. control, it was -0.08 (n.s.), for a mean effect size of -0.06. It is important to note that there were modest positive effects on the state mathematics test.

READ 180 – Targeted. READ 180 (Scholastic, 2006) 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 rBook – workshops on content area and literature introduced by video – for teaching reading strategies, vocabulary, writing, and grammar. READ 180 audiobooks and leveled paperbacks support modeled and independent reading.

In the Memphis Striving Readers program, READ 180 was implemented as a two-year supplemental reading intervention for 6th through 8th graders. All students received a whole-school intervention consisting mainly of professional development to enhance literacy across the content areas. In addition, in five of the eight schools, students who demonstrated the strongest need for intervention were randomly assigned to receive READ 180 or to just receive the school program. In five of the eight participating middle schools, READ 180 was given to struggling readers in addition to regular English Language Arts classes. In the three others, intervention students were enrolled in a two-hour English Language Arts class integrating 90 minutes of READ 180, while control students in these schools were enrolled in regular 45-55 minute English Language Arts classes. Schenck, Feighan, Coffey, & Rui (2011) evaluated this large randomized experiment (n=545 E, 728 C). After two years of intervention, there were no differences between experimental and control groups (ES=+0.02, n.s.). Results were similar after the first year of intervention.

The Wisconsin Striving Readers project randomly assigned struggling readers in grades 6 through 9 to READ 180 or to control conditions for a one-year supplemental reading intervention. Students scoring at least two years below their grade levels in five Title I Milwaukee Public Schools were the participants (n=335 E, 284 C). Swanlund, Dahlke, Tucker, Kleidon, Kregor, Davidson-Gibbs, & Hallberg (2012) found that students who received READ

33

180 outperformed control students on Reading Measure of Academic Progress outcome measures (ES=+0.14, p<.05).

READ 180 was one of three programs evaluated in comparison to a control group used in seven schools participating in the large randomized trial led by Lang, Torgesen, Vogel, Chanter, Lefsky, & Petscher (2009) in Florida. This research project was a multi-year and multi-tier intervention. All 9th grade students in each school received science and social science courses using content-enhancement routines and reading strategies. READ 180 supplemental reading instruction was provided to struggling readers, placed in "high risk" or "moderate risk" groups according to their reading levels. High risk students were reading below the fourth grade level, while moderate risk students were reading between the fourth and sixth grade levels. Struggling readers were randomly assigned to READ 180 or control conditions within schools and within levels. READ 180 was used for one year before the experiment, in order to familiarize teachers with the intervention before examining its effects. The control group received a daily supplemental 90-minute block of reading instruction designed by the schools themselves using commercial materials and a software with state test preparatory activities. READ 180 significantly improved the state reading test scores of the "moderate risk" students (ES=+0.30, p=.03; n=207 E, 202 C). The scores of highest-risk students who received READ 180 were significantly lower than those of their controls (ES= -0.27, p<.05; n=100 E, 90 C). The weighted average effect size was +0.12.

In the Springfield-Chicopee (Massachusetts) Striving Readers program (Sprague et al., 2012), two targeted interventions for struggling readers were compared to controls: READ 180 and Xtreme Reading. Both programs were used as a one-year supplemental reading intervention for struggling 9th graders in the context of a broader whole-school intervention aimed at

improving literacy instruction across all disciplines. Teachers of all students received professional development on reading strategies. Struggling readers were randomly assigned to READ 180, Xtreme Reading, or to a control group that provided supplemental services ordinarily available to students in need of additional reading support. This section only reports on the comparison of READ 180 to control groups. The majority of control students took regular ELA, while others were enrolled in elective courses in lieu of receiving additional reading supports. The randomized experiment was implemented in five high schools serving mostly low-income students. Combining the results of the five cohorts of 9th grade students involved in the Striving Readers program, Sprague, Zaller, Kite, & Hussar (2012) found a significant positive effect in favor of the 231 READ 180 students, compared to their 225 control counterparts (ES=+0.18, p=.03) on SRDT 4 posttests.

In the Newark Striving Readers project, Scholastic's READ 180 was chosen to replace the core language arts curriculum for low-achieving readers in Title I schools randomly assigned to the treatment or control condition (n schools=10 E, 9 C). The one-year intervention program was developed for struggling readers in grades 4 to 12. The school district purchased add-on READ 180 material and books from McDougal Littell, and provided a pacing guide with 3-week plans of instruction in order to expose the students to up to three years of intervention, from grades 6 to 8.

Four cohorts of students experienced the program, with incoming struggling sixth graders being added each year to the sample. An independent evaluation was performed by Westat (Meisch et al., 2011). Students could have received one, two, or three years of treatment, depending on their grade when the project began and ended. On SAT 10 reading comprehension subtests administered by the independent evaluator, students in the intervention schools

performed better than students in the control schools after two years of intervention (ES=+0.14, p=.02; n students=814 E, 706 C). Results for one and three years of intervention were small and not significant (1 year: ES=+0.04, n.s.; n students=1350 E, 1205 C; 3 years: ES=+0.06, n.s.; n students=552 E, 471 C). Overall, the average effect size was +0.06.

Across all five qualifying studies of Read180 the weighted effect size was +0.08.

System 44 – Targeted. Scholastic's System 44 (http://www.hmhco.com/products/system-44/) is a reading program developed for older struggling readers who have not mastered basic phonics and decoding skills. The program focuses on decoding, fluency, and comprehension. During a typical System 44 lesson provided 60 minutes daily, 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 material such as paperback and audio books.

RMC Research Corporation was asked by Scholastic to evaluate the impacts of System 44 on students in grades 4 to 8. The first study (Beam, Faddis, & Hahn., 2011) involved four middle and 3 high schools from a large suburban district in California. Struggling readers in need of additional phonetic intervention were identified, and 75 grades 6 to 8 randomly selected students received the intervention during their language arts instruction block. Results showed that control students performed better than intervention students. The effect on the Test of Silent Reading Efficiency and Comprehension posttests, controlling for pretests, was significantly negative (ES=-0.24, p<.05). There were no differences on the California Standards Test (ES= -0.04), for a mean of -0.14.

The second study (Beam & Faddis, 2012) included students in grades 6 to 8 in 6 schools in an urban district in Michigan. About half of the students had been identified as having specific learning disabilities. Struggling readers were randomly assigned to use System 44 or to continue their ordinary programs. Schools differed in the way they used the program: one of the schools used System 44 as a replacement reading class and the other pulled students out of a study skills class. This small study (n students=70 E, 75 C) found positive effects on reading comprehension. On the TOSREC, the effect size was +0.20 (p<.05). Across the two studies, the weighted mean effect size of System 44 was +0.03.

Accelerated Reader – Targeted. 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.) As used in this evaluation, Accelerated Reader provided 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, compared to baseline.

The Gorard et al. (2015) evaluation involved students who were very low achievers, as determined by Key Stage 2 (KS2) tests given at the end of primary school. 349 students were randomly assigned to receive the program (166 E, 183 C) within four schools across England. On on-line New Group Reading Tests (NGRT), controlling for KS2 pretests, students who received Accelerated Reader scored significantly higher (ES= +0.24, p<.05).

Prentice Hall Literature – Whole class/school. Prentice Hall Literature (2010) is a program for sixth to twelfth grade students. It combines off-line textbooks with online

components. The textbooks include units organized by grade level focused on a specific genre, combining classic and contemporary literature. Paired reading selections allow teachers to differentiate instruction according to the level of reading ability of the students. Questions are provided at the beginning of each unit, and connected activities are integrated throughout. The off-line component is supplemented by online material including vocabulary games, audios, and videos. A large cluster-randomized study involving eight schools from four states with high ethnic minority populations was performed by Eddy, Ruitman, Hankel, & Sloper (2010). Sixteen teachers and their grade 7, 8, and 10 classrooms were randomly assigned to Prentice Hall Literature while thirteen teachers kept their existing language arts programs. The impact on experimental students (n students=744 E, 774 C) on the Gates-MacGinitie was non-significantly negative at the cluster level (ES=-0.10, *n.s.*).

Comprehensive Circuit Training (CCT) – Whole class/school. Comprehensive 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 after-reading, which includes comprehension quizzes. Only the final component involves teacher instruction, although teachers help keep students on track in all components.

Fogarty et al. (2014) carried out a cluster randomized evaluation of CCT in three middle school English language arts classes in the Southwest. A total of 61 (30 E, 31 C) classes were randomly assigned within 14 teachers (n students= 411 E, 448 C). HLM analyses found that on GMRT-4 scores, controlling for GMRT-4 pretests, students in CCT did not differ significantly

from controls (ES=+0.12, n.s.). However, there was evidence that classes with higher implementation ratings had students who scored better than those in low-implementing classes.

In a later study of CCT, Fogarty et al. (2016) randomly assigned 228 students (112 E, 116 C) to 16 classes (9 E, 7 C) within three teachers in three middle schools in Texas. While a significant positive effect (ES=+0.14, p<.03) was found on a factor score derived from scores on GMRT, GRADE, and GORT, no significant differences were found on any of the outcomes taken one at a time. Significant positive effects were found on the TOSREC (ES=+0.24, p<.04). Nonsignificant effect sizes were reported for STAAR Reading (ES=+0.09), GMRT (ES=+0.12), GRADE (ES=+0.11), and GORT (ES=+0.18). The weighted mean effect size across all five qualifying measures was +0.15 (n.s.). Across the two studies, the mean effect size was +0.13.

Voyager Passport Reading Journeys – Targeted. *Voyager Passport Reading Journeys* (www.voyagersopris.com) is a supplemental literacy curriculum designed for adolescents who struggle with reading. This highly-structured intervention program relies on fifteen two-week sequences of lessons mixing whole-class and small group lessons as well as individualized computer-based practice. Formative assessments are conducted at the end of each sequence. The curriculum focuses on reading comprehension strategies, vocabulary, word study, and writing, using mainly science and social studies topics.

