



# WWC Intervention Report

A summary of findings from a systematic review of the evidence



## Adolescent Literacy

Updated November 2016

# READ 180®

### Program Description<sup>1</sup>

*READ 180*® is a reading program designed for struggling readers who are reading 2 or more years below grade level. It combines online and direct instruction, student assessment, and teacher professional development. *READ 180*® is delivered in 90-minute sessions that include whole-group instruction, three small-group rotations, and whole-class wrap-up. Small-group rotations include individualized instruction using an adaptive computer application, small-group instruction, and independent reading. *READ 180*® is designed for students in elementary through high school. This review of *READ 180*® focuses on students in grades 4–12.

### Research<sup>2</sup>

The What Works Clearinghouse (WWC) identified nine studies of *READ 180*® that both fall within the scope of the Adolescent Literacy topic area and meet WWC group design standards. Three studies meet WWC group design standards without reservations, and six studies meet WWC group design standards with reservations. Together, these studies included 8,755 adolescent readers in more than 66 schools in 15 school districts and 10 states.

The WWC considers the extent of evidence for *READ 180*® on the reading achievement of adolescent readers to be medium to large for four outcomes—comprehension, general literacy achievement, reading fluency, and alphabetics. (See the Effectiveness Summary on p. 7 for more details of effectiveness by domain.)

### Effectiveness

*READ 180*® was found to have positive effects on comprehension and general literacy achievement, potentially positive effects on reading fluency, and no discernible effects on alphabetics for adolescent readers.

**Table 1. Summary of findings<sup>3</sup>**

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
<b>Comprehension</b>	Positive effects	+6	–4 to +16	6	3,882	Medium to large
<b>General literacy achievement</b>	Positive effects	+4	0 to +7	6	6,235	Medium to large
<b>Reading fluency</b>	Potentially positive effects	+4	+4 to +4	2	561	Medium to large
<b>Alphabetics</b>	No discernible effects	0	–1 to +2	2	746	Medium to large

### Report Contents

Overview	p. 1
Program Information	p. 2
Research Summary	p. 4
Effectiveness Summary	p. 7
References	p. 11
Research Details for Each Study	p. 22
Outcome Measures for Each Domain	p. 39
Findings Included in the Rating for Each Outcome Domain	p. 41
Supplemental Findings for Each Outcome Domain	p. 47
Endnotes	p. 54
Rating Criteria	p. 56
Glossary of Terms	p. 57

This intervention report presents findings from a systematic review of *READ 180*® conducted using the WWC Procedures and Standards Handbook, version 3.0, and the Adolescent Literacy review protocol, version 3.0.

### Program Information

#### Background

*READ 180*® is currently distributed by Houghton Mifflin Harcourt. It was developed by Dr. Ted Hasselbring and a team from the Cognition and Technology Group at Vanderbilt University, the Orange County Literacy Project in Florida, and the development staff at Scholastic, Inc. in 1985. The first version of *READ 180*® was published in 1998. In 2006, Scholastic, Inc. released *READ 180*® *Enterprise* which added features to the program such as the *rBook*® (an interactive workbook that introduces reading skills and strategies), additional features for English learners, and a Scholastic Achievement Manager (SAM), which is an online learning management system designed to implement applications and collect data on a district-wide basis (currently known as the Student Achievement Manager). In 2011, Scholastic, Inc. released *READ 180*® *Next Generation*, which includes a suite of new technology, data analyses, content, and resources designed to maximize student engagement and teacher effectiveness. In 2015, Houghton Mifflin Harcourt acquired Scholastic's educational technology and services business, which included *READ 180*®. In 2016, Houghton Mifflin Harcourt released *READ 180*® *Universal*, which is based on research on the cognitive functioning of struggling readers. *READ 180*® *Universal* includes new adaptive learning software, new content, and a new learning management system called Teacher Central. The WWC refers to all of these packages as *READ 180*® in this intervention report, unless the version was noted in the original study.<sup>4</sup>

Address: Houghton Mifflin Harcourt, 524 Broadway, Ste. 920, New York, NY 10012. Attn: Francie Alexander, Chief Academic Officer, HMH Intervention Solutions Group. Email: [Francie.Alexander@hmhco.com](mailto:Francie.Alexander@hmhco.com). Web: <http://www.hmhco.com/products/read-180/>. Phone: 212-965-7233.

#### Program details

The *READ 180*® instructional model is 90 minutes long and is composed of three parts: whole-group direct instruction, small-group rotations, and whole-group wrap-up. The instruction begins with 20 minutes of whole-group direct instruction, in which the teacher provides instruction in reading, writing, vocabulary, and grammar to the entire class. This is followed by 20-minute rotations of smaller groups of students through three activities:

- Small-group direct instruction, in which the teacher works closely with individual students using interactive work texts (called *ReaL Books*), paperback books, and eBooks. Instruction focuses on vocabulary, writing, and fluency.
- Students' independent use of a computerized *READ 180*® Student Application that includes six components (called "zones"): (1) Explore, which includes anchor videos with vocabulary activities, (2) Reading, which involves reading of individualized texts based on a student's instructional reading level, (3) Language, which includes vocabulary practice, (4) Fluency, which includes practice in spelling and reading, (5) Writing, which includes spelling and writing activities, and (6) Success, which includes fluency and comprehension activities.
- Modeled and independent reading, in which students read paperbacks or eBooks, or listen to audiobooks.

The instruction ends with a 10-minute wrap-up discussion with the whole group. The goal of the *READ 180*® software is to continually adjust the level of instruction based on student performance.

Reports and periodic updates on student progress are intended to alert teachers to students' needs and direct them to resources for individualizing instruction. *READ 180*® includes professional development for teachers and leaders to evaluate and improve instruction to support students who are reading below proficiency and help them gain independence with grade-level text.

### Cost

As of September 2016, the initial start-up cost of a *READ 180*<sup>®</sup> *Universal* package for 60 students was approximately \$43,000. Houghton Mifflin Harcourt provides 2.5 days of in-person professional development with the purchase of the program. A *READ 180*<sup>®</sup> *Universal* upgrade kit for 30 students costs \$8,800 and includes teacher materials, 30 *ReaL Books*, six boxes of Independent Reading Library books, and access to the new online student application. An upgrade kit with 60 student licenses costs \$12,000.

## Research Summary

The WWC identified 39 eligible studies that investigated the effects of *READ 180*<sup>®</sup> on reading achievement for adolescent readers. An additional 117 studies were identified but do not meet WWC eligibility criteria for review in this topic area. Citations for all 156 studies are in the References section, which begins on p. 11.

The WWC reviewed 39 eligible studies against group design standards. Three studies (Fitzgerald & Hartry, 2008; Kim, Samson, Fitzgerald, & Hartry, 2010; Swanlund, Dahlke, Tucker, Kleidon, Kregor, Davidson-Gibbs, & Halberg, 2012) are randomized controlled trials that meet WWC group design standards without reservations, and six studies (Interactive Inc., 2002; Meisch et al., 2011; Sprague, Zaller, Kite, & Hussar, 2012; White, Haslam, & Hewes, 2006; White, Williams, & Haslam, 2005; Yurchak, 2013) are randomized controlled trials or quasi-experimental designs that meet WWC group design standards with reservations. Those nine studies are summarized in this report. The remaining 30 studies do not meet WWC group design standards.

**Table 2. Scope of reviewed research**

<b>Grades</b>	4–10
<b>Delivery method</b>	Whole class
<b>Program type</b>	Curriculum

### Summary of studies meeting WWC group design standards without reservations

Fitzgerald and Hartry (2008) conducted a randomized controlled trial that examined the effects of *READ 180*<sup>®</sup> *Enterprise Edition* on students in grades 4–6 in four elementary schools in Brockton, Massachusetts. Students were eligible for the study if they scored below proficient on the Massachusetts Comprehensive Assessment System (MCAS) English Language Arts (ELA) subtest; however, a small percentage of students who scored above proficiency level were also recruited to reach sample size targets. Students were randomly assigned either to receive *READ 180*<sup>®</sup> during an afterschool program or to participate in a standard afterschool program. The study was conducted over two academic years and included two cohorts of study participants. In the first year of the study (2006–07), the *READ 180*<sup>®</sup> afterschool program was provided to Cohort 1 students, and in the second year (2007–08), it was provided to Cohort 2 students and approximately a third of students in Cohort 1 who returned for a second year. The afterschool program included two full *READ 180*<sup>®</sup> lessons per week over approximately 23 weeks in each study year. For the first study year, the program was modified from its customary 90-minute session length to fit the 60-minute afterschool program’s schedule and was implemented 4 days per week, but was extended to the full 90 minutes in the second year. During the first study year, the afterschool program took place 4 days per week in all schools. During the second study year, it took place 2 days per week in three out of four schools and 4 days per week in the remaining school. The WWC based its effectiveness rating on findings from the first year for each cohort, which were measured in the spring of each school year, following completion of the program. The WWC based its effectiveness rating on 151 students in the *READ 180*<sup>®</sup> group and 146 students in the comparison group in Cohort 1, and 93 students in the intervention group and 94 students in the comparison group in Cohort 2.

Kim et al. (2010) conducted a randomized controlled trial in three elementary schools in Brockton, Massachusetts. This study was Phase 1 of a two-phase study; the study described above in Fitzgerald and Hartry (2008) was Phase 2. Because the three elementary schools that participated in Phase 1 were different from the four schools that participated in Phase 2, and because results were reported separately for both phases, the WWC considers these to be different studies. Students in grades 4–6 were eligible for the study if they scored below proficient on the MCAS ELA subtest. During the 2005–06 school year, students were randomly assigned either to receive the *READ 180*<sup>®</sup> program during the second half of a 2-hour afterschool session or to participate in the standard 2-hour afterschool program. Students attended these afterschool programs 4 days per week over a 23-week period, from October 2005 to May 2006. The WWC based its effectiveness rating on findings from 133 students in the *READ 180*<sup>®</sup> group and 131 students in the comparison group.

Swanlund et al. (2012) conducted a randomized controlled trial that examined the effects of *READ 180*® on students in five schools in Milwaukee, Wisconsin. During the 2010–11 school year, students in grades 6–10 were randomly assigned either to receive the *READ 180*® program as a 90-minute daily supplement to their regular reading instruction or to a comparison group which included regular ELA instruction plus an elective class or study hall. The WWC based its effectiveness rating on outcomes measured at the end of the school year (June 2011). These outcomes were gathered from 335 students in the *READ 180*® group and 284 students in the comparison group.

### Summary of studies meeting WWC group design standards with reservations

Interactive, Inc. (2002) conducted a randomized controlled trial that examined the effects of *READ 180*® on students in Boston (grade 6), Dallas (grade 8), Houston (grades 7–8), and Columbus, Ohio (grades 6–7).<sup>5</sup> The study was originally designed as a randomized controlled trial, but the authors note that the randomization was not implemented as planned. However, the authors demonstrated equivalence on the analytic sample and, therefore, the study meets WWC group design standards with reservations. Students were assigned within each school to either a *READ 180*® group or a business-as-usual comparison group in the beginning of 2000–01 school year. During the school year, the *READ 180*® program was generally delivered in daily 90-minute blocks; however, there was some variation in implementation (e.g., one school in Boston set aside 45 minutes of *READ 180*® instruction twice a week to focus on writing skills). Due to differences in assessments used, the WWC based its effectiveness rating on two separate samples: (1) a combined sample of students from Boston, Houston, and Dallas and (2) students from Columbus. Although the Boston and Houston samples individually did not meet WWC standards because baseline equivalence was not demonstrated, the combined Boston, Dallas, and Houston sample met WWC group design standards with reservations. The effectiveness rating on the combined sample of Boston, Houston, and Dallas was based on 387 students in the *READ 180*® group and 323 students in the comparison group. The effectiveness rating for the Columbus sample was based on 119 students in the *READ 180*® group and 52 students in the comparison group. All outcomes were measured in the spring of 2001.