Voyager Passport Reading Journeys was used in three of the Striving Readers programs (Boulay et al., 2015)**.** In each of these, struggling readers received 50 minutes of supplemental instruction daily over the course of a year. Control students received electives or other services (study hall, tutoring) not focused on literacy. Evaluations randomly assigned students within schools to Voyager or control treatments.

The Louisiana Striving Readers project (Vaden-Kiernan, Caverly, Bell, Sullivan, Fong, Atwood, Borman, Park, & Jones, 2012) included sixth and seventh graders (n=548 E, 554 C) scoring below basic on the state ELA test in ten Title I middle schools in four districts. Students were randomly assigned to Voyager or control conditions within schools. Class sizes for the Voyager intervention classes averaged 13:1. Intervention students made significant progress on GRADE compared to controls. The effect size was +0.27 (p<.001) at the student level. However, on the state English Language Arts test, the iLEAP, there were no differences between treatment and control groups (ES=+0.06, n.s.). The average effect size was +0.17, which is significant at the student level.

In the Virginia Striving Readers project, Schenck, Jurich, Frye, Lammert, Sayko, Najerat, & Willard (2012) evaluated the efficacy of Voyager Passport Reading Journeys for seventh and eighth grade readers at least two years below grade level in nine Title I urban middle schools. Students were randomly assigned to Voyager or control conditions within schools. Intervention students did not show more progress than their counterparts on Gates-McGinitie Reading tests (ES=+0.06; n=*279 E, 289 C*) nor on Virginia Standards of Learning English/Reading tests (ES=+0.06; n=*343 E, 358 C),* for an average effect size of +0.06.

In the Illinois Striving Readers project evaluated by Dimitrov, Jurich, Frye, Lammert, Sayko, & Taylor (2012), eligible ninth grade struggling readers were assigned in pairs matched on relevant characteristics and then randomly assigned within six high schools to Voyager or control conditions (n=427 E, 428 C). Effect sizes averaged +0.02 (n.s.) on Gates-McGinitie and -0.09 (n.s.) on EXPLORE, for a mean of -0.03 (n.s.). Results should be taken with caution due to the high level of attrition (about 40%), although the final sample remained adequately matched.

Across three studies, the weighted mean effect size of Voyager Passport was +0.06.

iLit – Whole class/school. iLit (inspire Literacy) is a digital instruction approach designed for struggling readers in grades 4-10. Students choose among more than 500 eBooks and work on vocabulary and comprehension strategies. After a teacher overview, students work independently, keeping on-line journals, answering questions, and discussing books in groups.

Gatti (2016) carried out a two-year longitudinal evaluation of iLit in six middle schools located in six different states. Within each school, students (114 E, 99 C) were randomly assigned to experimental or control conditions for a two-year period. Students were pre- and posttested each year on the GRADE. The average effect size across the two years was +0.09 (n.s.) for GRADE Total, +0.12 (n.s.) for Total Comprehension, and +0.01 (n.s.) for Vocabulary.

Texas Technology Immersion Pilot - Whole class/school. The Texas Technology Immersion Pilot (eTxTip) was an evaluation of a technology immersion intervention designed to cover language arts, math, science, and social studies in grades 6-8. The project was implemented over a three-year period. 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.

In a study by the Texas Center for Educational Research (Shapley, Sheehan, Maloney, & Caranika-Walker, 2009), 21 technology immersion schools were matched with 21 control schools based on prior test scores and demographic factors. Most students qualified for free- or reduced-price lunches (67%) and most were Hispanic (58%). Controlling for prior scores on the Texas Assessment of Knowledge and Skills (TAKS), TAKS posttest effect sizes averaged +0.08

41

for Cohort 1 (grades 6-9), +0.07 for Cohort 2 (grades 6-8), and +0.02 for Cohort 3 (grades 6-7), for a mean of +0.06. None of these differences were statistically significant.

The weighted mean effect size for three studies of CAI models was +0.09. For 17 studies of mixed-method models it was +0.08, and across all 20 studies of technology applications the weighted mean effect size was +0.08.

========================

TABLE 4 HERE

========================

Metacognitive Strategy Approaches

Table 4 lists 12 studies of 10 programs using metacognitive strategies to enhance students' reading, either teaching remedial, targeted groups of struggling readers, or whole classes, usually also primarily serving low achievers. In these programs, students are taught specific strategies to help them comprehend narrative and factual texts of various kinds, to study, and to write.

Strategic Instruction Model. The most extensively researched of all secondary reading models is the Strategic Instruction Model (SIM; Deshler & Schumacker, 2005) developed at the Center for Research on Learning at the University of Kansas. SIM 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, while others (especially Xtreme Reading and Fusion) are intensive supplemental programs for students who are two to five years below reading level.

In the Striving Readers grants to state departments and large districts (Boulay et al., 2015), several grantees adapted their approaches from SIM or Xtreme Reading and gave their versions new names. The complete set of SIM-related programs with at least one study meeting the standards of this review was as follows:

All Students	Targeted
Content Literacy Curriculum (CLC)	Xtreme Reading
Adolescent Literacy Model (ALM)	Fusion
	Learning Strategies Curriculum (LSC)

SIM: Xtreme Reading – Targeted. Xtreme Reading is a model derived from the University of Kansas Strategic Instruction Model (SIM; Deshler & Schumaker, 2005). The program teaches struggling adolescent readers step-by-step strategies for word identification, vocabulary, self-questioning, visual imagery, paraphrasing, and inference, among other objectives. Highly structured instructional material is provided to the teachers. A large randomized evaluation of Xtreme Reading was carried out by Somers, Corrin, Sepanik, Salinger, Levin, Zmach, & Wong (2010). The study compared Xtreme Reading and Reading Apprenticeship Academic Literacy (RAAL) to a control condition. Thirty-four high-poverty high schools in 10 districts were randomly assigned to receive either Xtreme Reading or RAAL. Then ninth graders who were reading 2 to 5 years below grade level were randomly assigned within the participating schools to a daily Xtreme Reading or RAAL reading class or to continue in an

elective class, such as music or art. Students were pre- and posttested on the GRADE. There were two cohorts treated in the same way. Combining across them, students in 17 schools assigned to use Xtreme Reading did not differ from controls on GRADE (ES=+0.04, n.s.) or on English Language Arts state tests (ES=+0.08, n.s.). There were positive effects for the RAAL treatment, described later in this article.

Three studies used Striving Readers funding to implement Xtreme Reading. One of the evaluations, in the Portland (OR) Public Schools (Faddis, Bean, Maxim, Vale Gandhi, Hahn, & Hale, 2011), reported an experiment initially involving 4 middle and 4 high schools. In the third project year, two more middle schools were added. In each school, grades 7 through 10 students reading at least two years below grade level were randomly assigned to participate in Xtreme Reading every day over a year, or to participate in a business-as-usual control group. All students in the schools, experimental as well as control, also received a content-area reading intervention. In the six middle schools, Xtreme Reading was used as a replacement model during the English Language Arts and Social Studies blocks. In the four high schools, Xtreme Reading was given as a supplemental intervention, while the control group participated in electives. A sample was built up over four years. All students were pre- and posttested on the GRADE (middle school: n= 401 E, 421 C and high school: n = 355 E, 402 C), and state standardized test (OAKS) scores were obtained as pre- and posttests for all in the sample excluding 9th graders who do not take the OAKS (middle school: n=472 E, 482 C, high school: n=260 E, 254 C). On the GRADE, the average effect size across the four years for the supplemental intervention within high schools was +0.12 (p=.04) and +0.02 (n.s.) on the OAKS. Combining the two measures produces an average effect size of +0.07. Within middle schools, where Xtreme Reading was used as a replacement model, the average effect size on the GRADE was +0.29 (p< .01) and on the OAKS

was +0.12 (p=.01). The average effect size for middle school students was +0.21. Across middle and high school students the average effect size on all measures was +0.15 (p< .01).

In the Springfield-Chicopee (Massachusetts) Striving Readers program, teachers and students were randomly assigned to Xtreme Reading, READ 180, or control conditions. The READ 180 vs. control results were presented earlier in this article. The interventions served struggling 9th graders during one year. Control students were enrolled in elective courses in lieu of receiving additional reading supports. The randomized experiment was implemented in five high schools with a high proportion of disadvantaged students. Sprague, Zaller, Kite, & Hussar (2012) averaged the results of the five cohorts of 9th grade students served with Xtreme Reading in comparison to controls. The intervention students did not make more progress than their control counterparts on SDRT 4 given at posttest, controlling for state tests given a year earlier (ES=0.00, n.s.; n students =223 E, 225 C).

Across the three studies of Xtreme Reading, the weighted mean effect size was +0.09 (p<.01).

SIM: Content Literacy Curriculum (CLC) – Whole class/school. The Content Literacy Curriculum (CLC) is another of the SIM variations. In a large cluster-randomized study by Corrin, Lindsay, Somers, Myers, Meyers, Condon, Smith (2012), 28 high schools (n=15 E, 13 C) from Midwest states were randomly assigned to receive CLC or to maintain their "business-as-usual" activities. Struggling readers received Fusion as a target intervention. The content area portion of the intervention (CERT) consisted of professional development and on-site support for four teachers of English Language Arts, mathematics, science, and social studies. The training was devoted to grade 9 teachers during the first year, and to grade 9 and 10 teachers the second

year. The GRADE test was used as an outcome measure. After one year, grade 9 students did not make significant progress compared to comparison school students (average ES=+0.11, *n.s.*; n students=5844 E, 3953 C, 2 cohorts). Results were similar after two years of treatment (ES=+0.10, *n.s.*; n students=2908 E, 1638 C).

SIM: Learning Strategy Curriculum (LSC) – Targeted. In the Kentucky Striving Readers study (Cantrell, 2016), struggling readers two or more years below reading level within 21 rural middle and high schools were randomly assigned within schools to receive a targeted version of SIM every day for a year, or to participate in the electives they would have ordinarily received (e.g., band or civics).

The evaluation focused on four successive groups of struggling readers in grades six and nine, each of which participated for one year. Students were randomly assigned within schools to receive LSC. All students were pre- and posttested on the GRADE. The effect size was +0.08 for sixth graders (n.s.), and +0.12 for ninth graders (p=.03), for an average effect size of +0.10.

SIM: Fusion Reading Program – Targeted. Fusion is a highly structured course based on explicit teaching and ongoing assessment for grade 6-12 struggling readers. Fusion Reading contains advanced phonics instruction, decoding and fluency, and strategies to improve comprehension.

Fusion Reading was implemented in grades 6 through 10 in four middle and three high schools from three districts in suburban areas of Michigan under Striving Readers 2009 funding. A large sample of students scoring two years below grade level were randomly assigned to Fusion Reading for a one-year daily supplemental class period, or were assigned to a control

condition engaged in "business-as-usual" non-literacy activities. An independent evaluation was performed by Schiller, Wei, Thayer, Blackorby, Javitz, & Williamson (2012). The average effect size on the three GRADE sub-tests was +0.05 (n.s.; n=285 E, 296 C). The effect was positive but not significant on the Michigan state test, MEAP (ES=+0.11, n.s.), but results should be read with caution since the sample sizes were half as large as those of the GRADE tests (n=118 E, 138 C). The weighted mean effect size was +0.07.

SIM: Adolescent Literacy Model (ALM) – Whole class/school. In a Striving Readers study in seven rural districts in Kentucky, Cantrell, Almasi, Carter, & Rintamaa (2011) evaluated two programs. One was a schoolwide reading approach, the Collaborative for Teaching and Learning Adolescent Literacy Model (ALM). Within schools using ALM, struggling readers were randomly assigned to receive an extra daily reading class using the Learning Strategies Curriculum (LSC). This section describes the experiment evaluating the ALM intervention and outcomes.

ALM (Awbrey, 2008) is a professional development approach designed to help teachers of all subjects use effective literacy strategies. As part of the Kentucky Striving Readers study, ALM was evaluated in a matched design (Cantrell, Almasi, Carter, & Rintamaa, 2011). Nineteen control schools across Kentucky were matched with 19 experimental schools based on state test scores and demographics. ALM did not provide specific texts or other materials, but gave all teachers in grades 6-12 at least 11 days of professional development over a 4-year period, as well as ongoing support provided by on-site coaches and mentoring. The PD focused on vocabulary development, reading comprehension, fluency, writing, and academic dialogues.

Yearly composite school scores on KCCT were used to compare progress of intervention and control schools. Reading effect sizes, adjusted for pretests, averaged +0.10 (n.s.).

Across all seven studies of SIM variations, the average effect size was +0.09.

Reading Apprenticeship – Whole class/school. Reading Apprenticeship refers to 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 teachers' 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.

Greenleaf, Hanson, Herman, Litman, Rosen, Schneider, & Silver (2011) evaluated Reading Apprenticeship in grade 9-11 biology and history classes, but because there were substantial pretest differences in the history classes, we only report the results for biology classes. Also, because of a lack of comparable tests in other states, only schools in California were included. Two subgroups were studied, those who had parent permission to be followed longitudinally and those who did not, who were only measured cross-sectionally, over a three-year period from grades 9 to 11. 78 schools (39 E, 39 C) and 111 teachers (56 E, 55 C) were randomly assigned to conditions, and pre- and post-tested on Degrees of Reading Power (DRP) and California State Test (CST)-ELA and Reading Comprehension. There were no significant differences between experimental and control groups. Effect sizes were -0.04 (n.s.) on DRP. Averaging the longitudinal and cross-sectional samples, effect sizes were +0.10 on CST-ELA (n.s.) and +0.11 on Reading Comprehension (n.s.).

Reading Apprenticeship Academic Literacy (RAAL) – Targeted. Reading

Apprenticeship Academic Literacy (RAAL) (Schoenbach, Greenleaf, Cziko, & Hurwitz, 1999) is

an intensive program for struggling secondary readers created by WestEd. It focuses on

motivation and engagement, on the study of language and text structures, and writing. A detailed

curriculum is provided, but teachers have the freedom to choose the aspects of the curriculum

they want to work on.

RAAL was evaluated in a large, randomized experiment by Somers et al. (2010). In this

study, 34 high schools were randomly assigned to use either RAAL or Xtreme Reading,

described previously. Then ninth graders who were reading between 2 and 5 years below grade

were randomly assigned within 17 schools to receive RAAL or to remain in their ordinary non-

literacy electives. Two cohorts were evaluated in the same design. Averaging across them,

students who participated in RAAL scored significantly better on GRADE Reading

Comprehension (ES=+0.12, p<.01) than did controls, but there were no differences on GRADE

Vocabulary, for a mean effect size of +0.08. There were also significant differences on the ELA

state test (ES=+0.15, p<.01). In a one-year follow-up, however, no significant effects were found

to maintain into tenth grade. Students in Xtreme Reading did not perform better than controls.

Reading Apprenticeship Improving Secondary Education (RAISE) – Whole

class/school. Reading Apprenticeship Improving Secondary Education (RAISE) is a whole-class

program related to RAAL, both of which were developed by overlapping teams at WestEd.

RAISE integrates literacy learning into content areas by offering professional development to

high school teachers. Teachers are instructed to reconsider the role of the students and of the

teachers (who become resources for learning in a collaborative environment), and to encourage

the students to establish personal reading goals. Teachers are trained to develop text-based problem-solving strategies, to work on discipline-specific discourse, and engage students in "metacognitive conversations," discussing several dimensions of the reading process.

A large quasi-experimental study was carried out by Fancsali et al. (2015) in California and Pennsylvania. Voluntary teachers from 42 high schools were offered to implement RAISE in their ELA, biology, or history classes, or to the control group, pursuing their business-as-usual activities (E=22 schools, 130 teachers; C=20 schools, 122 teachers). Despite randomization, the use of volunteer teachers within randomly-selected schools makes this a quasi-experiment. Grade 9 to 12 students received up to two years of treatment. At the end of the second year, students in the experimental schools did not perform significantly better (ES=+0.14, n.s.; n students=5531 E, 4642 C) than students in the control schools on the assessment of literacy achievement developed by ETS, adjusted for scores on the 8th grade state reading/ELA tests and demographic characteristics.

iRAISE – Whole class/school. 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. iRAISE teachers were given 65 hours of Reading Apprenticeship professional development, starting with a 5-day training in the summer and monthly follow-up meetings.

Jaciw, Schellinger, Lin, Zacamy, & Toby (2016) evaluated iRAISE over one year, in a cluster randomized trial. 69 teachers of grades 9-12 were randomly assigned to treatments (35 E, 34 C) within 26 schools. There were 1468 students (751 E, 717 C). Students were pre- and post-

tested on the ETS Literacy Assessment. There was no significant impact on reading test scores (ES= 0.00, n.s.).

Across four studies of variations of Reading Apprenticeship, the weighted mean effect size was +0.11.

Content Knowledge-Building and Student-Regulated Comprehension Practices - Whole class/school. Content Knowledge-Building and Student-Regulated Comprehension Practices is designed to improve students' reading comprehension. Each cycle of the intervention includes three main steps. 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. During the second phase, students work in pairs to analyze the text and answer critical questions. The last part alternates individual, partner, and group work on text synthesis. In a study by Simmons et al. (2014), teachers were asked to implement the project three days a week (120-150 minutes) during their regular English Language Arts classes over one semester. Teachers participated in one day of professional development, met research staff for feedback over the course of the project, and received lesson plans and materials for expository and narrative texts.

The research team conducted a cluster randomized experimental study in six Title I middle and high schools from one southwestern state. Grade 7 to 10 classrooms were assigned at random to experimental or control conditions (n classes=36 E, 29 C). The 17 participating teachers taught both intervention and control classrooms (n students=413 E, 373 C). Control groups were asked to maintain their typical practices (including partner work and small-group

discussions) and material. On the Gates MacGinitie Reading Comprehension subtest, treatment students performed at the same level as control students (ES=-0.01, n.s.).

Across all 12 studies of metacognitive approaches, the weighted mean effect size was +0.09.

Project CRISS – Whole class/school. Project CRISS (Creating Independence Through Student-Owned Strategies) is a professional development approach designed to help teachers of all subjects use proven reading comprehension strategies. The project provides summer institutes on Project CRISS strategies to local facilitators, who then create local teacher-to-teacher study groups who help each other implement the approach. A large randomized evaluation of CRISS in ninth grades in the Northwest (Kushman, Hanita, & Raphael, 2011) found no significant differences on the Stanford Diagnostic Reading Test (ES=+0.05, n.s.)

REACH – Targeted. REACH (not to be confused with the unrelated U.K. tutoring program of the same name) is a comprehensive reading/language arts program designed to provide explicit, intensive instruction to students in grades 6-12. It uses a direct instruction approach for grades 6 through 12 students reading significantly below grade level. REACH is a combination of three programs distributed by McGraw Hill Education (https://www.mheonline.com): Corrective Reading, Reasoning and Writing, and Spelling Through Morphographs. REACH also incorporates placement and progress assessments, and chapter books. Corrective Reading focuses on phonics, fluency, word knowledge, and reasoning skills. Reasoning and Writing works on knowledge of narrative structure and writing skills.