Meisch et al. (2011) conducted a cluster randomized controlled trial that examined the effects of *READ 180*® on students in 19 middle schools in Newark, New Jersey. In May 2006, 20 schools that were Title I eligible, categorized as “in need of improvement” under the No Child Left Behind Act, and had at least 25 eligible students were randomly assigned either to deliver *READ 180*® or to serve as a comparison group. Students in grades 6–8 were eligible based on their score on the reading subtest of the New Jersey Assessment of Skills and Knowledge. *READ 180*® instruction was provided 90 minutes per day for 1–3 years. Students in comparison schools received the regular language arts curriculum. After randomization took place, two schools in the comparison group merged, which left 10 schools in the intervention group and nine in the comparison group. The integrity of the random assignment was jeopardized because students who entered schools after random assignment was conducted were included in the analytic sample. Because the authors discuss the effects of the intervention on students (not on schools) and the study demonstrated equivalence on the analytic sample at baseline, the study *meets WWC group design standards with reservations*. The WWC based its effectiveness rating on outcomes from students who had 3 years of exposure to the *READ 180*® intervention, which included 552 students in the *READ 180*® group and 471 students in the comparison group.

Sprague et al. (2012) conducted a randomized controlled trial that examined the effects of *READ 180*® on students in five high schools located in two school districts in western Massachusetts. Beginning in the 2006–07 school year, students that were at least 2—but less than 4—years behind grade level were randomly assigned either to receive *READ 180*® as a 90-minute daily supplement to the standard ninth-grade ELA course or to serve in a comparison group. The comparison group received standard ninth-grade ELA instruction and had access to supplemental services available to all students. Across all five annual cohorts (2006–07 school year through the 2010–11 school year), a total of 548 students were randomly assigned to the *READ 180*® group, and 566 students were randomly assigned to the comparison group. The WWC based its effectiveness rating on outcomes measured in the spring of

each school year, following the completion of the 125–145 day *READ 180*<sup>®</sup> program, for 231 students in the *READ 180*<sup>®</sup> group and 225 students in the comparison group. Because this study had high attrition by WWC standards, but demonstrated baseline equivalence on the analytic sample, the study meets WWC group design standards with reservations.

White et al. (2006) conducted a quasi-experimental study that examined the effects of *READ 180*<sup>®</sup> on students in the Phoenix Union High School District.<sup>6</sup> Students in grades 9 and 10 were eligible to receive *READ 180*<sup>®</sup> if they were reading one or more grades below their assigned grade level. Students in the *READ 180*<sup>®</sup> group were matched to nonparticipants based on prior reading proficiency assessments, English learner (EL) status, special education eligibility, gender, and ethnicity. Four cohorts of students were studied. Two cohorts did not meet Adolescent Literacy protocol or WWC eligibility requirements. Cohort 1 did not meet eligibility requirements for the Adolescent Literacy review protocol, since more than half of participating students (53%) were eligible for EL services. Cohort 4 did not include a comparison group and was thus ineligible for WWC review. The WWC based its effectiveness rating on Cohort 2 and Cohort 3 findings, which were measured at the end of each school year. Cohort 2 included 815 *READ 180*<sup>®</sup> students and 815 matched comparison students who were in ninth grade in the 2004–05 school year. Cohort 3 included 1,029 students in the *READ 180*<sup>®</sup> group and 1,029 students in the comparison group who were ninth graders in the 2005–06 school year.

White et al. (2005) conducted a quasi-experimental study that examined the effects of *READ 180*<sup>®</sup> on students in grades 4–8 at 16 schools in New York City.<sup>7</sup> Students receiving *READ 180*<sup>®</sup> instruction in the 2001–02 school year were compared to students in the same schools who had never participated in *READ 180*<sup>®</sup>. The combined analysis sample and the individual subsamples by grade did not meet WWC baseline equivalence standards. However, subgroup analyses were conducted by grade level and proficiency level (level 1=Below Basic; level 2=Basic; level 3=Proficient; and level 4=Advanced). Three subgroup analyses had no baseline differences between the intervention and comparison groups and met WWC group design standards with reservations: (1) grade 6, proficiency level 2 at baseline; (2) grade 8, proficiency level 2 at baseline; and (3) grade 8, proficiency level 3 at baseline. The WWC based its effectiveness rating on findings from the three referenced subgroup analyses. The grade 6, proficiency level 2 subsample included 64 students in the intervention group and 407 students in the comparison group. The grade 8, proficiency level 2 subsample included 47 students in the intervention group and 378 students in the comparison group. The grade 8, proficiency level 3 subsample included 10 students in the intervention group and 191 students in the comparison group.

Yurchak (2013) conducted a quasi-experimental study that examined the effects of *READ 180*<sup>®</sup> on students in a single urban high school in northern New Jersey. Students with 1 year of exposure to *READ 180*<sup>®</sup> in ninth grade were matched with students in regular ninth-grade English classes based on eighth-grade pretest scores from the Language Arts Literacy portion of the state assessment. This design included three consecutive cohorts from the 2007–08, 2008–09, and 2009–10 school years. Students in 15 *READ 180*<sup>®</sup> sections received 80 minutes of daily instruction that closely mirrored the standard 90-minute *READ 180*<sup>®</sup> model. Students in the comparison group received the standard ninth-grade English course, which was 40 minutes long. The WWC based its effectiveness rating on the findings from the three cohorts combined. The analytic sample included 67 students in the *READ 180*<sup>®</sup> group and 67 students in the comparison group.

### Effectiveness Summary

The WWC review of *READ 180*<sup>®</sup> for the Adolescent Literacy topic area includes outcomes in four domains: comprehension, general literacy achievement, reading fluency, and alphabetics. The nine studies of *READ 180*<sup>®</sup> that meet WWC group design standards reported findings in all four domains. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *READ 180*<sup>®</sup> on adolescent readers. Additional comparisons are presented as supplemental findings in Appendix D. These supplemental findings do not factor into the intervention's rating of effectiveness. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 56.

#### Summary of effectiveness for the comprehension domain

Six studies that meet WWC group design standards with or without reservations reported findings in the comprehension domain.

Fitzgerald and Hartry (2008) reported findings from the Stanford Achievement Test, Tenth Edition (Stanford 10) Vocabulary and Reading Comprehension subtests. For Cohort 1, the authors reported statistically significant positive differences between the *READ 180*<sup>®</sup> *Enterprise Edition* and comparison groups on both outcomes, and the result for the Reading Comprehension subtest was large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25). The WWC confirmed that the substantively important result for the Reading Comprehension subtest was statistically significant. However, when the result for the Vocabulary subtest was adjusted for multiple comparisons, the result was no longer statistically significant. The authors also reported, and the WWC confirmed, no statistically significant differences between the intervention and comparison groups for Cohort 2. The effect sizes for the Cohort 2 findings were not large enough to be considered substantively important. The WWC characterizes this study finding as a statistically significant positive effect.

Interactive, Inc. (2002) reported findings from the Stanford 9 Total Reading assessment for both the combined Boston, Houston, and Dallas sample (grades 6–8) and the Columbus sample (grades 6–7). The authors reported, and the WWC confirmed, positive and statistically significant differences between the *READ 180*<sup>®</sup> group and the comparison group. The average effect size across samples is large enough to be considered substantively important. The WWC characterizes this study finding as a statistically significant positive effect.

Kim et al. (2010) reported findings on the Group Reading Assessment and Diagnostic Evaluation (GRADE) total score. The authors reported, and the WWC confirmed, no statistically significant or substantively important findings between the *READ 180*<sup>®</sup> group and the comparison group. The WWC characterizes this study finding as an indeterminate effect.

Meisch et al. (2011) reported findings on the Stanford 10 Vocabulary and Reading Comprehension subtests. The authors reported, and the WWC confirmed, no statistically significant differences between students with 3 years of exposure to *READ 180*<sup>®</sup> and the comparison group, and the average effect size across these findings was not substantively important. The WWC characterizes this study finding as an indeterminate effect.

White et al. (2005) reported findings for three eligible subgroups of students (one in grade 6 and two in grade 8) on the CTB/McGraw Hill Reading Assessment (grade 6) and the New York State end-of-year test in ELA (grade 8). The authors did not report the statistical significance of findings, but the WWC found that none of the findings were statistically significant after correcting for multiple comparisons. The average effect size for students in the *READ 180*<sup>®</sup> group was positive and substantively important. The WWC characterizes these study findings as having a substantively important positive effect.

Yurchak (2013) reported findings on the New Jersey High School Proficiency Assessment (HSPA) Analyzing Text cluster score and the HSPA Reading cluster score. The author did not report the statistical significance of these

findings, but the WWC-computed calculations indicated that findings were not statistically significant or substantively important between students in the *READ 180*<sup>®</sup> group and students in the comparison group. The WWC characterizes this study finding as an indeterminate effect.

Thus, for the comprehension domain, one study that meets WWC group design standards without reservations showed a statistically significant positive effect, one study that meets WWC group design standards with reservations showed a statistically significant positive effect, one study that meets WWC group design standards with reservations showed a substantively important positive effect, and three studies that meet WWC group design standards with or without reservations showed an indeterminate effect. This results in a rating of positive effects, with a medium to large extent of evidence.

**Table 3. Rating of effectiveness and extent of evidence for the comprehension domain**

Rating of effectiveness	Criteria met
<b>Positive effects</b> <i>Strong evidence of a positive effect with no overriding contrary evidence.</i>	In the six studies that reported findings, the estimated impact of the intervention on outcomes in the <i>comprehension</i> domain was positive and statistically significant for two studies, one of which meets WWC group design standards without reservations, positive and substantively important for one study, and indeterminate for three studies.
Extent of evidence	Criteria met
<b>Medium to large</b>	Six studies that included 3,882 students in 61 schools reported evidence of effectiveness in the <i>comprehension</i> domain.

### Summary of effectiveness for the general literacy achievement domain

Six studies that meet WWC group design standards with or without reservations reported findings in the general literacy achievement domain.

Fitzgerald and Hartry (2008) reported findings on the Stanford 10 Total Reading Score for Cohort 2. The authors reported, and the WWC confirmed, no statistically significant or substantively important differences between students in the *READ 180*<sup>®</sup> group and students in the comparison group. The WWC characterizes this study finding as an indeterminate effect.

Kim et al. (2010) reported findings on the MCAS ELA assessment. The authors reported, and the WWC confirmed, no statistically significant or substantively important differences between students in the *READ 180*<sup>®</sup> group and students in the comparison group. The WWC characterizes this study finding as having an indeterminate effect.

Meisch et al. (2011) reported findings on the Stanford 10 Language Arts subtest. The authors reported, and the WWC confirmed, no statistically significant or substantively important differences between students with 3 years of exposure to *READ 180*<sup>®</sup> and students in the comparison group. The WWC characterizes this study finding as having an indeterminate effect.

Sprague et al. (2012) reported findings on the Stanford Diagnostic Reading Test (SDRT-4). The authors reported, and the WWC confirmed, that differences in test scores between students in Cohorts 1–5 of the *READ 180*<sup>®</sup> group and students in the comparison group were positive and statistically significant, but not substantively important. The WWC characterizes this study finding as having a statistically significant positive effect.

Swanlund et al. (2012) reported findings on the Measures of Academic Progress (MAP) outcome. The authors reported, and the WWC confirmed, that differences in MAP scores between students in the *READ 180*<sup>®</sup> group and students in the comparison group were positive and statistically significant, but not substantively important. The WWC characterizes this study finding as having a statistically significant positive effect.



White et al. (2006) reported findings on the TerraNova reading test for two cohorts of students. The authors reported, and the WWC confirmed, that differences between students in the *READ 180*<sup>®</sup> group and students in the comparison group were positive and statistically significant, but not substantively important. The WWC characterizes these study findings as having a statistically significant positive effect.

Thus, for the general literacy achievement domain, one study that meets WWC group design standards without reservations showed statistically significant positive effects, two studies that meet WWC group design standards with reservations showed statistically significant positive effects, and three studies showed indeterminate effects. This results in a rating of positive effects, with a medium to large extent of evidence.

**Table 4. Rating of effectiveness and extent of evidence for the general literacy achievement domain**

Rating of effectiveness	Criteria met
<b>Positive effects</b> <i>Strong evidence of a positive effect with no overriding contrary evidence.</i>	In the six studies that reported findings, the estimated impact of the intervention on outcomes in the <i>general literacy achievement</i> domain was positive and statistically significant for three studies, one of which meets WWC group design standards without reservations, and no studies showed statistically significant or substantively important negative effects.
Extent of evidence	Criteria met
<b>Medium to large</b>	Six studies that included 6,235 students in at least 37 schools reported evidence of effectiveness in the <i>general literacy achievement</i> domain.

**Summary of effectiveness for the reading fluency domain**

Two studies that meet WWC group design standards without reservations reported findings in the reading fluency domain.

Fitzgerald and Hartry (2008) reported findings on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency assessment from Cohort 1. The authors reported, and the WWC confirmed, no statistically significant or substantively important differences between students in the *READ 180*<sup>®</sup> group and students in the comparison group. The WWC characterizes this study finding as having an indeterminate effect.

Kim et al. (2010) reported findings on the DIBELS Oral Reading Fluency assessment. The authors reported, and the WWC confirmed, statistically significant differences between students in the *READ 180*<sup>®</sup> group and students in the comparison group. The WWC characterizes this study finding as a statistically significant positive effect.

Thus, for the reading fluency domain, in the two studies that meet WWC group design standards without reservations, one study showed a statistically significant positive effect and one study showed an indeterminate effect. This results in a rating of potentially positive effects, with a medium to large extent of evidence.

**Table 5. Rating of effectiveness and extent of evidence for the reading fluency domain**

Rating of effectiveness	Criteria met
<b>Potentially positive effects</b> <i>Evidence of a positive effect with no overriding contrary evidence.</i>	In the two studies that reported findings, the estimated impact of the intervention on outcomes in the <i>reading fluency</i> domain was positive and statistically significant for one study that meets WWC group design standards without reservations, and one study showed indeterminate effects.
Extent of evidence	Criteria met
<b>Medium to large</b>	Two studies that included 561 students in seven schools reported evidence of effectiveness in the <i>reading fluency</i> domain.

### Summary of effectiveness for the alphabetics domain

Two studies that meet WWC group design standards without reservations reported findings in the alphabetics domain.

Fitzgerald and Hartry (2008) reported findings on the Stanford 10 Spelling subtest separately for two cohorts of students. The authors reported, and the WWC confirmed, no statistically significant or substantively important differences between *READ 180*® students in Cohorts 1 and 2 and students in the comparison groups for each cohort. The WWC characterizes this study finding as having an indeterminate effect.

Kim et al. (2010) reported findings on the Test of Word Reading Efficiency. The authors reported, and the WWC confirmed, no statistically significant or substantively important differences between students in the *READ 180*® group and students in the comparison group. The WWC characterizes this study finding as an indeterminate effect.

Thus, for the alphabetics domain, two studies that meet WWC group design standards without reservations reported indeterminate effects. This results in a rating of no discernible effects, with a medium to large extent of evidence.

**Table 6. Rating of effectiveness and extent of evidence for the alphabetics domain**

Rating of effectiveness	Criteria met
<b>No discernible effects</b> <i>No affirmative evidence of effects.</i>	In the two studies that reported findings, the estimated impact of the intervention on outcomes in the <i>alphabetics</i> domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	Criteria met
<b>Medium to large</b>	Two studies that included 746 students in seven schools reported evidence of effectiveness in the <i>alphabetics</i> domain.

### References

#### Studies that meet WWC group design standards without reservations

Fitzgerald, R., & Hartry, A. (2008). *What works in afterschool programs: The impact of a reading intervention on student achievement in the Brockton Public Schools (phase II)*. Berkeley, CA: MPR Associates, Inc. and the National Partnership for Quality Afterschool Learning at SEDL.

**Additional sources:**

Kim, J. S., Capotosto, L., Hartry, A., & Fitzgerald, R. (2011). Can a mixed-method literacy intervention improve the reading achievement of low-performing elementary school students in an after-school program? Results from a randomized controlled trial of READ 180 Enterprise. *Educational Evaluation and Policy Analysis*, 33(2), 183–201.

Vaden-Kiernan, M., Hughes Jones, D., & Rudo, Z. (2008). *The National Partnership for Quality Afterschool Learning randomized controlled trial studies of promising afterschool programs: Summary of findings*. Austin, TX: SEDL Afterschool Research Consortium. <http://files.eric.ed.gov/fulltext/ED513822.pdf>.

Kim, J. S., Samson, J. F., Fitzgerald, R., & Hartry, A. (2010). A randomized experiment of a mixed-methods literacy intervention for struggling readers in grades 4–6: Effects on word reading efficiency, reading comprehension and vocabulary, and oral reading fluency. *Reading and Writing: An Interdisciplinary Journal*, 23(1), 1109–1129.

**Additional source:**

Hartry, A., Fitzgerald, R., & Porter, K. (2008). Implementing a structured reading program in an afterschool setting: problems and potential solutions. *Harvard Educational Review*, 78(1), 181–210.

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

**Additional source:**

Scholastic, Inc. (2013). *Urban students in Milwaukee public schools demonstrate improved reading achievement after READ 180® instruction*. New York, NY: Scholastic, Inc.

#### Studies that meet WWC group design standards with reservations

Interactive, Inc. (2002). *An efficacy study of READ 180: A print and electronic adaptive intervention program, grades 4 and above*. Ashland, VA: Author.

Meisch, A., Hamilton, J., Chen, E., Quintanilla, P., Fong, P., Gray-Adams, K., ... Thornton, N. (2011). *Striving Readers study: Targeted and whole-school interventions—Year 5*. Rockville, MD: Westat.

**Additional sources:**

Carr, M., Daft, J., Ek, P., Falk-Smith, A., Fong, P., Hamilton, J., ... Thornton, N. (2010). *Striving Readers study: Targeted and whole-school interventions—Year 3*. Rockville, MD: Westat.

Hamilton, J., Gray-Adams, K., Meisch, A., & Petta, A. (2009). *Striving Readers study: Targeted and whole school interventions—Year 2*. Rockville, MD: Westat.

Hamilton, J., Meisch, A., Chen, E., Quintanilla, P., Fong, P., Gray-Adams, K., ... Thornton, N. (2011). *Striving Readers study: Targeted and whole school interventions—Year 4*. Rockville, MD: Westat.

Hamilton, J., Meisch, A., Gray-Adams, K., Petta, I., & Chen, E. (2010). *Evaluation of the Striving Readers program in Newark, New Jersey*. Rockville, MD: Westat.

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2012). *Springfield-Chicopee School Districts Striving Readers program final report Years 1–5: Evaluation of implementation and impact*. Providence, RI: The Education Alliance at Brown University.

**Additional sources:**

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2009). *Springfield-Chicopee School Districts Striving Readers (SR) program Year 2 report: Evaluation of implementation and impact*. Providence, RI: The Education Alliance at Brown University.

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2010). *Springfield-Chicopee School Districts Striving Readers (SR) program Year 3 report: Evaluation of implementation and impact*. Providence, RI: The Education Alliance at Brown University.

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2011). *Springfield-Chicopee School Districts Striving Readers (SR) program Year 4 report: Evaluation of implementation and impact*. Providence, RI: The Education Alliance at Brown University.

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

**Additional source:**

Scholastic Research and Results. (2008). *READ 180: Longitudinal evaluation of a ninth-grade reading intervention (2003–2006)*. New York, NY: Scholastic, Inc.

White, R., Williams, I., & Haslam, M. B. (2005). *Performance of District 23 students participating in Scholastic READ 180*. Washington, DC: Policy Studies Associates.

Yurchak, S. M. (2013). *The effect of READ 180 on the reading achievement of struggling readers in a large, public, urban high school in northern New Jersey* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3613825)

### Studies that do not meet WWC group design standards

Barbato, P. F. (2006). *A preliminary evaluation of the READ 180 program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3240964) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Bebon, C. D. (2007). *The impact of a reading program designed to increase comprehension and proficiency of middle school migrant students in a south Texas school district* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3274051) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Bonville, J. K. (2013). *The impact of two response-to-intervention tier 2 literacy programs on middle level achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3573287) The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Brennan, T., Leuer, M., Boyer, D., Dalessi, M., Newman, D., & Yepes-Baraya, M. (2006). *Rhetoric to reality: Addressing reading achievement in secondary education*. Palo Alto, CA: Empirical Education, Inc. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Caggiano, J. A. (2007). *Addressing the learning needs of struggling adolescent readers: The impact of a reading intervention program on students in a middle school setting* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3257319) The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Caldwell, R. L. (2010). *The study of predictive factors of reading in low-performing readers in an urban setting* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3370031) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Campbell, Y. C. (2006). *Effects of an integrated learning system on the reading achievement of middle school students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3228194) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

- Cannon, A. C. (2011). *A comparison between READ 180 students and non-READ 180 students reading and math scores by classroom structure* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3462041) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Davis, D. A. (2014). *Examining the effects of READ 180 with sixth grade students in a southwest United States school district based on a formative assessment—measures of academic progress—and its impact on leadership decisions* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3537764) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Denman, J. S. (2004). *Integrating technology into the reading curriculum acquisition, implementation, and evaluation of a reading program with a technology component (READ 180) for struggling readers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No.3133785) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Ferguson, J. M. (2005). *The implementation of technology in reading classrooms and the impact of technology integration and student perceptions on reading achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3170209) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Gentry, L. (2006). *An evaluation of READ 180 in an urban secondary school* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3234282) The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.
- Gray, D. (2014). *FCAT comparisons of students enrolled or not enrolled in Read 180* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3578019) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Haag Guyne, R. J. (2010). *The implementation of interventions and strategies for children who struggle with reading utilizing the Read 180 program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3389735) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 20(2), 123–144. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Hubbard, A. J. (2012). *An investigation of the effects of READ 180® on fourth grade reading achievement and how selected teachers implement the program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3478948) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Jacobs, R. S. (2012). *The impact of technology-based reading programs on lower achievers' state reading assessment* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3495151) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Lang, L. H., Torgesen, J. K., Petscher, Y., Vogel, W., Chanter, C., & Lefsky, E. (2008, March). *Exploring the relative effectiveness of reading interventions for high school students*. Presented at the annual research conference of the Society for Research on Educational Effectiveness, Crystal City, VA. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

- Lawson, S. (2012). *The Read 180 program: Analysis of program effect on the reading achievement, motivation, engagement, and self-efficacy of sixth grade middle school students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3481667) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Lombardi, D. (2015). *READ 180 evaluation: Balanced literacy in a low-income, underperforming urban high school* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3688004) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- McWhorter, H. (2009). *Facilitating high school student success through READ 180: Analysis of program impact using measures of academic progress (MAP)* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3379873) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Nave, J. (2007). *An assessment of READ 180 regarding its association with the academic achievement of at-risk students in Sevier County Schools* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3271894) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- O'Hare, J. M. (2012). *Instructional leadership: Using a research approach to evaluate the impact of a reading intervention program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3515553) The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.
- Papalewis, R. (2002). *A study of the intensive academic support program and READ 180 in the Los Angeles Unified School District*. New York, NY: Scholastic, Inc. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Additional source:**
- Papalewis, R. (2004). Struggling middle school readers: Successful, accelerating intervention. *Reading Improvement*, 41(1), 24–37.
- Parker, C. A., Holland, G., & Jones, D. (2013). The effectiveness of two reading intervention programs in a south Texas urban school district. *National Forum of Applied Educational Research Journal*, 26(3), 1–9. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.
- Additional source:**
- Parker, C. (2011). *The effectiveness of READ 180 and Voyager Journeys III in a south Texas urban school district* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3534090)
- Plony, D. A. (2014). *The effects of Read 180 on student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3563453) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Rakestraw, K. R. (2013). *The impact of the Read 180 program on response to intervention services* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3601111) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Scholastic Research and Validation. (2009). *Huntington Beach Unified School District research update*. New York, NY: Scholastic, Inc. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

- Thomas, D. M. (2005). *Examining the academic and motivational outcomes of students participating in the READ 180 program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3200340) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.
- Woods, D. E. (2007). *An investigation of the effects of a middle school reading intervention on school dropout rates* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3256138) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

### Studies that are ineligible for review using the Adolescent Literacy Evidence Review Protocol

- Admon, N. (2003). *READ 180 stage C: An evaluation within the federal Job Corps program*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Admon, N. (2003). *READ 180 stages A and B: Iredell-Statesville schools, North Carolina*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Admon, N. (2004). *A study of READ 180 in partnership with Guajome Park Academy/School for Integrated Academics and Technologies and Job Corps*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Admon, N. (2004). *READ 180 stage B: St. Paul School District, Minnesota*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Aguhob, M. (2007). *READ 180 in Miami-Dade County, Florida 2005–2006*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Aguhob, M. (2007). *READ 180 in Seminole County, Florida*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Bishop-Kallmeyer, N. (2009). *A critical examination of the READ 180 program among a sample of English as a second language students and special education students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3340152) The study is ineligible for review because it does not use a sample aligned with the protocol.
- Bishop-Temple, C. (2008). *The effects of interactive read-alouds on the reading achievement of middle grade reading students in a core remedial program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3287381) The study is ineligible for review because it is out of scope of the protocol.
- Blasewitz, M. R., & Taylor, R. T. (1999). Attacking literacy with technology in an urban setting. *Middle School Journal*, 30(3), 33–39. The study is ineligible for review because it is out of scope of the protocol.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125–230. <http://files.eric.ed.gov/fulltext/ED472569.pdf>. The study is ineligible for review because it does not use an eligible design.
- Brown, S. H. (2006). *The effectiveness of READ 180 intervention for struggling readers in grades 6–8* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3232344) The study is ineligible for review because it does not use an eligible design.
- Carter, L. H. (2015). *A secondary-level READ 180 comparison study in an urban/suburban high school* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3719728) The study is ineligible for review because it is out of scope of the protocol.
- Casey, B. L. (2011). *The influence of an interactive reading program on adolescent students in middle school* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3428715) The study is ineligible for review because it does not use an eligible design.
- Cheatham, P. A. (2012). *A correlational study between learning links and Read 180 using MetaMetrics Lexile measures in a county in southeastern United States* (Unpublished doctoral dissertation). Austin Peay State University, Clarksville, TN. The study is ineligible for review because it does not use an eligible design.