Spelling Through Morphographs provides strategies to spell words. Tests are administrated regularly to evaluate progress and determine areas in need of additional work.

REACH was field-trialed one year before the large experiment by Lang, Torgesen, Vogel, Chanter, Lefsky, & Petscher (2009). In the evaluation, REACH was used 90 minutes daily, replacing an elective course in seven schools, by ninth grade students randomly assigned to one of three possible interventions or to a control group, which used the schools' existing supplemental program. Experimental as well as control students also experienced a reading strategies approach in social studies and sciences given to all ninth graders in the school. REACH did not yield significant results for treated students, whatever their initial level of performance. The effect size on the reading component of the FCAT was -0.19 (n.s.) for the "high risk" readers (n=91 E, 90 C), and +0.06 (n.s.) for the "moderate risk" readers (n=199 E, 202 C), for a weighted average of -0.02.

Across all 12 studies of metacognitive strategy programs, the weighted mean effect size was +0.09.

==================

TABLE 5 HERE

==================

Mixed-Model Professional Development

Studies of mixed-model professional development strategies, where teachers received PD and coaching on a variety of instructional strategies, were particularly common among Striving Readers grantees, especially when the state and local district grantees developed their own programs. Seventeen studies, summarized in Table 5, evaluated programs assigned to this category.

53

Teacher Effectiveness Enhancement Programme (TEEP) – Whole class/school. The 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. A smaller group receives two additional days and members of this group work with their peers to embed the TEEP strategies schoolwide. The strategies include formative assessment, thinking skills, cooperative learning, and effective use of technology.

In a large, cluster randomized experiment, the Institute for Effective Education (IEE; 2016) evaluated outcomes of TEEP on GCSE scores, the test taken by all English students at the end of secondary school. Forty-five schools (23 E, 22 C) were randomly assigned to TEEP or control conditions in two cohorts, for a treatment period of 1.5 to 2 years. There were 10,893 students in total (5,327 E, 5,058 C). On GCSE English scores, controlling for scores at the end of primary school (Key Stage 2), outcomes were essentially zero (ES= -0.04).

Word Generation - Whole class/school. 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.

Previous studies of Word Generation have found positive effects on measures of knowledge of the target words, but not on general vocabulary or comprehension. Lawrence, Francis, Paré-Blagoev, & Snow (2016) carried out a large cluster randomized trial to determine whether the program had impacts beyond the target words. Forty-four middle schools (n=25 E, 19 C) in three urban districts were randomly assigned to conditions. Teachers of all grades (6-8) and all academic subjects were trained in Word Generation strategies. Students (n=4796 E, 3670 C) were pre- and posttested on Gates McGinitie Reading Comprehension and Vocabulary scales.

HLM analyses showed no significant effects both for Reading Comprehension (ES=+0.07) and Vocabulary (ES=0.00).

ANet (Achievement Network)-Whole class/school. ANet is a program designed to help teachers and school leaders use data more effectively to identify and close gaps in student achievement. The program provides quarterly interim assessments in English and math, data tools to report on students' progress, coaching of school leaders in effective data use, and networks of peer schools to share results and engage in joint professional development.

Under i3 funding, West, Morton, & Herlihy (2016) carried out a large-scale randomized evaluation of ANet in reading and math in grades 3-8. The grade 6-8 reading results are reported here.

Initially, 89 schools (45 E, 44 C) in urban districts in Massachusetts, Jefferson Parish (LA), and Chicago were randomly assigned to receive ANet immediately (experimental) or two years later (control). However, 10 schools dropped out of ANet after random assignment. Their matched pairs were also dropped, leaving 69 schools (35 E, 34 C; 8070 students). Data were obtained from the districts and converted to z-scores to allow for combining. An intent-to-treat analysis involving all 89 schools found no significant effect of ANet on reading in grades 6-8 (ES= -0.03). Analyses of the 69 "reduced sample" schools also found no differences (ES=+0.02).

Chicago Striving Readers – Whole class/school. The Chicago Striving Readers project was conceived as a multitier intervention. In the schools randomly assigned to the project, every grade 6 through 8 student received a whole-school intervention (tier 1). Students who could reach grade level with support in their regular English Language Arts classroom received, in addition to the whole-school intervention, a targeted intervention (tier 2), offered mostly to 6th

graders. Sixth grade struggling readers (more than one year below grade level) received an intensive intervention given after the regular school day (tier 3) to supplement the two first tiers.

The whole-school intervention involved changes in the organization of the classes: during English Language Arts block, as well as during content-area classes, teachers were encouraged to start and conclude the lesson with the entire class, but the main instruction was devoted to practice in small groups where students rotated among three different workshops. Technology tools (media and listening centers, handheld computers) were provided to support student learning during small group rotation. All teachers were encouraged to use partner reading to foster fluency, comprehension and vocabulary, and Marzano's (2004; Marzano & Pickering, 2005) vocabulary model was used to provide direct instruction of subject-specific and technical words. Words Their Way was implemented for word study. Engaging informational texts regrouped in text set units were developed and provided to social studies, science, and math teachers.

Professional development was provided to accompany these instructional changes. Control schools were engaged in school-wide literacy initiatives, using some similar components such as data-driven instruction, and implementing literacy programs such as READ 180, Reading First, or the Balanced Literacy model, among others.

A large cluster randomized evaluation in 30 schools of 8127 students (n students=4074 E, 4053 C) was carried out by Simon, Tunik, Alemany, Zhu, Zacharia, Ramsay, Swann, Fields, & Mendes (2011). It failed to find an effect of the whole-school intervention in the participating schools, after 1-3 years of the intervention, compared to the progress made on the state reading test by the 31 comparison schools. The effect size was near zero (ES=-0.01, n.s.) at the end of the fourth year. Effect sizes computed on sixth graders only were similar.

REWARDS – Targeted. The Reading Excellence Word Attack and Rate Development Strategies (REWARDS) program created by Archer, Vachon, and Gleason is a family of reading and writing intervention materials for grade 4 to 12 struggling readers.

Three REWARDS components were used in the New York Striving Readers project: the REWARDS Secondary, aimed at mastering fluency and academic vocabulary; the REWARDS Plus, aimed at developing reading strategies in science and social studies passages; and REWARD Writing. REWARDS was implemented as a one-year supplemental intervention given by trained teachers five times a week during the school day. Control students attended additional class sessions in the content subjects (i.e., science, social studies) or talent/enrichment classes (e.g., art, music).

Newman, Kundert, Spaulding, White, & Gifford (2012) performed a large independent evaluation of the New York Striving Readers project. The eleven participating schools were located in four boroughs of New York City and served mostly disadvantaged, ethnically diverse sixth to eighth grade students. Struggling seventh graders were randomly assigned to the intervention. The effects of the intervention were positive and marginally significant for treated students on the NYS-English Language Arts Assessment (ES=+0.15, p=.06; n=253 E, 264 C) but close to zero on Gates-MacGinitie Reading Comprehension (ES=+0.02; n.s.; n=232 E, 237 C), with a mean of +0.09 across the two measures.

Kentucky Cognitive Literacy Model – Targeted. The Kentucky Cognitive Literacy Model (KCLM) is a targeted intervention for struggling readers in high school. Designed by the Kentucky Department of Education (KDE), the KCLM provided an intervention teacher to each participating school to teach a year-long reading class, replacing an elective such as art or music. Participating teachers in 9 rural high schools received 11 days of professional development and

up to 66 hours of support visits from KDE staff. The intervention focused on teaching comprehension strategies, vocabulary, study skills, and writing. Teachers received templates to create or adapt units of instruction, rather than specified instructional materials.

In the evaluation of KCLM (Cantrell, Carter, & Rintamaa, 2012), ninth graders who scored two levels or more below grade level were identified in the spring of 8th grade and randomly assigned to receive KCLM or to participate in their usual electives. On spring GRADE scores, the overall effect size for reading was not significantly different from zero (ES=-0.06, n.s.; students n= 232 E, 253 C).

Reading Intervention through Strategy Enhancement (RISE) - Targeted . RISE (Lefsky, 2004) relies on teachers' capacity to build effective curriculum for struggling readers, provided they are given adequate time and strong professional development. Teachers are taught how to create rich collections of thematically-related texts varying in difficulty and genre, plans to assess students' progress, and plans for differentiated instruction when students are not making the desired improvements. During the RISE classes, students are given the opportunity to read independently, to work in small groups, and to receive whole group lessons.

RISE, developed by the Florida State Department of Education, was one of three interventions evaluated by Lang, Torgesen, Vogel, Chanter, Lefsky, & Petscher (2009) in seven high schools in a large Florida district. To prepare teachers, RISE was first implemented the year before the experiment began. Then a large randomized experiment was undertaken with grade 9 students placed in "high risk" or "moderate risk" intervention classes according to their scores on the Florida Comprehensive Achievement Test. Control students, it is important to note, were also receiving a 90-minute supplemental reading block instead of an elective, but this supplemental instruction varied by school, mostly using commercial materials, as a continuation of what they

had already been doing. Both experimental and control students were exposed to reading strategies instruction in science and social studies.

Among the high risk students (n=104 E, 90 C), those who received RISE had no significantly lower scores on FCAT reading than control students who received the school-designed interventions (ES=-0.06, n.s.). By contrast, among "moderate risk" students (n=204 E, 202 C), those who received RISE made more progress than the control group on FCAT Reading. The effect size was +0.27 (p=.04). The weighted average effect size was +0.16.

Building Assets Reducing Risks (BARR) - Whole class/school. Building Assets Reducing Risks (BARR) is a whole-school reform approach focused on "developmental, academic, and structural challenges during the ninth grade year." BARR is not limited to reading, but is used in all subjects across the ninth grade to attempt to increase student achievement by improving students' social-emotional skills, 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 schools 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) carried out a within-school randomized evaluation of BARR in a large high school in Southern California. Ninth graders were randomly assigned to receive BARR services or to continue their ordinary program for one year. Students were pre- and

posttested on NWEA Reading. Controlling for pretests, the effect size was +0.14, p<.01

(students n = 261 E, 234 C).