- Cheung, A. C. K., & Slavin, R. E. (2012). *The effectiveness of educational technology applications for enhancing reading achievement in K–12 classrooms: A meta-analysis. Revised*. Baltimore, MD: Center for Research and Reform in Education. <http://files.eric.ed.gov/fulltext/ED527573.pdf>. The study is ineligible for review because it does not use an eligible design.
- Cheung, A. C. K., & Slavin, R. E. (2013). Effects of educational technology applications on reading outcomes for struggling readers: A best-evidence synthesis. *Reading Research Quarterly*, 48(3), 277–299. The study is ineligible for review because it does not use an eligible design.
- Chmielewski, T. (2005). *Differentiating reading instruction in an alternative high school using READ 180* (Unpublished doctoral dissertation). Mount Mary College, Milwaukee, WI. The study is ineligible for review because it does not use an eligible design.
- Coffey, D. J. (2013). *Helping struggling adolescent readers: Is implementation of different components of Scholastic's READ 180 associated with differences in student achievement gains?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3612959) The study is ineligible for review because it does not use an eligible design.
- Daviess County Public Schools, Assessment, Research and Curriculum Department (2005). *READ 180 implementation year study*. Owensboro, KY: Author. The study is ineligible for review because it does not use an eligible design.
- DeSchryver, D. (2012). *Improving outcomes and reducing costs: Napa Valley Unified School District*. Washington, DC: Whiteboard Advisors. The study is ineligible for review because it does not use an eligible design.
- Dunn, C. A. (2014). *An investigation of the effects of computer-assisted reading instruction versus traditional reading instruction on selected high school freshmen* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3027824) The study is ineligible for review because it is out of scope of the protocol.
- Edington, J. L. (2008). *Measuring the reading self efficacy of Ohio's incarcerated youth: A comparison between READ 180 students and traditional English students* (Unpublished doctoral dissertation). The Ohio State University, Columbus. The study is ineligible for review because it is out of scope of the protocol.
- Eiberg, M., Knudsen, A. D., Sonne-Schmidt, C., & Klint Jørgensen, A. (2014). *Academic intervention programmes for improving school outcomes for educationally disadvantaged youth and low achieving students in grade 7 to 12: A systematic review*. Oslo, Norway: The Campbell Collaboration. The study is ineligible for review because it does not use an eligible design.
- Felty, R. L. (2008). *READ 180 implementation: Reading achievement and motivation to read within an alternative education middle school program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3300297) The study is ineligible for review because it does not use an eligible design.
- Flatland, T. A. (2010). *Computer-based reading programs and their effectiveness as reading instruction* (Unpublished doctoral dissertation). Bethel University, Saint Paul, MN. The study is ineligible for review because it does not use an eligible design.
- Frey, D. A. (2011). *Reading proficiency of middle and high school students: The effects of instructional duration, reading-related professional development, and administrative leadership in Read 180 classrooms* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3443002) The study is ineligible for review because it does not use an eligible design.
- Gagliardi, L. (2012). *Examining the Scholastic READ 180 program teachers' perceptions regarding local setting factors and role of the teacher impacting the program's implementation in seventh grade at three middle schools* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3468186) The study is ineligible for review because it does not use an eligible design.
- Giedd, D. (2008). *A study of the effectiveness of the READ 180 program measured by student test scores* (Unpublished doctoral dissertation). Northwest Missouri State University, Maryville. The study is ineligible for review because it does not use an eligible design.



- Gober, C. (2014). *READ 180: Is it an effective reading intervention for English language learners?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3668712) The study is ineligible for review because it does not use a sample aligned with the protocol.
- Goin, L., Hasselbring, T., & McAfee, I. (1998). *DoDEA/Scholastic READ 180 project: An evaluation of the READ 180 intervention program for struggling readers*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Hamilton, J., & Chen, E. (2012). *Secondary analysis of Striving Readers READ 180 data: Special education students*. Rockville, MD: Westat. The study is ineligible for review because it does not use a sample aligned with the protocol.
- Haslam, B. M., White, R., & Klinge, A. (2006). *Improving student literacy: READ 180 in the Austin Independent School District 2004–05*. Washington, DC: Policy Studies Associates. The study is ineligible for review because it does not use a sample aligned with the protocol.
- Hasselbring, T. S., Goin, L. I., Taylor, R., Bottge, B., & Daley, P. (1997). The computer doesn't embarrass me. *Educational Leadership*, 55(3), 30–33. The study is ineligible for review because it does not use an eligible design.
- Hewes, G., Palmer, N., Haslam, M. B., & Mielke, M. (2006). *Five years of READ 180 in Des Moines: Improving literacy among middle school and high school special education students*. Washington, DC: Policy Studies Associates. The study is ineligible for review because it does not use an eligible design.
- Hollingsworth, E. S. (2014). *An investigation into the effectiveness of Scholastic's READ 180 as an intervention program at the secondary level in a middle Tennessee school district* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3582788) The study is ineligible for review because it does not use an eligible design.
- Holyoke Public Schools. (2006). *READ 180 overview*. Holyoke, MA: Author. The study is ineligible for review because it does not use an eligible design.
- Johnson, L. R. (2012). *Evaluating the effectiveness of READ 180: A reading program for struggling middle school students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3486971) The study is ineligible for review because it does not use an eligible design.
- Johnson, M., McGinley, N., Rose, J., & Kokkins, A. (2006). *Implementation of READ 180 (reading intervention program) in CCSD schools, Brief No. 06-Brief 014*. Charleston, SC: Charleston County School District Department of Assessment and Accountability. The study is ineligible for review because it is out of scope of the protocol.
- Knutson, K. A. (2006). *Because you can't wait until spring: Using the SRI to improve reading performance*. New York, NY: Scholastic, Inc. The study is ineligible for review because it is out of scope of the protocol.
- Letson, K. S. (2013). *A study of the Gauge Middle School reading intervention program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3522208) The study is ineligible for review because it does not use an eligible design.
- Levin, H. M., Catlin, D., Elson, A., & Carnegie Corporation of New York. (2010). *Adolescent literacy programs: Costs of implementation. Final report from Carnegie Corporation of New York's Council on Advancing Adolescent Literacy*. New York, NY: Carnegie Corporation of New York. <http://files.eric.ed.gov/fulltext/ED535296.pdf>. The study is ineligible for review because it does not use an eligible design.
- Loadman, W., Sprague, K., Hamilton, J., Coffey, D., & Faddis, B. (2010). *Using randomized clinical trials to determine the impact of reading intervention on struggling adolescent readers: Reports of research from five nationally funded Striving Readers grants*. Evanston, IL: Society for Research on Educational Effectiveness. <http://files.eric.ed.gov/fulltext/ED513389.pdf>. The study is ineligible for review because it does not use an eligible design.

- Loadman, W. E., Moore, R. J., Ren, W., Zhu, J., Zhao, J., & Lomax, R. (2011). *Striving Readers year 5 project evaluation report: Ohio—An addendum to the year 4 report*. Columbus: The Ohio State University. The study is ineligible for review because it does not use a sample aligned with the protocol.
- Additional source:**
- Loadman, W. E., Moore, R. J., Ren, W., Zhu, J., Zhao, J., & Lomax, R. (2010). *Striving Readers year 3 project evaluation report: Ohio*. Columbus: The Ohio State University.
- Lupino, E. (2005). Taking place: The teacher in reading. *Journal of Adolescent & Adult Literacy*, 49(1), 4. The study is ineligible for review because it is out of scope of the protocol.
- Menendez, R. M. (2010). *Cultural-historical activity perspectives on the effects of participation in teacher-mediated, computer-mediated reading instruction* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3392603) The study is ineligible for review because it does not use a sample aligned with the protocol.
- Nelson, T. (2008). *Predictive factors in student gains in reading comprehension using a reading intervention program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3318825) The study is ineligible for review because it does not use an eligible design.
- Palmer, N. (2003). *READ 180 middle-school study: Des Moines, Iowa 2000–2002*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Palubinsky, R. C. S. (2008). *Factors impacting the effectiveness of Pennsylvania’s Educational Assistance Program (EAP) for eighth grade students as determined by increased reading proficiency on the Pennsylvania State System of Assessment (PSSA)* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3313319) The study is ineligible for review because it is out of scope of the protocol.
- Papalewis, R. (2003). *A study of READ 180 in middle schools in Clark County School District, Las Vegas, Nevada*. New York, NY: Scholastic Research and Evaluation Department. The study is ineligible for review because it does not use an eligible design.
- Pearson, L. M., & White, R. N. (2004). *Study of the impact of READ 180 on student performance in Fairfax County Public Schools*. Washington, DC: Policy Studies Associates. The study is ineligible for review because it does not use an eligible design.
- Pittman-Windham, S. P. (2015). *Addressing gaps in student reading: READ 180 program evaluation* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3718851) The study is ineligible for review because it does not use an eligible design.
- Policy Studies Associates. (2002). *Final report: A summary of independent research on READ 180*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Robby, M. A. (2009). *Evaluation of the academic effectiveness of the READ 180 program: Educational software intervention in reading for at risk high school students in Riverside County, California* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3323282) The study is ineligible for review because it does not use a sample aligned with the protocol.
- Santa Rosa County School District. (n.d.). *READ 180 update: Santa Rosa County School District*. Santa Rosa, FL: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (n.d.). *A summary of READ 180 effectiveness research*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (n.d.). *Platinum performers: Blackhawk Middle School*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (n.d.). *Platinum performers: Carthage Central School District*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (n.d.). *Platinum performers: Glendale Union High School District*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.

- Scholastic, Inc. (n.d.). *Platinum performers: Indian River School District*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (n.d.). *Platinum performers: Madison Middle School*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (n.d.). *Platinum performers: Martin Luther King Elementary School*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (n.d.). *Platinum performers: Sebastopol Attendance Center*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (n.d.). *Platinum performers: Westwood Middle School, Alachua County Public Schools, Gainesville, Florida*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2004). *A study of READ 180 at Shiprock High School in Central Consolidated School District on the Navajo Indian Reservation, New Mexico*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2005). *1999–2004 compendium of READ 180 research*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2005). *READ 180 implementation year study*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2006). *Compendium of READ 180 research*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2006). *READ 180: A heritage of research*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2006). *Research foundation paper: READ 180 aligned to No Child Left Behind*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2006). *Special education students, Selbyville Middle and Sussex Central Middle Schools, Indian River School District (Delaware)*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2008). *Compendium of READ 180 research*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic, Inc. (2011). *Compendium of READ 180 research*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research and Evaluation Department, & Interactive, Inc. (2002). *Final Report: Study of READ 180 in the Council of Great City Schools*. New York, NY: Author. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (n.d.). *New York City Public Schools, region 8 (third grade)*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation (n.d.). *Northeastern Florida School District research update*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Calera Elementary School, Calera, Alabama*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Carter Junior High School, Arlington, Texas*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Harding Senior High School, St. Paul, Minnesota*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Highview Elementary School, Nanuet, New York*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Holmes Middle School, Alexandria, Virginia*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.

- Scholastic Research & Validation. (2003). *Lufkin High School, Lufkin, Texas*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Marshall Middle School, Beaumont, Texas*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Miami Lakes Educational Center, Miami Lakes, Florida*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Raba Elementary School, San Antonio, Texas*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *READ 180: Compilation of school studies: Impact study. READ 180 reports guide: Using data to drive instruction* (p. 80). New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Rochester Elementary School, Rochester, Pennsylvania*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Rogers Middle School, Boston, Massachusetts*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *South Ocean Middle School, Patchogue, New York*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Transitional Resource Education Center, Kansas City, Kansas*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Two Eagle River School, Pablo, Montana*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Washington Middle School, Kenosha, Wisconsin*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2003). *Wyandotte County Juvenile Center, Kansas City, Kansas*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2007). *Department of Defense Education Activity Schools (DoDEA)*. New York, NY: Scholastic, Inc. The study is ineligible for review because it is out of the scope of the protocol.
- Scholastic Research & Validation. (2007). *An evaluation of a community college pilot*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use a sample aligned with the protocol.
- Scholastic Research & Validation. (2007). *Phoenix College, AZ*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use a sample aligned with the protocol.
- Scholastic Research & Validation. (2008). *Desert Sands Unified School District, CA*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use a sample aligned with the protocol.
- Scholastic Research & Validation. (2008). *Martin County Public Schools, FL*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2008). *Peoria Public School District 150, IL*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2008). *Socorro Independent School District, TX*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Additional Source:**
- Scholastic, Inc. (n.d.). *Platinum performers: Socorro Independent School District, TX*. New York, NY: Author.
- Scholastic Research & Validation. (2008). *Traverse City Area Public Schools, MI*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2009). *Colton Joint Unified School District research update*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2009). *Lawrence Public Schools research update*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.

- Scholastic Research & Validation. (2010). *Columbus City Schools research update*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Scholastic Research & Validation. (2012). *Deer Valley Unified School District research update*. New York, NY: Scholastic, Inc. The study is ineligible for review because it does not use an eligible design.
- Sigears, K. A. (2009). *The impact of the implementation of the Scholastic READ 180 model on reading skills development of middle school students with learning disabilities as compared to those using the Traditional Resource Reading model* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3320196) The study is ineligible for review because it does not use a sample aligned with the protocol.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, 79(4), 1391–1466. The study is ineligible for review because it does not use an eligible design.
- Sousa, D. A. (2005). *How the brain learns to read*. Thousand Oaks, CA: Corwin Press. The study is ineligible for review because it is out of scope of the protocol.
- Teja, D. L. (2014). *The effects of the READ 180 program on oral reading fluency, linguistic comprehension, and reading comprehension with secondary special education students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (UMI No. 3633503). The study is ineligible for review because it does not use an eligible design.
- Thomas, J. (2003). *Reading program evaluation: READ 180 grades 4–8*. Kirkwood, MO: Kirkwood School District. The study is ineligible for review because it does not use an eligible design.
- Thorpe, P. K. (2003). *Escher's intersecting worlds: Evaluation as a reflection of the evaluator, the evaluator being reflected in the evaluation*. Wichita, KS: Wichita Public Schools. <http://files.eric.ed.gov/fulltext/ED479692.pdf>. The study is ineligible for review because it does not use an eligible design.
- Vintinner, J. P. (2009). *A content analysis of vocabulary instruction in high school commercial literacy programs* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI no 3356488) The study is ineligible for review because it does not use an eligible design.
- Vogel, J. T. (2013). *A case study on the impact of the READ 180 reading intervention program on affective and cognitive reading skills for at-risk secondary level students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3557305) The study is ineligible for review because it does not use an eligible design.
- Wahl, M. (2008). *READ 180 Enterprise Edition*. Tallahassee, FL: Florida Center for Reading Research. The study is ineligible for review because it does not use an eligible design.
- White, L. M. (2007). *A middle school under review: A study on the effectiveness of two implementation models of the READ 180 program on special education students' reading performance* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3291717) The study is ineligible for review because it does not use an eligible design.
- Witkowski, P. M. (2004). *A comparison study of two intervention programs for reading-delayed high school students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (UMI No. 3135832). The study is ineligible for review because it does not use a sample aligned with the protocol.
- Zhu, J. (2008). *Investigating intervention effect of a reading program for low-achieving incarcerated youth including simulation studies for longitudinal research* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3325830) The study is ineligible for review because it does not use a sample aligned with the protocol.
- Zvoch, K., & Letourneau, L. (2006). *Closing the achievement gap: An examination of the status and growth of ninth grade READ 180 students*. Las Vegas, NV: Clark County School District. The study is ineligible for review because it does not use an eligible design.

**Appendix A.1: Research details for Fitzgerald and Hartry (2008)**

Fitzgerald, R., & Hartry, A. (2008). *What works in afterschool programs: The impact of a reading intervention on student achievement in the Brockton Public Schools (phase II)*. Berkeley, CA: MPR Associates, Inc. and the National Partnership for Quality Afterschool Learning at SEDL.

*Additional sources:*

Kim, J. S., Capotosto, L., Hartry, A., & Fitzgerald, R. (2011). Can a mixed-method literacy intervention improve the reading achievement of low-performing elementary school students in an after-school program? Results from a randomized controlled trial of READ 180 Enterprise. *Educational Evaluation and Policy Analysis*, 33(2), 183–201.

Vaden-Kiernan, M., Hughes Jones, D., & Rudo, Z. (2008). *The National Partnership for Quality Afterschool Learning randomized controlled trial studies of promising afterschool programs: Summary of findings*. Austin, TX: SEDL Afterschool Research Consortium. <http://files.eric.ed.gov/fulltext/ED513822.pdf>.

**Table A1. Summary of findings**

**Meets WWC group design standards without reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Comprehension	4 schools/483 students	+6	Yes
General literacy achievement	4 schools/185 students	0	No
Reading fluency	4 schools/297 students	+4	No
Alphabetics	4 schools/482 students	+2	No

**Setting** The study included students in grades 4, 5, and 6 in four elementary schools in Brockton, Massachusetts.

**Study sample** Brockton Public Schools identified four of its 16 elementary schools to participate in the study. Schools were chosen because they had a large number of students reading below grade level, they had adequate facilities, and afterschool programs already existed in the schools. Students who enrolled in the afterschool program at each of these four schools were randomly assigned within school- and grade-blocks to be in either a *READ 180*® classroom or a comparison classroom.

The study took place over 2 school years (2006–07 and 2007–08). In each study year, 24 afterschool classrooms participated: 12 *READ 180*® classrooms and 12 comparison group classrooms. The sizes of these afterschool classes ranged from eight to 17 students. A total of 36 teachers participated in the study in Year 1, and 30 teachers participated in Year 2.

There are three analytic samples of interest in this study: (1) Cohort 1, first year sample (297 students); (2) Cohort 2, first year sample (187 students); and (3) Cohorts 1 and 2, combined second year sample (294 students). Findings from the Cohort 1, first year sample are presented in Kim et al. (2011). Although findings from this sample were also presented in Fitzgerald and Hartry (2008), sample sizes and findings differed slightly between the two sources, and the WWC opted to use the most recent reference to use in this report. Findings from the Cohort 2, first year sample and the Cohorts 1 and 2, second year sample are presented in Fitzgerald and Hartry (2008).

As reported in Kim et al. (2011), there were 155 students in the *READ 180*<sup>®</sup> group at baseline in the fall of 2007 (Cohort 1). Of these students, 67% were eligible for free or reduced-price lunch; 52% were female; and the average age of students was 10.6 years. At baseline in the fall of 2007, there were 157 students in the comparison group: 71% were eligible for free or reduced-price lunch; 56% were female; and the average age of students was 10.6 years. Across both groups in Cohort 1, 28% of students were White, 54% of students were African American, 12% were Hispanic, and 6% were other races or ethnicities. Across both groups, 36% of students were in grade 4, 44% of students were in grade 5, and 20% of students were in grade 6.

Detailed information on the Year 2 sample, which is a combination of the Cohort 1, second year and Cohort 2, first year samples, is provided in Fitzgerald and Hartry (2008). The intervention group in Year 2 included 152 students. Of these students, 49% were female; 92% were eligible for free or reduced-price lunch; 19% were in special education; 55% were African American, 32% were White, 7% were Hispanic, 5% were Asian American, and 2% were from other ethnic backgrounds. The comparison group in Year 2 also included 152 students. Of these students, 57% were female; 90% were eligible for free or reduced-price lunch; 18% were in special education; 43% were African American, 38% were White, 10% were Hispanic, 5% were Asian, and 5% were from other ethnic backgrounds.

### Intervention group

The study tested the *READ 180*<sup>®</sup> *Enterprise* intervention. Students in the intervention condition received the *READ 180*<sup>®</sup> structured reading program in an afterschool setting. Although the *READ 180*<sup>®</sup> program was implemented in an afterschool setting, the key program components were implemented, including the structuring of time to include whole-class instruction, as well as three rotations focused on (1) time using *READ 180*<sup>®</sup> software, (2) modeled and independent reading, and (3) small-group direct instruction. Because of the reduced 60-minute session length (relative to the standard *READ 180*<sup>®</sup> 90-minute session length), the program developer devised a schedule in which, on any given day, students would rotate through two rather than three of the small-group centers. Student workbooks (“*rBooks*<sup>®</sup>”) were also provided in keeping with the program design, and the intended class size of 15 or fewer students was generally maintained. In Year 1, *READ 180*<sup>®</sup> students received the program 4 days per week in 60-minute sessions for 23 weeks. In Year 2, three of the four study schools changed the schedule so that the program was implemented for only 2 days per week in 90-minute sessions. The fourth school provided the program 4 days per week and in 90-minute sessions.

### Comparison group

Students in the comparison group attended Brockton Public Schools’ standard afterschool program, which generally includes 40 minutes of homework, 1 hour of another structured learning activity such as math or reading, and the remainder of the time in physical exercise or recreation. Instructors could choose from 16 structured learning activities, including math games, reading, art projects, or science activities, or they could develop their own activities. In Year 1, comparison group students attended the regular afterschool program for 4 days each week. In Year 2, three of the four schools switched to a 2-day-per-week schedule for the regular afterschool program, while the fourth school retained the 4-day-per-week schedule.

### Outcomes and measurement

Baseline reading skills were measured using state test scores from the spring prior to enrollment in the study. Outcomes were measured by study-administered reading assessments (Stanford 10 and DIBELS) in the spring following enrollment. The study reported several outcomes that met WWC standards in relevant domains for this protocol: general literacy achievement (Stanford 10 Total Reading Score [Cohort 2 only]), alphabetics (Stanford 10 Spelling subtest), reading fluency (DIBELS Oral Reading Fluency subtest), and comprehension (Stanford 10 Reading Comprehension and Vocabulary subtests). DIBELS outcomes are reported for the full sample for Cohort 1 only; findings from Cohort 2 on the DIBELS were separated by grade level and are presented as supplemental findings in Appendix D. Supplemental findings are also presented on the above-referenced outcomes for the second-year findings for the combined cohorts (i.e., Cohort 1 after 2 years and Cohort 2 after 1 year). The supplemental findings do not factor into the intervention's rating of effectiveness.

This study includes afterschool program attendance, attitudes toward reading, a test of exposure to print, and implementation measures that are not eligible for review under the Adolescent Literacy review protocol.<sup>8</sup>

For a more detailed description of these outcome measures, see Appendix B.

### Support for implementation

Scholastic, Inc., the publisher of *READ 180*<sup>®</sup>, provided professional development services to participating teachers. These services consisted of a full day of training prior to the launch of the *READ 180*<sup>®</sup> intervention, as well as a half-day of training after approximately 6 weeks of implementation. During the implementation period, a Scholastic trainer periodically met with all of the teachers implementing *READ 180*<sup>®</sup> to discuss challenges and identify solutions. All teachers also had access to an online professional development program, called RED, provided by Scholastic.



**Appendix A.2: Research details for Kim et al. (2010)**

Kim, J. S., Samson, J. F., Fitzgerald, R., & Hartry, A. (2010). A randomized experiment of a mixed-methods literacy intervention for struggling readers in grades 4–6: Effects on word reading efficiency, reading comprehension and vocabulary, and oral reading fluency. *Reading and Writing: An Interdisciplinary Journal*, 23(1), 1109–1129.