Read to Achieve – Targeted. Read to Achieve

(https://www.mheonline.com/onlinesamples/program.php?subject=1&program=11&p=3) is a

program that emphasizes comprehension, vocabulary, and fluency. It provides units of five class

lessons each on content area and narrative texts, and incorporates small group collaboration and

independent activities.

In the Washington State Striving Readers project, two groups of sixth to eighth graders

were separately assigned to experimental and control conditions. Experimental students in Group

1, which consisted of students who scored two grades below their grade level on the state reading

assessment and experienced decoding difficulties, were assigned to receive supplemental reading

instruction starting with a separate program called Phonics Blitz, followed by Read to Achieve.

Experimental students in Group 2, struggling readers who did not need phonics intervention,

received Read to Achieve the entire year. The two control groups were placed in study hall or an

elective. Both groups were randomly selected from six Title I secondary schools from three

districts in Western Washington.

Deussen, Scott, Nelsestuen, Roccograndi, & Davis (2012) found nonsignificant positive

effect sizes for Group 1 on the GMRT (ES=+0.13; n=32 E, 31 C) and on the state reading state

test (ES=+0.11, n.s.; n=37 E, 39 C), for an average of +0.12 across the two measures. For Group

2 there were no differences between the treatment and control groups on the GMRT (ES=+0.02,

n.s.; n=144 E, 151 C), but there was a marginally significant positive effect (ES=+0.16, p=.07;

n=192 E, 191 C) on the state test. The average effect size for Group 2 was +0.09, and the mean

weighted effect size across Groups 1 and 2 was +0.10.

Expert 21 – Whole class/school. Expert 21 is a comprehensive English language arts curriculum for grades 6-9. It 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. Curriculum-embedded assessments based on the online version of the Scholastic Reading Inventory are used. Teachers receive two days of professional development.

Sivin-Kachala & Bialo (2012) carried out a year-long evaluation of Expert 21 in an urban, disadvantaged middle school in New Jersey. Students were randomly assigned to classes (n=276; 137 E, 139C) and their teachers were randomly assigned to conditions (n=6; 3E, 3C). On the state test, the NJ Ask, controlling for pretests, students in Expert 21 gained non-significantly more than controls on Language and Literacy (ES=+0.22, p<.07) and Reading Comprehension (ES=+0.18, n.s.). On GMRT Comprehension, there were also no significant differences (ES=+0.10, n.s.). The average effect size across the NJ Ask and GMRT was +0.15 (n.s.).

Strategies for Literacy Independence Across the Curriculum (SLIC) – Targeted. A San Diego Striving Readers project evaluated a program called Strategies for Literacy Independence Across the Classroom, or SLIC (Hofstetter et al., 2011). The model was adapted from a New Zealand program (McDonald & Thornley, 2004). In San Diego, SLIC was used and evaluated in two forms. The whole school version is described below. This section describes a supplemental, targeted model, in which students experienced SLIC in daily classes replacing one elective (e.g., art or music).

In the targeted form of SLIC, students reading at least two years below grade level on any of three standardized tests were taught literacy strategies such as authors' use of different text

forms for different types of information and how the surface features of a text convey information about the text. Students used a variety of persuasive, expository, and narrative texts, including content-area texts, magazines, newspaper articles, short stories, and novels.

In the evaluation, students in grades 7 to 10 in 8 middle and high schools were randomly assigned to the targeted intervention or to their ordinary elective. Students were eligible for up to three years of intervention. Teachers received extensive professional development and coaching.

Pre- and posttest measures were available on the California Standards Test (CST) and the Degrees of Reading Power (DRP). For students who received two years of treatment (n=782 E, 792 C) the effect size on the CST was +0.01 (n.s.), and +0.08 (n.s.) for DRP. For students who received three years of intervention (n=305 E, 301 C), the effect size was -0.03 for CST, and +0.03 for DRP. None of these differences were statistically significant.

Strategies for Literacy Independence across the Curriculum (SLIC) – Whole class/school. The San Diego evaluation of the whole-school version of SLIC (Hofstetter et al., 2012), compared all students in eight SLIC and eight matched control schools. The model provided professional development to all content teachers of grades 6-12 on literacy strategies derived from the SLIC strategies, described earlier. A comparison using hierarchical linear modeling (HLM) found no significant difference, with an effect size of +0.05 (n.s.) on the CST (n students=4915 E, 6823 C) and +0.02 (n.s.) for the DRP (n students=2234 E, 2859 C).

Every Classroom, Every Day (ECED) – Whole class/school. Every Classroom, Every Day (ECED) is a program for math and literacy developed by the Institute for Research and Reform in Education (IRRE) for high school students. It relies on a professional development program based on Desimone's (2009) model, which focuses on active learning. "Literacy

matters" – a structured literacy curriculum based on authentic expository texts – supplements the regular English course, doubling the amount of English Language Arts exposure.

ECED was implemented in ten high schools over two years (starting at grade 9) as part as a large cluster randomized experiment independently evaluated by Early, Berg, Alicea, Si, Aber, Ryan, & Deci (2015). Almost all ninth graders were eligible for this two-year supplemental literacy program. IRRE sustained the process by visiting the schools four times yearly to provide on-site professional development, and by supporting instructional coaches hired by the schools for this project. Control schools were asked to pursue their "business as usual" activities, including coaching and professional development. Over the two years, the progress made by the intervention students were not significant compared to their counterparts in control schools on the state English Language Arts tests (ES=+0.06, n.s.; n students=3935 E, 4315 C).

Expository Reading and Writing Course – Whole class/school. The 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 or can go directly to credit-bearing English coursework. Because it was designed for students expecting to go to college, the program (and study) excluded students in special education for learning difficulties as well as other state-defined categories of low achievers.

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. ERWC replaces ordinary English classes for the 12th grade year.

A quasi-experimental evaluation of ERWC was carried out by Fong, Finkelstein, Jaeger, Diaz, & Broek (2015). Using propensity matching, students in ERWC were matched on prior achievement and demographic variables with similar students in ordinary English classes. ERWC teachers had to have at least one year of experience with the model. There were a total of 56 ERWC and 58 non-ERWC teachers in 24 high schools throughout California. The final analytic sample comprised 3309 ERWC and 3309 closely matched non-ERWC students. The sample was 45% Hispanic, 27% Asian, 24% White, and 4% African-American. On EPT posttests at the end of the school year, ERWC students scored modestly higher (ES=+0.13). This difference was significant after adjustment for clustering was done by the What Works Clearinghouse (Fong & Finkelstein, 2016).

Schoolwide Enrichment Model – Reading Framework (SEM-R)- Whole class/school. SEM-R is an instructional program in which students are exposed 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 large cluster randomized study involving 47 sixth to eighth grade teachers randomly assigned to experimental or control conditions was carried out by Little, McCoach, & Reis (2014). Four middle schools serving a large proportion of low-income students participated. Students (n students=1198 E, 830 C) were pre- and posttested on the Gates MacGinitie Reading comprehension subtest (GMRT). The GMRT posttests, controlling for pretests, showed a small, non-significant advantage (ES=+0.10, *n.s.*) for students taught by experimental teachers.

Across all 17 studies of programs categorized as mixed-model professional development, the weighted mean effect size was +0.05.

===============

TABLE 6 HERE

===============

Programs for English Language Learners

Four studies of two programs, Pathways and ALIAS, focused on improving reading outcomes for English language learners and other language minority students. Table 6 summarizes the findings of these studies.

Pathway – Whole class/school. Pathway is a professional development program used primarily with mainstreamed Latino English learners able to participate in regular English classes. It was created in a partnership between the National Writing Project at the University of California-Irvine and the Santa Ana (CA) Unified School District. Pathway is primarily focused on writing rather than reading, but it does also focus on reading comprehension, and studies of Pathway include measures of reading. The programs provide substantial PD. In Kim et al. (2011), 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.

In a large cluster randomized study carried out over two years, teachers in 15 disadvantaged schools were randomly assigned to experimental or control conditions and then students were randomly assigned to classes (Kim et al., 2011; Olson et al., 2012). During the first year, 48 classrooms of intervention teachers were assessed on the California Standards Test (n

65

classes=48 E, 47 C; n students=1421 E, 1305 C). In year two, 31 new intervention classrooms and 35 controls were evaluated (n students=779 E, 954 C). After the first year of intervention, treatment classrooms performed significantly better than controls. The effect size on the CST for grade 6 to 11 students was +0.07 (p<.05). The second year, the effect size was identical, but not statistically significant.

A second evaluation of Pathway by Olson, Matuchniak, Chung, Strumpf, & Farkas (2016) involved California tenth graders. Sixty eight percent were Hispanic and 21% qualified as English learners. Teachers were randomly assigned to treatments within 16 secondary schools (n students=575, 313 E, 262 C). On California High School Exit Examination (CAHSEE) English Language Arts scores, Pathway students scored non-significantly higher at the cluster level (ES=+0.19), controlling for California Standards Test pretests.

The weighted average across the two Pathway studies was +0.08.

Academic Language Instruction for All Students – ALIAS – Whole class/school.
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 on which students work deeply. 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. Two large cluster randomized quasi-experimental studies have been performed by Lesaux and her colleagues in grade 6 classrooms including a large proportion of Hispanic students. The first (Lesaux, Keiffer, Faller, & Kelley, 2010) involved volunteer teachers from seven middle schools in a southwestern state, selected at

random into experimental or control groups (n students=296 E, 180 C). Students were pre- and posttested on Gates-MacGinitie (GMRT) Reading Comprehension. Experimental-control differences were marginally significant (ES=+0.15, p=.06). In the second study (Lesaux, Keiffer, Kelley, & Russ Harris, 2014), 50 teachers and their classes were randomly assigned to either the treatment or control conditions within 14 urban middle schools in California (n students=971 E, 1111 C). In this study, the effect size on GMRT Comprehension was only +0.04 (n.s.), but on GMRT Vocabulary, the effects were positive and significant (ES=+0.17, p<.002). Combining these produces an overall GMRT effect size of +0.08.

The weighted mean effect size across the two ALIAS studies was +0.09, and the weighted mean for the four studies of programs for ELLs was +0.08.

Differences by Design Factors

Targeted vs. whole-class/whole-school interventions. In many of the qualifying studies, the intervention was provided to target groups of low achievers. 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 targeted treatments, group sizes were typically small (usually 12 to 20), and instruction was intensive. In contrast, many other studies provided interventions to entire classes or schools. In these studies, experimental teachers were doing something different in their classes from what control teachers were doing, but there was no additional teaching time, and group sizes were the same in experimental and control classes. For these reasons, one might expect that targeted treatments would have greater impacts on reading than whole class or whole school approaches.

An analysis comparing targeted to whole-class/whole-school interventions among the six categories of programs other than tutoring (Tables 2 to 6) found no difference between targeted

and whole-class interventions in achievement impacts. Thirty-seven whole-class studies had a mean weighted effect size of +0.07, while 23 extra-time targeted programs had a weighted mean effect size of +0.08.

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). Some studies that included both grade levels failed to distinguish between their outcomes, and others showed separate middle and high school outcomes. Taken together, weighted mean effect sizes were non-significantly higher for middle schools (n=36 studies, ES=+0.12) than for high schools (n=25 studies, ES=+0.08).

Differences by Research Design

We compared effect sizes between studies that used random assignment to conditions (n=56 studies, ES=+0.04) and quasi-experiments, which used matching (n=9 studies, ES=+0.11). This non-significant difference is similar to the differences reported by Cheung & Slavin (2016). Effect sizes were non-significantly lower for studies using clustered designs (n=33, ES=+0.06) than for those using individually randomized designs (n=32, ES=+0.11).

Discussion

This review of rigorous research on programs 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, two quite different categories contained programs with more positive impacts. One of these was tutoring programs, all done in England, in which teachers worked with groups of one to four students (or in one case, two adults to 6-8 students).

Three of the five tutoring studies found significant positive effects. The weighted mean effect size across all five studies was +0.23. It is not surprising that tutoring would be very effective, as it has also been in elementary reading (Slavin, Lake, Davis, & Madden, 2011). Tutors are able to build individual relationships with students, and to personalize instruction to their individual needs. Because tutors see so few students at a time, one might argue that the large effect sizes do not mean as much as they appear to. For example, if a program for four students at a time gets an effect size of +0.30, while a program for 16 students at a time gets an effect size of +0.15, which is more effective? Considered at the student level, it is the tutoring, but at the class level the more distributed benefit might be preferred.

Another category with particularly positive outcomes is cooperative learning, in particular The Reading Edge, with a weighted mean effect size of +0.29, and the related Talent Development High School's Strategic Reading Approach (ES=+0.32). 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 a very different 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.

Overall impacts of technology approaches were modest (mean ES=+0.08), but there were some promising programs. The weighted mean effect size for Achieve3000 was +0.29 (though not significant at the cluster level). An English study of a form of Accelerated Reader using technology reported an effect size of +0.24. Comprehensive Circuit Training had a mean effect

size of +0.13. Of five studies of READ 180, two found significant but modest 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 Voyager Passport found significant positive impacts but two did not, and the mean weighted effect size was only +0.06.

Beyond tutoring, cooperative learning, and technology, there were other programs with statistically significant but usually modest effect sizes, around +0.10 to +0.14, that could best be described as promising. A few programs reported significant positive outcomes in some studies but not others, and weighted means for these programs were less than +0.10.

Among the programs focusing on teaching metacognitive skills, two families of approaches are worthy of particular notice. One is the Strategic Instruction Model (SIM), based on development and research at the University of Kansas. These programs teach struggling adolescent readers step-by-step strategies for phonics, comprehending, writing, note-taking, and other skills, each with a series of mnemonics to help students recall the steps. Seven studies evaluated variations of this approach. The outcomes of these studies, in very different locations and circumstances, were remarkably consistent, with most effect sizes clustering around a mean of +0.09. What these findings indicate is a quite modest but highly reliable impact.

Another interesting family of programs is Reading Apprenticeship, from WestEd (Greenleaf et al., 2011). Reading Apprenticeship also focuses on comprehension and writing strategies. Unlike the SIM programs, however, the four variations of Reading Apprenticeship have very different outcomes. The only one of them to have significant positive effects is Reading Apprenticeship Academic Literacy (RAAL; Somers et al., 2010), a targeted version (which provided an additional instructional period each day for low achievers). However, the mean effect size was only +0.10. Reading Apprenticeship Improving Secondary Education

(RAISE; Fancsali et al., 2015) had an effect size of +0.14, but due to its use of a clustered design this was not quite significant. The other two RA variations had effect sizes near zero.

In the Mixed-Model Professional Development category were three models with single studies showing promise. 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. Another was BARR, a whole-school model that emphasizes social-emotional development, building relationships between teachers and students, and closely monitoring students' academic progress (ES=+0.14). Expository Reading and Writing Course (ERWC), a program designed to help twelfth graders prepare for the test they will take as freshmen in California State Universities required to gain access to credit-bearing (i.e., not remedial) English courses, reported a significant positive effect size of +0.13. Pathway, a professional development program for teaching mainstreamed English learners, also showed some promise (ES=+0.07).

The findings of this review provide some hints about a general theory of action for secondary reading. This theory of action begins with a conception of the adolescent learner. The studies that met the inclusion criteria for the present review primarily involved low-skilled readers, either students specifically targeted for a remedial class or students in high-poverty, low-achieving schools taught in whole classes or schools. Either way, an adolescent lacking confidence in reading is unlikely to be eager to read, unlikely to have a positive self-esteem as a reader, and perhaps good at finding ways to avoid reading or avoid exposing his or her poor reading skills.

Most of the programs evaluated under Striving Readers grants, and those in the metacognitive and mixed-model professional development categories, are based on a theory of

action emphasizing a need for teaching skills previously taught in elementary school to students who did not learn them previously. Teaching students in groups of 12 to 20 using teaching methods and content students are likely to have seen before does not seem to be a formula for engendering motivation or enhanced self-concept. The modest effect sizes generally found in studies of targeted programs, which often added an hour a day all year for extra teaching, is a repudiation of the idea that what struggling adolescent readers need is more and better remedial teaching. Effect sizes for these extra-time interventions were no better than those for programs that used existing time in different ways.

Now consider the approaches that did seem to make a difference. These apparently dissimilar approaches share several important features. One relates to motivation. For example, tutoring and cooperative learning programs, plus programs focused on social-emotional development and relationships such as BARR (ES=+0.14), engage students with valued others. Successful approaches also provide extensive personalization. A student's tutor and teammates know what he or she can do, they know what he or she needs to do next, and they can adapt instruction to his or her learning style and needs. The modest effects of most technology approaches, with the notable exceptions of Achieve3000, Accelerated Reader, and Comprehensive Circuit Training, all of which combine technology applications with teaching by teachers, may suggest that personalization by machine is not sufficient, or at least not as effective as personalization by caring teachers and peers. Note that these successful strategies are very different from ordinary teaching. Not only does this make them novel, which may contribute to motivation, but it also may give otherwise despairing students a belief that this time, things will be different, and that this time, if they apply themselves, they may see significant learning gains, perhaps enough to make learning a source of pride rather than shame. An excellent example of

this is ERWC (ES=+0.13), a program designed to prepare twelfth graders to succeed on a test that enables them to skip freshman remedial English in college. Taking ERWC is probably seen as an honor, not a remediation, to most students, because it is preparing them directly for a bright future. Similarly, SEM-R (Schoolwide Enrichment Model-Reading; ES=+0.10) provides high-poverty schools with instruction posed as enrichment, not remediation.

While explanations for the impacts of tutoring, cooperative learning, and other promising approaches are somewhat speculative, there is one conclusion from this review that seems well-justified. No program that showed positive effects in this review involved anything like traditional teaching. If secondary schools are to make real breakthroughs with struggling readers, they are going to have to do something much more motivating, more personalized, and more likely to give students a belief in their own capacity for learning. Most of the innovations tested in the studies that met the inclusion criteria may have been too much like what students had likely experienced for years before. A secondary student taking a course he or she perceives as remedial, no matter how small the class size, how much extra time is allocated, or how well-designed the content and teaching, is likely to be watching the clock and wishing he or she were in the control group, which is usually taking band or art at that time.

The research reviewed here may provide evidence of what *not* to do to accelerate the learning of struggling secondary readers. It provides some 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.

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.

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.* Research report prepared for the U. S. Department of Education, Washington, D. C.

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.

Cantrell, S. C., Carter, J. C., & Rintamaa, M. (2012). *Striving Readers Cohort II Evaluation Report: Kentucky*. KY: Collaborative Center for Literacy Development, University of Kentucky.

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. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45* (5), 283-292.

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.

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.* i3 Development Grant. Final Report. Grantee Submission.

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: Author.

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.

Desimone, L.M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualization and measures. *Educational Research*, *38*, 181-199.

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

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.* 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

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.

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 11-01-16 from http://www.tandfonline.com/doi/abs/10.1080/19345747.2016.1227412?journalCode=uree20.

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.

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: 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.

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.

 high school. Washington, DC: Alliance for Excellent Education.