**Table A2. Summary of findings**

**Meets WWC group design standards without reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Comprehension	3 schools/264 students	+2	No
General literacy achievement	3 schools/264 students	+2	No
Reading fluency	3 schools/264 students	+4	Yes
Alphabetics	3 schools/264 students	-1	No

**Setting** The study included students in grades 4, 5, and 6 in three elementary schools in Brockton, Massachusetts. These three schools differed from the four schools studied in Fitzgerald and Hartry (2008).

**Study sample** Students were recruited from three elementary schools with a large percentage of struggling readers. To be eligible for the study, students must have been in grades 4–6 and have scored below the proficiency level on their most recent MCAS ELA test. Eligible students whose parents provided active consent were randomly assigned to an afterschool program that either used a modified *READ 180*<sup>®</sup> program or the district’s standard curriculum.

The baseline study sample was evenly distributed between students in grades 4, 5, and 6 (34.4%, 37.1%, 28.6%, respectively) and between girls and boys (50.3% and 49.7%, respectively). Over 80% of students were eligible for free or reduced-price lunch. Just over a fifth (21.1%) of students in the baseline sample had disabilities, and over 75% were minority students (51.5% African American, 22.2% White, 20.8% Hispanic, and 5.5% other).

**Intervention group** The intervention group attended a 2-hour afterschool program 4 days per week for 23 weeks from October 2005 through April 2006. The first hour was dedicated to a snack and homework. The second hour was dedicated to *READ 180*<sup>®</sup>. In this study, the standard 90-minute *READ 180*<sup>®</sup> model (version 1.6) was shortened to 60 minutes to accommodate the district’s afterschool program. Teachers implemented three 20-minute rotations, but did not implement the whole-group lesson. The first rotation consisted of a 20-minute individualized computer-assisted *READ 180*<sup>®</sup> instruction, which included structured reading practice with videos, leveled text, and word reading and fluency activities. The rotation focused on a substantive area selected by the student. The second rotation consisted of independent reading of books that were matched to student’s Lexile level. The third rotation consisted of small-group teacher-directed lessons that were tailored to the reading level of the students in each group.

**Comparison group**

The comparison condition was also implemented 4 days per week over 23 weeks from October 2005 through April 2006. Like the intervention group, the first hour of the comparison condition’s afterschool program was dedicated to a snack and homework. The second hour included both literacy and non-literacy activities; however, the amount of time dedicated to these activities varied each day. Teachers were instructed to implement activities that encouraged attendance in the afterschool program. Each teacher was provided with a selection of 16 activities, including informal art-based projects, games, and commercially-developed materials for afterschool programs in various subject areas (e.g., astronomy, history, geography, space exploration, math, or literacy). The teachers had flexibility in choosing and tailoring which activities to use.

**Outcomes and measurement**

The study measured four outcomes: (1) the Test of Word Reading Efficiency (TOWRE) total score, which is in the alphabetic domain; (2) the Group Reading Assessment and Diagnostic Evaluation (GRADE) total score, which is in the comprehension domain; (3) the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency assessment, which is in the reading fluency domain; and (4) the Massachusetts Comprehensive Assessment System (MCAS) English Language Arts assessment, which is in the general literacy achievement domain.

Supplemental findings are presented for the full sample on the GRADE Comprehension and Vocabulary subtests and on the TOWRE Sight Word Reading and Phonetic Decoding subtests (GRADE and TOWRE total scores are presented in Appendix C). Supplemental findings are also presented for grade 4, 5, and 6 samples on the DIBELS Oral Reading Fluency test. The supplemental findings do not factor into the intervention’s rating of effectiveness.

For a more detailed description of these outcome measures, see Appendix B.

**Support for implementation**

Classrooms were observed twice during the study period and rated from 1 to 3 (low to high fidelity to the intervention). Ratings ranged from 2.9–3 in observations at the beginning of the intervention period and from 2.3–2.8 in observations at the end of the intervention period.

**Appendix A.3: Research details for Swanlund et al. (2012)**

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

**Table A3. Summary of findings**

**Meets WWC group design standards without reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General literacy achievement	5 schools/619 students	+6	Yes

**Setting**

The intervention was implemented in five schools in the Milwaukee Public Schools district.

**Study sample** *READ 180*<sup>®</sup> was implemented in fall 2010 through spring 2011. Students were eligible for the study if they met the guidelines established by Milwaukee Public Schools for entrance into the *READ 180*<sup>®</sup> program. More specifically, students were eligible if they scored at the Minimal or Basic level on the Wisconsin Knowledge and Concepts Examination (WKCE) in the fall of 2009. If WKCE scores were not available, students could still be eligible for the study if they scored at Minimal or Basic on the Discovery Education Assessment Predictive Benchmark Assessment or if teacher assessments indicated that students were performing at least two grade levels below expectations. Students with disabilities were eligible for the study if they completed a 1-year remedial language course, and English learners (ELs) were eligible for the study if they had a Language Acquisition Unit level of 3.0 or higher.

Eligible students in grades 6–10 were randomly assigned to the intervention or comparison group in two stages. The first stage was completed in July 2010, and randomization was conducted within each school-by-grade block, controlling for special education status. This randomization process resulted in 434 students assigned to the *READ 180*<sup>®</sup> group and 375 students assigned to the comparison group. Following the receipt of an updated school enrollment file at the end of July, a second randomization was conducted in August 2010. This second randomization process, which was designed to fill the remaining *READ 180*<sup>®</sup> slots in each school, involved assigning each eligible student a random number, sorting those numbers by school and grade, and then selecting the appropriate number of students based on their assigned number. The second randomization resulted in 158 students assigned to the *READ 180*<sup>®</sup> group and 159 students assigned to the comparison group.

Including both randomizations, a total of 592 students were assigned to the intervention group and 534 to the comparison group. The analysis was conducted on 335 intervention group students and 284 comparison group students.

Among the students for whom data were available, the majority of students in both the *READ 180*<sup>®</sup> and comparison groups was eligible for free or reduced-price lunch (88%) and was African American (70%). About 36% were special education students, and 8% were English learners. Less than half of the students (39%) were female.

**Intervention group** Students were given *READ 180*<sup>®</sup> instruction for 90 minutes each day for the 2010–11 school year. Classes began with 20 minutes of whole-group instruction. Next, students broke out into three groups that provided 20 minutes each of small-group instruction, instructional software, and modeled and independent reading. The class concluded with a 10-minute whole-group wrap-up. Students were to remain in the *READ 180*<sup>®</sup> intervention between 1 and 2 years. If students reached district-approved proficiency levels, they could exit the program early.

Eight reading intervention teachers were hired to teach the supplemental *READ 180*<sup>®</sup> classes, with 15–21 students assigned to each teacher.

**Comparison group** The planned comparison condition called for students to attend their regular ELA class, plus an elective (non-reading related) class or study hall. However, multiple students in the comparison condition enrolled in reading or ELA-related electives, and two comparison students enrolled in the *READ 180*<sup>®</sup> course.

**Outcomes and measurement**

Outcomes in the general literacy achievement domain were measured using the Measures of Academic Progress (MAP) test.

The authors present treatment on the treated (TOT) estimates of *READ 180*<sup>®</sup> impacts on the MAP outcome. This finding meets WWC complier average causal effect (CACE) guidance; however, since the CACE guidance indicates that the ITT estimates should be prioritized when both ITT and TOT estimates are presented, the TOT results are included as supplemental findings in Appendix D.2.

The authors conducted subgroup analyses by special education status and EL status. These subgroup analyses are not eligible for review under the Adolescent Literacy review protocol. The authors also present analyses of intervention effect accounting for different levels of intervention take-up (dose). These analyses included only students in the intervention group, and therefore are not eligible for review under WWC group design standards.

The study also addressed student outcomes related to self-efficacy and constructs of behavioral engagement, emotional engagement, and cognitive engagement with reading, all of which are outside of the relevant domains within the Adolescent Literacy protocol.

For a more detailed description of these outcome measures, see Appendix B.

**Support for implementation**

Teachers received 3 days of *READ 180*<sup>®</sup> training and ongoing training throughout the year. Teachers were also required to participate in monthly roundtable discussions. Building administrators for each school also attended a half-day orientation to the program.

**Appendix A.4: Research details for Interactive, Inc. (2002)**

*Interactive, Inc. (2002). An efficacy study of READ 180: A print and electronic adaptive intervention program, grades 4 and above. Ashland, VA: Author.*

**Table A4. Summary of findings**

**Meets WWC group design standards with reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Comprehension	18 schools/881 students	+16	Yes

**Setting**

The study took place in seven districts in six states: Atlanta, Georgia; Boston, Massachusetts; Columbus, Ohio; Dallas, Texas; Houston, Texas; Miami-Dade, Florida; and San Francisco, California. Outcome data were not available for Atlanta, Miami-Dade, and San Francisco, so the study’s findings are available for only four of the seven districts.

**Study sample**

The study was designed as a randomized controlled trial with assignment at the student level, but students were not assigned entirely by chance. The original study included middle school students from seven districts, but data are reported for only four of these districts.

Students in different grade levels participated across districts. The authors report findings for the following districts by grade level combinations:

- Boston, sixth grade: This sample included 115 students in the intervention group and 105 in the comparison group. Students in the intervention group were from four schools. Students in the comparison group were from seven middle schools, with 30 students in the comparison group attending the same four middle schools as the intervention group, while the others attended three middle schools that did not participate in the intervention.
- Dallas, eighth grade: This sample included 101 students in the intervention group and 142 in the comparison group, all from the same four schools.
- Houston, seventh grade: This sample included 112 students in the intervention group and 40 in the comparison group, all from the same two schools.
- Houston, eighth grade: This sample included 59 students in the intervention group and 36 in the comparison group, all from the same two schools.
- Columbus, sixth and seventh grade (combined): This sample included 119 students in the intervention group and 52 in the comparison. Students in the intervention group came from two schools; students in the comparison group came from three other schools.
- The authors also present findings for a combined sample of Boston, Dallas, and Houston students (all grades).

The study demonstrated baseline equivalence on the Dallas sample, the Columbus sample, and the combined Boston, Dallas, and Houston analytic sample described above and, therefore, received a rating of meets WWC group design standards with reservations. Among the four districts for which outcomes are reported, there were a total of 506 students in the intervention group and 375 in the comparison group.

### Intervention group

The intervention was delivered during the 2000–01 school year. *READ 180*<sup>®</sup> included daily whole-group, small-group, and individual instruction. Literacy instruction was delivered in 90-minute blocks. During the first 10 minutes of the block, students met together with the teacher to receive language arts instruction. The class then broke into three smaller groups that proceeded through 20-minute rotations of small-group instruction (the teacher sat with 5–6 students doing group reading and/or language arts instruction), independent reading (students read leveled paperbacks with the option of adding audio through headphones as modeled reading), and direct instruction (through nine topic-focused CD-ROMs). In using the CD-ROMs, students were presented with a reading passage based on a video that was tailored to the student’s ability level as determined by an electronic placement test administered at the beginning of the program. After the video and the reading passage, students worked through three “zones” on each CD: the word zone (instruction for developing basic decoding skills), the spelling zone (instruction on spelling patterns and sounds), and the success zone (individual assessment for comprehension, word recognition, and fluency skills).

There was some variation across sites in how *READ 180*<sup>®</sup> was implemented. For example, in one school in Boston, teachers set aside 45 of the 90 minutes twice a week to focus on writing skills.

### Comparison group

The comparison condition varied both within and across districts (and in some cases, within schools). For example, the authors report that the Houston Independent School District conducted an audit of their middle school reading curricula and identified 50 to 60 different programs being implemented across the district. In Columbus, the district offered a “Safety Net” program for students who performed at low levels on tests of reading proficiency; schools with a significant number of low-performing students could choose to implement one of a variety of literacy interventions.

### Outcomes and measurement

Reading comprehension was measured in spring 2001 using the Stanford Achievement Test, Ninth Edition (Stanford 9) Total Score in reading (a composite of the Stanford 9 Reading Vocabulary subtest and Reading Comprehension subtest). Three of the four districts included in analyses used the Stanford 9 Total Score as a baseline and outcome measure. The remaining district (Columbus) used only the Stanford 9 Reading Comprehension subtest for the pretest and posttest.

In addition to completing a Stanford 9 multiple choice reading test, students were also supposed to have completed a Stanford 9 open-ended reading assessment. However, some districts did not administer the open-ended assessment. Dallas and Atlanta only administered the multiple-choice reading assessment as the pretest, and Miami implemented only the multiple-choice reading test for both pretest and posttest.