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.

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*

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.

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. *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.

Lawrence, Francis, Pare-Blagoev, & Snow (2016). The poor get richer: Heterogeneity in the

    efficacy of a school-level intervention for academic language. *Journal of Research on*

    Educational Effectiveness. Retrieved 11-01-16 from

    www.tandfonline.com/doi/full/10.1080/19345747.2016.1237596

Lefsky, E. B. (2004, April). The effect of text set use in the secondary content classroom on

    adolescent literacy achievement and self-efficacy. Paper presented at the annual meeting

    of the American Educational Research Association, San Diego, CA.

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.

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.

Lord, P., Bradshaw, S., Stevens, E. and Styles, B. (2015). *Perry Beeches Coaching Programme:

  Evaluation Report and Executive Summary*. London: Education Endowment Foundation.

Marzano, R. J. (2004). *Building background knowledge for academic achievement.* Alexandria,

Marzano, R. J., & Pickering, D. J. (2005). *Building academic vocabulary: Teacher's manual.*

  Alexandria, VA: Association for Supervision and Curriculum Development.

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.

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*.

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

OECD (2013). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: 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.

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*.

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.

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

Schenck, A., Feighan, K., Coffey, D., & Rui, N. (2011). Memphis Striving Readers project: Evaluation report, Year 4. 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.

Scholastic, Inc. (2006). *Scholastic's READ 180: A heritage of research*. New York, NY.

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). *Texas Center for Educational Research*.

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

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

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

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. (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.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*.*

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. 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. (1992). *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.

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.

VA: Association for Supervision and Curriculum Development.

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.

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.

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

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 11/07/2016 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.

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- Targeted								
Rutt et al. (2015)	Student randomized	30 weeks	Students: 557 (286 E, 271 C)	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- Targeted								
Lord et al. (2015)	Student randomized	1 year	Students:291 (149 E, 142 C)	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) – Targeted								
Sibieta (2016)	Student randomized	20 weeks	Students: 202 (70 REACH, 69 REACH + LC, 63 C)	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- Targeted								
Merrell & Kasim (2015)	Student randomized	4 months	Students:310 (161 E, 149 C)	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 - Targeted								
King & Kasim (2015)	Student randomized	7 months	Students:178 (86 E, 92 C)	6-7	Students from 22 primary and 13 secondary schools in Norfolk Country, England who were reading at least one year below grade level and had decoding difficulties. 50% W, 50% SPED, 50% ELL, 50% FRL.	NGRT	-0.05	-0.05

Notes for Tables 1-6

CAHSEE: California High School Exit Examination, CAT: California Achievement Test, 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), FCAT: Florida Comprehensive Assessment Test, GORT: Gray Oral Reading Test, GRADE: Group Reading Assessment and Diagnostic Evaluation, GMRT: Gates-MacGinitie Reading Tests, iLEAP, Louisiana State Reading Assessment, ISAT: Illinois Student Achievement Test, 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, 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. Cooperative Learning Programs

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
The Reading Edge/Student Team Reading – Whole Class/School								
Slavin, Chamberlain, Daniels, & Madden (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. >60% FRL.	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)	6,7,8	5 mostly AA, mostly FRL middle schools in Baltimore, Maryland.	CAT Comprehension	+0.34	+0.38
						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	6 mostly AA middle schools in Baltimore, Maryland. 75% AA, 58% FRL.	CAT Comprehension	+0.13	+0.08
						CAT Vocabulary	-0.02	
Talent Development High School (Strategic Reading and Student Team Writing) – Whole Class/School								
Balfanz et al., 2014	Cluster quasi-experiment	1 year	Schools: 6 (3 E, 3 C) Teachers: 20 E Students: 457 (257 E, 200 C)	9	High schools in Baltimore. 89% AA, 9% W, >90% FRL.	CTBS Terra Nova	+0.32	+0.32
Collaborative Strategic Reading (CSR) – Whole Class/School								
Denver Public Schools (2016)	Cluster randomized (teachers randomized)	1 year	Schools : 16 Students : 5660 (3101 E, 2559 C)	6, 7, 8	16 Denver middle schools, 62% H, 19% W, 11% AA, 30% ELL, 11% SPED, 76% FRL.	GMRT	+0.04	+0.03
						State Reading Test (TCAP)	+0.02	

Vaughn et al. (2011)	Cluster randomized (sections randomized within teachers)	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 (sections randomized within teachers)	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, 2%A, 2%AA, 6% LEP, 7% SPED	GMRT	+0.10	+0.10
						TOSREC	+0.11	

Table 3. Programs Incorporating Technology

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
					Computer-Assisted Instruction			
The Thinking Reader – Whole Class/School								
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 (> 33% FRL) from 16 districts in Connecticut, Massachusetts, and Rhode Island. 37% W, 28% H, 11% SPED, 10% ELL.	GMRT	+0.01	+0.01
						Comprehension	+0.03	
						Vocabulary	-0.04	
SuccessMaker – Whole Class/School								
Gatti (2011)	Cluster randomized	1 year	Classes: 22 (11 E, 11 C) Students: 453 (254 E, 199 C)	7	5 schools from 8 urban and suburban school districts in 4 states (Arizona, Kansas, Michigan, Missouri). 51% W, 53% FRL, 32% >=1 grade below grade level.	GRADE	+0.11	+0.11
						Comprehension	+0.10	
						Vocabulary	+0.12	
Achieve3000 – Whole Class/School								
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	
					Mixed-Model Approaches			
eMINTS-Whole Class/School								
Meyers, Molefe, Dhillon, & Zhu (2015)	Cluster randomized	3 years	Schools:59 (20 E, 20 E+, 19 C) Students: 3295 (1208 E, 1216 E+, 871 C)	6-8	59 rural middle schools across Missouri. 93%W, 7% non-white, 3% LEP, 60% FRL	MAP		-0.06
						eMINTS	-0.04	
						eMINTS + Intel	-0.08	

READ 180 - Targeted

Study	Design	Duration	Sample	Grade	Population	Measure	Effect	Overall
Schenck et al. (2011)	Student randomized	2 years	Students: 1295 (556 E, 739 C) (3 cohorts)	6-7 7-8	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)	6,7,8,9	Students from 5 Title I schools in Milwaukee who performed below proficient on standardized reading tests or were identified as performing at least two grade levels below expectations. 70% AA, 19% H, 36% SPED, 8% ELL, 88% FRL.	MAP Reading	+0.14*	+0.14*
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*	
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*
Meisch et al. (2011)	Cluster randomized	3 years	Schools: 19 (10 E, 9C) Students 1023 (552 E, 471 C) (4 cohorts)	6-8	19 Title I middle schools in Newark, NJ. Eligible students: below the state minimum level of proficiency in reading.	SAT 10 Comprehension	+0.06	+0.06
						Vocabulary	+0.05	

System 44- Targeted

Study	Design	Duration	Sample	Grade	Setting	Measure	Effect	Overall
Beam, Faddis, & Hahn (2011)	Student randomized	1 year	Students: 147 (75 E, 72 C)	6, 7, 8	7 schools (4 middle, 3 high) from one large suburban school district in southern California. 7% SPED, 63%W, 33% FRL.	TOSREC	-0.24*	-0.14
						CST	-0.04	
Beam & Faddis (2012)	Student randomized	1 year	Students: 145 (70 E, 75 C)	6, 7, 8	6 middle schools from one urban district in Michigan. 53% SPED, 78% AA, 96% FRL.	TOSREC	+0.20*	+0. 20*

Accelerated Reader - Targeted

Study	Design	Duration	Sample	Grade	Setting	Measure	Effect	Overall
Gorard, Siddiqui & See (2015)	Student randomized	22 weeks	Students: 349 (166 E, 183 C)	Year 7	Low-achieving students in four schools in England. 88% W, 4% ELL, 23% SPED, 35% FRL	NGRT	+0.24	+0.24

Prentice Hall Literature (2010) – Whole-Class/School

Study	Design	Duration	Sample	Grade	Setting	Measure	Effect	Overall
Eddy et al. (2010)	Cluster randomized	1 year	Teachers: 29 (16 E, 13 C) Students: 1518 (744 E, 774 C)	7,8,10	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 – Whole Class/School

Study	Design	Duration	Sample	Grade	Setting	Measure	Effect	Overall
Fogarty et al. (2014)	Cluster randomized	1 year	Classes: 61 (30 E, 31 C) Students: 859 (411 E, 448 C)	6, 7, 8	61 classes in 3 middle schools in the Southwest. 43% H, 35% W, 22% AA, 9% EL, 6% SPED, 67% FRL	GMRT	+0.12	+0.12
Fogarty et al. (2016)	Cluster randomized	1 year	Classes: 16 (9 E, 7 C) Students: 228 (112 E, 116 C)	6, 7, 8	16 classes in three middle schools in Texas. 30% AA, 27% W, 26% H, 3% A	STAAR	+0.09	+0.15
						GMRT	+0.12	
						GRADE	+0.11	
						GORT	+0.18	
						TOSREC	+0.24*	

Voyager Passport Reading Journeys - Targeted

Study	Design	Duration	Sample	Grade	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, 4% LEP, 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 855: (427 E, 428 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 – Whole Class

Study	Design	Duration	Sample	Grade	Population	Measure	Effect	Overall
Gatti (2016)	Student randomized	2 years	Students: 213 (114 E, 99 C)	7-8	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	

Texas Technology Immersion Pilot (eTxTip) – Whole Class/School

						TAKS		
Shapley, Sheehan, Maloney, & Caranikas-Walker (2009)	Cluster quasi-experiment	2-3 years	Schools: 42 (21 E, 21 C) Students: Cohort 1 (Gr. 6, 9): 3311 (1506 E, 1805 C) Cohort 2 (Gr. 6-8): 3268 (1571 E, 1697 C) Cohort 3 (Gr. 6-7): 3655 (1690 E, 1965 C)	6-9	42 middle schools across Texas. 70% H, 22% W, 7% AA, 15% LEP, 70% FRL	6-9	+0.08	+0.06
						6-8	+0.07	
						6-7	+0.02	

Table 4. Metacognitive Strategy Approaches

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
Strategic Instruction Model (SIM)								
SIM: Xtreme Reading - Targeted								
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	High School Students GRADE: 757 (355 E, 402 C) (4 cohorts)	7,8, 9,10	High School: 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*	+0.15*
						Comprehension	+0.15*	
						Vocabulary	+0.07	
			High School Students OAKS: 514 (260 E, 254 C) (4 cohorts)			OAKS (10th grade only)	+0.02	
			Middle School Students GRADE: 822 (401 E, 421 C) (4 cohorts)		Middle School: 6 Title I 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*	
			Middle School Students OAKS: 954 (472 E, 482 C)			OAKS	+0.12*	
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

						Outcomes		
SIM: Content Literacy Curriculum (CLC) – Whole Class/School								
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, 9-10	28 urban high schools in urban context from midwest states (Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio and Wisconsin). > 33% of students below proficient on state reading tests. > 25% FRL.	GRADE 1 year Comprehension +0.06 Vocabulary +0.09 2 years Comprehension +0.10 Vocabulary +0.10		+0.09
SIM: Learning Strategies Curriculum (LSC) - Targeted								
Cantrell et al. (2016)	Student randomized	1 year	Grade 6 Students: 1135, 605 E, 530 C (4 cohorts) / Grade 9 Students: 1128, 593 E, 535 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 Grade 6 +0.08 Grade 9 +0.12*		+0.10*
SIM: Fusion Reading - Targeted								
Schiller et al. (2012)	Student randomized	1 year	GRADE: Students: 581 (285 E, 296C) / MEAP: Students: 256 (118 E, 138 C)	6, 7, 8, 9, 10 / 6,7	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 Comprehension +0.08 Vocabulary 0.00 MEAP Reading +0.11		+0.07
SIM: Adolescent Literacy Model (ALM) – Whole Class/School								
Cantrell et al. (2011)	Cluster quasi-experimental	4 years	Middle schools: 18 (9 E, 9 C) High schools : 20 (10 E, 10 C) Students: not available	6-12	38 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

Reading Apprenticeship – Whole Class								
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	+0.03
						CST ELA	+0.10	
						Reading Comprehension	+0.13	
Reading Apprenticeship Academic Literacy (RAAL) - Targeted								
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*
						Comprehension	+0.12*	
						Vocabulary	0.00	
			State Test: Students: 1053 (2 cohorts)			State Tests ELA	+0.15*[s]	
Reading Apprenticeship Improving Secondary Education (RAISE) – Whole Class/School								
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)	9,10,11, 12	42 high schools in California and Pennsylvania. 49% AA, 33% H, 10% ELL, 40% FRL.	ETS Literacy Achievement Assessment	+0.14	+0.14
iRAISE – Whole Class/School								
Jaciw et al. (2016)	Cluster randomized	1 year	Schools: 26 Teachers: 69 (35 E, 34 C) Students: 1468 (751 E, 717 C)	9,10,11, 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 – Whole Class/School								
Simmons et al. (2014)	Cluster randomized	1 semester	Classes: 65 (36 E, 29 C) Students: 786 (413 E, 373 C)	7,8,9, 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 5. Mixed-Model Professional Development

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
Teacher Effectiveness Enhancement Programme (TEEP) – Whole Class/School								
Institute for Effective Education (2006)	Cluster Randomized	1.5-2 years	Schools: 45 (23 E, 22 C) Students: 10,385 (5327 E, 8058 C)	Year 9	45 secondary schools across England. 30% FRL, 16% ELL.	CEM Insight-English	-0.04	-0.04
Word Generation - Whole Class/School								
Lawrence, Francis, Pare-Blagoev, & Snow (2016)	Cluster randomized	1 year	Schools: 44 (25E, 19C) Students: 8466 (4796E, 3670C)	6, 7, 8	Schools in 2 Northeast, 1 Western urban districts. 81% FRL.	Gates-MacGinitie		+0.05
						Reading Comprehension	+0.07	
						Vocabulary	0.00	
ANet – Whole Class/School								
West, Morton, & Herlihy (2016)	Cluster randomized	2 years	Schools: 89 (45E, 44C) 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
Chicago Striving Readers – Whole Class/School								
Simon et al. (2011)	Cluster randomized	1 to 3 years	Schools: 61 (30 E, 31 C) Students: 8127 (4074 E, 4053 C) (2 cohorts)	6-8	60 middle schools in Illinois. 55% AA, 40% H, 15% SPED. > 95% low income.	ISAT Reading	-0.01	-0.01
Project CRISS – Whole Class/School								
Kushman, Hanita, & Raphael (2011)	Cluster randomized	1 year	Schools: 49 (23E, 26C) Students: 4959 (2460 E, 2499 C)	9	Schools in rural and urban fringe Northwest	SDRT	+0.05	+0.05

REACH - Targeted								
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
			Moderate Risk Students: 401 (199 E, 202 C)			High risk	-0.19	
						Moderate risk	+0.06	
Reading Excellence: Word Attack and Rate Development Strategies (REWARDS) - Targeted								
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% Asian, 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) - Targeted								
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
Reading Intervention through Strategy Enhancement (RISE) - Targeted								
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*
			Moderate Risk Students: 406 (204 E, 202 C)			High risk	-0.06	
						Moderate risk	+0.27*	
Building Assets Reducing Risk (BARR) – Whole Class/School								
Corsello & Sharma (2015)	Student randomized	1 year	Schools: 1 Students: 495 (261 E, 234 C)	9	School in Southern California 52% W, 37% H, 11% AA, 17% ELL, 68% FRL	NWEA	+0.14*	+0.14*

Read to Achieve - Targeted

Deussen et al. (2012)	Student randomized	1 year	Read to Achieve + PhonicsBlitz	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		+0.10
			GMRT Students: 63 (32 E, 31 C)			GMRT	+0.13	
			MSP Students: 76 (37 E, 39 C)			MSP	+0.11	
			Read to Achieve			Read to Achieve		
			GMRT Students: 295 (144 E, 151 C)			GMRT	+0.02	
			MSP Students: 325 (192 E, 191 C)			MSP	+0.16	

Expert 21 – Whole Class/School

Sivin-Kachala & Bialo (2012)	Student randomized	1 year	Teachers: 6 (3 E, 3 C) Students: 276 (137 E, 139 C)	6, 7, 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	

Strategies for Literacy Independence Across the Curriculum (SLIC) - Targeted

Hofstetter et al. (2011)	Student randomized	Up to 3 years	CST 2 years: Students: 1574 (782 E, 792 C)	7-10	Students from 8 middle and high schools in San Diego who were reading at least two years below grade level. 39% H, 16% AA, 22% ELL, 64% FRL	CST ELA		+0.04
			CST 3 years: Students: 606 (305 E, 301 C)			2 years	+0.01	
						3 years	-0.03	
			DRP 2 years: Students: 1178 (587 E, 591 C)			DRP		
						2 years	+0.08	

			DRP 3 years: Students: 324 (168 E, 156 C)			3 years	+0.03	

Strategies for Literacy Independence Across the Curriculum (SLIC) – Whole Class/School

Hofstetter et al. (2011)	Cluster quasi-experimental	1 year	CST: Schools:16 (8 E, 8 C) Students: 11738 (4915 E, 6823 C)	6,7,8, 9,10, 11,12	16 middle and high schools in San Diego. 39% H, 18% W, 17% Filipino.	CST ELA	+0.05	+0.04
			DRP: Students: 5093 (2234 E, 2859 C)			DRP	+0.02	

Every Classroom, Every Day (ECED) – Whole Class/School

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

Expository Reading and Writing Course (ERWC) – Whole Class/School

Fong, Finkelstein, Jaeger, Diaz, & Broek (2015)	Within-School quasi-experimental	1 year	Students: 6618 (3309 E, 3309 C)	12	Twelfth graders in 24 schools across California matched. 45% H, 27% A, 24% W, 4% AA	EPT	+0.13*	+0.13*

Schoolwide Enrichment Model-Reading (SEM-R) – Whole Class/School

Little et al. (2014)	Cluster randomized	1 year	Teachers: 47 (27E, 20C) Students: GMRT: 2028 (1198E, 830C) ORF: 2011 (1179E, 832C)	6, 7, 8	4 middle schools with 48% to 79% of students eligible for free or reduced price lunch	GMRT Comprehension	+0.10*	+0.10*
						Oral Reading Fluency	+0.09*	

Table 6: Programs for English Language Learners

Study	Design	Duration	N	Grades	Sample characteristics	Posttest	Effect sizes	Overall effect size
Pathway – Whole Class/School								
Kim et al. (2011); Olson et al. (2012)	Cluster randomized	1 year	*1st cohort* Teachers: 95 (48 E, 47 C) Students: 2726 (1421 E, 1305 C)	6, 7, 8, 9, 10, 11	15 schools (9 middle, 6 high) from a large school district in California. Eligible students: mainstreamed Latino ELLs able to participate in regular English classes. 95% H, 88% ELL, 79% FRL.	CST ELA	+0.07*	+0.07*
			2nd cohort Teachers: 66 (31 E, 35 C) Students: 1733 (779 E, 954 C)			CST ELA	+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.	CAHEE	+0.19	+0.19
Academic Language Instruction for All Students (ALIAS) – Whole Class/School								
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. H 49%, 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	+0.04	+0.04
						Comp.	-0.04	
						Vocabulary	+0.17*	