The Scholastic Reading Inventory (SRI) was administered only to students in the intervention group. These scores were not used to evaluate the effectiveness of *READ 180*<sup>®</sup>. The authors also report the results of a teacher survey which measured teachers’ attitudes toward *READ 180*<sup>®</sup>, their utilization of various aspects of the program, and their perceptions of student attitudes toward *READ 180*<sup>®</sup>. Teacher outcomes are not eligible for review under the Adolescent Literacy protocol.

For a more detailed description of these outcome measures, see Appendix B.

### Support for implementation

While district staff from the seven participating districts selected the schools that would participate in the study, the school staff were responsible for the implementation of *READ 180*<sup>®</sup>. Teachers from each site generally reported receiving “good” support from school administrators, though this support declined in some cases over the course of the school year. In the four districts in which *READ 180*<sup>®</sup> was considered to be well implemented (Boston, Dallas, Houston, and Columbus), a district administrator was assigned to be the *READ 180*<sup>®</sup> liaison and oversaw implementation of the program. Teachers in the intervention group were trained in the summer or early fall prior to initial implementation of the program. Although districts could initiate follow-up training, the authors note that teachers were mostly on their own. In responding to a teacher survey, approximately two-thirds of teachers reported that the professional development provided for *READ 180*<sup>®</sup> was not sufficient.

Appendix A.5: Research details for Meisch et al. (2011)

Meisch, A., Hamilton, J., Chen, E., Quintanilla, P., Fong, P., Gray-Adams, K.,...Thornton, N. (2011). *Striving Readers study: Targeted and whole-school interventions—year 5*. Rockville, MD: Westat.

Table A5. Summary of findings

Meets WWC group design standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Comprehension	19 schools/1,023 students	+2	No
General literacy achievement	19 schools/1,023 students	+3	No

**Setting** The study took place in 20 public middle schools (19 after two schools merged) in Newark, New Jersey.

**Study sample** The schools were selected based on several eligibility criteria: being Title I eligible, not already using *READ 180*<sup>®</sup>, serving at least two of the three middle school grades (6, 7, and 8), being categorized as “in need of improvement” under the No Child Left Behind Act, and serving a minimum of 25 eligible students.

Schools were grouped into blocks based on the number of eligible students, the number of years the school had been identified as “in need of improvement”, the number of eligible students whose home language was not English, and the number of eligible students with an Individualized Education Program (IEP). Schools were then randomly assigned within each block to intervention and comparison groups.

This cluster randomized controlled trial included 20 schools at randomization in May 2006, 19 after two comparison schools merged. For the outcomes measured in the analysis, the number of students varied, with larger numbers having 1 year of exposure (1,305 intervention, 1,255 comparison), somewhat fewer having 2 years of exposure (814 intervention, 706 comparison), and even fewer with 3 years exposure (552 intervention, 471 comparison). Students were eligible for *READ 180*<sup>®</sup> if they scored one standard deviation or more below the norm on the New Jersey Assessment of Skills and Knowledge (NJASK) reading subtest.

The majority of students were African American (ranging from 51% in Year 5 to 58% in Year 1) and over 40% of students were Hispanic (ranging from 41% in Year 1 to 45% in Year 5). The sample was roughly equally split between students in grades 6, 7, and 8, with a slightly larger proportion of students in grade 6.

**Intervention group** Eligible students were assigned to classes of 21 students or fewer. *READ 180*<sup>®</sup> was implemented in classrooms as a replacement to the regular curriculum. The instructional model for *READ 180*<sup>®</sup> included five parts, totaling 90 minutes, which included whole-group instruction and small-group instruction with equally sized groups. Each 90-minute session included 20 minutes of whole-group instruction, 20 minutes of small-group instruction in reading comprehension strategies, 20 minutes of independent reading, 20 minutes of software use, and 10 minutes of whole-group wrap-up. Instruction lasted 1 to 3 years.

**Comparison group** Students in the business-as-usual comparison condition received the regular language arts curriculum.

**Outcomes and measurement** Primary findings are based on the study-administered test, the Stanford 10, after 3 years of exposure to the intervention. In the comprehension domain, outcomes include the Reading and Vocabulary subscales of the Stanford 10 assessment. In the general literacy achievement domain, outcomes include the Stanford 10 Language Arts test.

Supplemental findings are presented on Stanford 10 scores for all students after 1 or 2 years of exposure to the intervention, and for African-American, Hispanic, male, and female students after 1, 2, or 3 years of exposure to the intervention. The supplemental findings do not factor into the intervention’s rating of effectiveness.

School attendance was measured using district administrative data; however, this outcome was not eligible for review under the Adolescent Literacy protocol.

For a more detailed description of these outcome measures, see Appendix B.

**Support for implementation** Professional development was provided to teachers of the *READ 180*<sup>®</sup> curriculum and their supporting staff. For teachers, this included 1 to 3 days of large-group training. Classroom support was provided by five Resource Teacher Coordinators (RTCs), who were teacher’s aides. RTCs also attended the teacher training. Technology coordinators for the *READ 180*<sup>®</sup> software provided support for technical issues encountered by the teachers. These technology coordinators had half day of training in Years 1 and 2. Finally, principals of *READ 180*<sup>®</sup> schools received 2 hours of training in Years 1 and 2.

**Appendix A.6: Research details for Sprague et al. (2012)**

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2012). *Springfield-Chicopee School Districts Striving Readers program final report Years 1-5: Evaluation of implementation and impact*. Providence, RI: The Education Alliance at Brown University.

*Additional sources:*

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2009). *Springfield-Chicopee School Districts Striving Readers (SR) program Year 2 report: Evaluation of implementation and impact*. Providence, RI: The Education Alliance at Brown University.

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2010). *Springfield-Chicopee School Districts Striving Readers (SR) program Year 3 report: Evaluation of implementation and impact*. Providence, RI: The Education Alliance at Brown University.

Sprague, K., Zaller, C., Kite, A., & Hussar, K. (2011). *Springfield-Chicopee School Districts Striving Readers (SR) program Year 4 report: Evaluation of implementation and impact*. Providence, RI: The Education Alliance at Brown University.

**Table A6. Summary of findings**

**Meets WWC group design standards with reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General literacy achievement	5 schools/456 students	+7	Yes



**Setting** The study was conducted in two school districts, Chicopee and Springfield, in western Massachusetts.

**Study sample** In each of the 5 study years, students in five study schools were screened prior to random assignment. Students at least two—but less than four—grade levels behind in reading performance were selected to participate. Students were excluded from the sample if (a) they had an IEP that specified reading supports not compatible with *READ 180*<sup>®</sup>, (b) they lacked sufficient English language proficiency, (c) their parents opted out of the study, (d) they were enrolled in an off-campus evening school, (e) they were deemed not to be a “struggling reader” based on grade history and MCAS scores, or (f) they could not be located in school enrollment records.

Over the five annual cohorts, a total of 548 ninth-grade students with five teachers per year (one in each of five schools) were randomly assigned to the *READ 180*<sup>®</sup> group. The *READ 180*<sup>®</sup> analysis sample included 231 students taught by five teachers in five schools. This analysis sample was comprised of 74% racial and/or ethnic minorities, 61% female students, 18% special education students, and 3% English learners. A majority of students (69%) were eligible for free or reduced-price lunch.

A total of 566 students with five teachers per year (one in each of five schools) were randomly assigned to the comparison group. The analysis sample for the comparison group includes 225 students taught by five teachers in five schools. This analysis sample was comprised of 71% racial and/or ethnic minorities, 53% female students, 19% special education students, and 4% English learners. A majority of students (74%) were eligible for free or reduced-price lunch.

Results for additional samples were reported in Year 2, Year 3, and Year 4 reports. In the Year 2 report, which includes impact estimates for a sample combining Cohorts 1–2, there were 128 students in the intervention group and 113 students in the comparison group. The Year 3 report presents findings for Cohorts 1–3, which included 175 students in the intervention group and 159 in the comparison. The Year 4 report presents findings on Cohorts 1–4, which included 186 students in the intervention group and 178 in the comparison. These supplemental findings do not factor into the intervention’s rating of effectiveness.

**Intervention group** The *READ 180*<sup>®</sup> intervention was delivered as a 90-minute daily supplement to the standard ninth-grade ELA course. A typical daily session included 20 minutes of whole-class instruction, 60 minutes of small-group breakouts involving direct instruction, independent work using program software, and modeled or independent reading. In addition, the intervention included recommended instructional strategies and instructional materials, including videos and interactive work texts. The *READ 180*<sup>®</sup> curriculum was paced to be completed over 125–145 school days; the average number of sessions attended by each student was not reported.

**Comparison group** Students in the comparison condition received the standard ELA course (as did students in the intervention condition), as well as supplemental services ordinarily available to all students. In practice, comparison group students had minimal access to supplemental services.

None of the comparison group teachers reported having any past experience with the *READ 180*<sup>®</sup> program, and they did not receive formal professional development in literacy instruction beyond what was customarily provided to all teachers. Use of multimedia appears to have been much more limited in the comparison group than in the intervention group.

**Outcomes and measurement**

This study used the Stanford Diagnostic Reading Test, fourth edition (SDRT-4) as a measure of general literacy achievement. The overall score on the SDRT-4 combines measures of phonetic analysis, vocabulary, comprehension, and scanning, but only the overall normal curve equivalent and scaled scores are reported in this study. The test was administered to study participants in the spring of their ninth-grade year, the year following random assignment.

Supplemental findings are reported on the SDRT-4 for Cohorts 1–2, Cohorts 1–3, and Cohorts 1–4. These supplemental findings do not factor into the intervention’s rating of effectiveness.

For a more detailed description of these outcome measures, see Appendix B.

**Support for implementation**

Teachers implementing the intervention were required to participate in professional development activities. Those implementing *READ 180*® for the first time were required to complete 52 hours of professional development over the course of the year in online training (seven sessions), group seminars (up to 30 hours), and individual face-to-face sessions (up to 16 hours). Less professional development was required of more experienced users: teachers with 3 years of prior *READ 180*® experience had to complete only 8 hours, and those implementing their fifth year had no such requirement.

**Appendix A.7: Research details for White et al. (2006)**

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

*Additional source:*

Scholastic Research and Results. (2008). *READ 180: Longitudinal evaluation of a ninth-grade reading intervention (2003–2006)*. New York, NY: Scholastic, Inc.

**Table A7. Summary of findings**

**Meets WWC group design standards with reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General literacy achievement	3,688 students	+7	Yes

**Setting**

The study took place in the Phoenix Union High School District in Arizona.

**Study sample**

All students in grades 9 and 10 who were reading one or more grade levels below their assigned grade level were considered for the study; however, the *READ 180*® program did not have space for all eligible students. Students in the *READ 180*® program were included in the study if they met all of the following criteria:

- had two or more SRI scores at least 45 days apart (to allow for analysis of changes in SRI scores).
- had Stanford 9 and/or TerraNova scores from both eighth and ninth grades.
- had a matched nonparticipant available for the purposes of comparison.

Students were matched on eighth-grade reading proficiency (measured by the Stanford 9 in 2003–04 and the TerraNova in 2004–05 and 2005–06), EL status, special education eligibility, gender, and ethnicity.

Four cohorts of students were studied:

**Cohort 1:** This cohort included ninth graders in the 2003–04 school year. This cohort did not meet eligibility requirements specified in the Adolescent Literacy protocol because 53% of students from this cohort were eligible for EL services.

**Cohort 2:** This cohort included 1,630 students in grade 9 in the 2004–05 school year. The sample included 815 students in each condition, among whom: 40% of the intervention (*READ 180*<sup>®</sup>) group and 44% of comparison group students were eligible for EL services, 7% of the intervention group and 10% of comparison group students were eligible for special education, 48% of the intervention group and 49% of comparison group were female, and 84% of the intervention group and 86% of comparison group students were Hispanic. Follow-up outcomes were collected 1 year later in tenth grade (2005–06). Although the additional source for this study (Scholastic Research and Results, 2008) indicated that there were 821 students in each condition, a query response received from the authors confirmed that there were 815 students in each group (as reported in White et al., 2006).

**Cohort 3:** This cohort, as described in Scholastic Research and Results (2008), included 2,058 students in grade 9 in the 2005–06 school year. The White et al. (2006) article indicated Cohort 3 included students in grade 10 in the 2003–04 school year, but this sample did not have a comparison group and was thus determined to be ineligible for review. Outcomes for this cohort are only available for ninth grade; tenth-grade follow-up outcomes are not available.

**Cohort 4:** This cohort, as described in Scholastic Research and Results (2008), included students in tenth grade in the 2004–05 school year; however, this cohort did not have a comparison group, and therefore, is ineligible for review.

### Intervention group

No details were provided about the intervention except its name and version: Scholastic *READ 180*<sup>®</sup> program, Stage C, Version 1.6.

### Comparison group

No information was provided about the comparison condition.

### Outcomes and measurement

One outcome was included in the domain of general literacy achievement (TerraNova Reading Test). All TerraNova scores were reported as normal curve equivalent scores, and were available for ninth grade students in both Cohort 2 and Cohort 3.

Supplemental findings on the TerraNova Reading Test are presented for students in Cohort 2 that scored below 40 NCE on the pretest and students that scored above 40 NCE on the pretest. These supplemental findings do not factor into the intervention's rating of effectiveness.

Scholastic Reading Inventory (SRI) posttest scores were collected only from the intervention group and thus are not eligible for review. The study also addressed two outcomes that meet review requirements in the domain of reading comprehension: the Stanford 9 and the AIMS Reading Test. However, the Stanford 9 was administered as an outcome measure to Cohort 1 only, which was not eligible for review, and baseline equivalence was not established for the AIMS.

For a more detailed description of these outcome measures, see Appendix B.

### Support for implementation

Support for implementation was not described in the report.

Appendix A.8: Research details for White et al. (2005)

White, R., Williams, I., & Haslam, M. B. (2005). *Performance of District 23 students participating in Scholastic READ 180*. Washington, DC: Policy Studies Associates.

Table A8. Summary of findings

Meets WWC group design standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Comprehension	16 schools/1,097 students	+14	No

**Setting** The study took place in 16 schools in New York City’s District 23.

**Study sample** Students receiving *READ 180*® instruction in the 16 participating schools were compared to students within the same schools who had never participated in *READ 180*®.

The full sample of 617 *READ 180*® students and 4,619 students in the comparison group had similar percentages of African-American students (86% intervention, 84% comparison), Hispanic students (14% intervention, 15% comparison), female students (54% intervention, 51% comparison), students eligible for special education (6% intervention, 11% comparison), and students eligible for free or reduced-price lunch (91% intervention, 90% comparison). Both groups had the same percentages of students who were eligible for EL services (3%) and who were recent immigrants (3%).

Main analysis samples were excluded from review because either they were not eligible or they did not meet WWC group design standards. For example, there were no intervention students in the grade 7 analysis sample; therefore, grade 7 students were excluded from this review. Moreover, results of an author query revealed that the samples of students in grades 4, 5, 6, and 8 did not establish baseline equivalence on the analytic sample, either combined or separately by grade.

This review is based on the analytic sample which consists of three subgroups of students that were found to be equivalent at baseline:

- **Grade 6, proficiency level 2 [Basic]:** This subgroup consisted of 64 students in the intervention group and 407 in the comparison group.
- **Grade 8, proficiency level 2 [Basic]:** This subgroup consisted of 47 students in the intervention group and 378 in the comparison group.
- **Grade 8, proficiency level 3 [Proficient]:** This subgroup consisted of 10 students in the intervention group and 191 in the comparison.

**Intervention group** The intervention group received *READ 180*® during the 2001–02 school year.

**Comparison group** The comparison group received business-as-usual instruction in the same schools that served the intervention group during the 2001–02 school year.

**Outcomes and measurement**

The study reported outcomes after 1 year of program implementation.

For the pretest, students took a reading test developed by CTB/McGraw-Hill for the City of New York. This test produces scores that can be aligned with and compared to the New York State Department of Education end-of-year tests. For the posttest, students in grade 6 took the CTB-McGraw Hill Reading Test developed for the City of New York. Students in grade 8 took the New York State Department of Education end-of-year test in ELA (NYSDE/ELA).

For a more detailed description of these outcome measures, see Appendix B.

**Support for implementation**

Support for implementation was not described in the report.

**Appendix A.9: Research details for Yurchak (2013)**

Yurchak, S. M. (2013). *The effect of READ 180 on the reading achievement of struggling readers in a large, public, urban high school in northern New Jersey (Doctoral dissertation)*. Available from ProQuest Dissertations and Theses database. (UMI No. 3613825)

**Table A9. Summary of findings**

**Meets WWC group design standards with reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Comprehension	1 school/134 students	-4	No

**Setting**

The study took place in a single, large urban high school in northern New Jersey.

**Study sample**

This study used a quasi-experimental design, matching students in grade 9 receiving *READ 180*<sup>®</sup> instruction with students in regular English 9 classes on pretest Language Arts Literacy (LAL) scores from the grade 8 state assessment. Students were eligible for the study if they did not meet proficiency levels on the LAL portion of the grade 8 state assessment, and if they were on the general education track in school. The overall sample is made up of students in grade 9 from three consecutive cohorts from the 2007–08, 2008–09, and 2009–10 school years. Only students with complete data (those who were in the same school district in grades 8–11) were eligible to be matched and be in the study.

The study took place in one school. *READ 180*<sup>®</sup> was offered in six class sections the first year, four class sections the second year, and five class sections the third year. Across the cohorts, 67 students had complete data and were able to be matched to students who had participated in English 9.

The intervention and comparison groups were both 52% male. The intervention group was 52% White, 27% Hispanic, and 20% African American. The comparison group was 52% Hispanic, 34% White, and 13% African American. The majority of students in both the intervention group (61%) and the comparison group (72%) qualified for free or reduced-price lunch.

### Intervention group

Students in the intervention group were exposed to the *READ 180*<sup>®</sup> intervention for a full school year. Classes were 80 minutes daily, which closely resembled the prototypical 90-minute five-class instructional model. Of the 15 *READ 180*<sup>®</sup> sections, 13 were inclusion-based classrooms, and two were general education. Inclusion classes were taught by a content-certified English teacher and a special education teacher; general education sections were taught by a content-certified English teacher.

### Comparison group

Comparison students took part in the standard English 9 course, which was 40 minutes long.

### Outcomes and measurement

Outcomes in the comprehension domain were measured using the LAL portion of the New Jersey High School Proficiency Assessment (HSPA), which included a Reading Cluster and an Analyzing Text Cluster.

Supplemental findings are presented for the Reading Cluster and the Analyzing Text Cluster for male, female, and African-American students. These supplemental findings do not factor into the intervention's rating of effectiveness.

The authors also presented outcomes on the HSPA Interpreting Text Cluster (comprehension domain); however, it does not meet reliability requirements.

The authors presented grade 9, 10, and 11 final English grades for the intervention and comparison students. Teacher-reported grades are not eligible based on the Adolescent Literacy protocol. The authors also included SRI Lexile scores for the 2009–10 intervention cohort; however, since SRI Lexile scores were not available from the comparison group, this design is not eligible for review under the WWC group design standards.

For a more detailed description of these outcome measures, see Appendix B.

### Support for implementation

Teachers delivering the intervention were trained by *READ 180*<sup>®</sup> personnel or others in the district who were previously trained in *READ 180*<sup>®</sup>.

Appendix B: Outcome measures for each domain

General literacy achievement	
<i>Massachusetts Comprehensive Assessment System (MCAS) English Language Arts (ELA) assessment</i>	The MCAS is the standardized assessment for students in Massachusetts. The MCAS ELA assessment is designed to evaluate student knowledge and mastery of ELA, and results are presented as scale scores. A scale score of 240 was used as the cut point for proficiency determinations (as cited in Kim et al., 2010).
<i>Measures of Academic Progress (MAP)</i>	The Northwest Evaluation Association (NWEA) MAP benchmark assessment is a computer-adaptive assessment that is aligned to state standards in Wisconsin. It was administered three times per year (October, February, and June) district wide in grades 3–10 (as cited in Swanlund et al., 2012).
<i>Stanford Diagnostic Reading Test, Fourth Edition (SDRT-4)</i>	The SDRT-4 assesses four indicators of reading achievement: decoding, vocabulary, comprehension, and scanning. This assessment was administered to all students school-wide in the spring of each school year (as cited in Sprague et al., 2012).
<i>Stanford 10 Language Arts subtest</i>	The Stanford 10 Language Arts subtest is designed to assess language mechanics (e.g., capitalization, punctuation), language expression (e.g., writing strategies, sentence structure), and students' assessment of language for extraneous information, descriptive language, and the combining of simple sentences (as cited in Meisch et al., 2011).
<i>Stanford 10 Total Reading Score</i>	The Stanford 10 Total Reading Score is a composite of the vocabulary and reading comprehension subtests. The assessment also includes a Word Study Skills subtest for grade 4; however, this subtest was only administered in Year 2 of the study (as cited in Fitzgerald & Hartry, 2008).
<i>TerraNova Reading Test</i>	The TerraNova Reading Test is a multiple-choice, standardized assessment. Number of correct responses (NCR) scores were reported for this assessment (as cited in White et al., 2006).
Reading fluency	
<i>Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency assessment</i>	The DIBELS Oral Reading Fluency assessment is a standardized, individually-administered assessment that measures students' reading accuracy and reading rate. Reading rates are measured as the number of words read correctly per minute. Test-retest reliabilities for this assessment range from .92 to .97 (as cited in Kim et al., 2011).
Comprehension	
<i>CTB/McGraw Hill Reading</i>	The CTB/McGraw Hill Reading assessment is administered annually by the New York City Department of Education. This assessment, which is administered to students in grades 3, 5, 6, and 7, includes three subtests: Information and Understanding; Literary Response; and Expression and Critical Analysis. Student performance on each component is reported as the percent of items answered correctly. Scale scores are aligned to the New York State ELA assessment, so proficiency level cut points are the same; however, this assessment is not vertically scaled (as cited in White et al., 2005).
<i>Group Reading Assessment and Diagnostic Evaluation (GRADE) Total Score</i>	The GRADE is a group administered assessment that includes subtests in vocabulary, sentence comprehension, and passage comprehension. Reported alternate form reliabilities were above .87 for grades 4–6 (as cited in Kim et al., 2010).
<i>New Jersey High School Proficiency Assessment (HSPA) Analyzing Text Cluster</i>	The HSPA is a state-mandated assessment, required of every student entering eleventh grade in New Jersey. It is designed to assess students' level of proficiency in language arts literacy, and the Analyzing Text Cluster consists of two reading passages: narrative and persuasive. Students answered 10 multiple choice questions for each passage (worth one point each) and two open-ended questions for each passage (worth four points each). The 2009 HSPA reliability estimates were .750 (Cronbach's Alpha) for the Analyzing Text Cluster (as cited in Yurchak, 2013).
<i>New Jersey HSPA Reading Cluster</i>	The HSPA is a state-mandated assessment, required of every student entering eleventh grade in New Jersey. The HSPA Reading Cluster is an overall assessment that incorporates two smaller clusters: Interpreting Text and Analyzing Text. These two clusters assessed two reading passages: a narrative passage and a persuasive passage. Each narrative had both multiple choice and open-ended questions (as cited in Yurchak, 2013).
<i>New York State end-of-year test in ELA</i>	The New York State end-of-year test in ELA is administered annually to students in grades 4 and 8. This standardized test is published by McGraw-Hill and contains multiple-choice questions based on brief reading passages. A performance assessment is also included, in which students listen to and read passages and write responses to open-ended questions based on the passages. This assessment is administered by the New York State Education Department and is not vertically scaled (as cited in White et al., 2005).

<i>Stanford Achievement Test, Ninth Edition (Stanford 9) Total Reading</i>	This assessment is a composite of the Stanford 9 Reading Comprehension subtest and the Stanford 9 Vocabulary subtest (as cited in Interactive, Inc., 2002).
<i>Stanford Achievement Test, Tenth Edition (Stanford 10) Reading Comprehension subtest</i>	The Stanford 10 Reading Comprehension subtest is a multiple-choice assessment that measures students' comprehension of text read for enjoyment (e.g., fiction, poetry), text read for information purposes (e.g., textbook material), and functional text (e.g., directions, labels). There are six to nine passages per subtest, and each passage is designed to be more complex than the last (as cited in Fitzgerald & Hartry, 2008; Kim et al., 2011; and Meisch et al., 2011).
<i>Stanford 10 Vocabulary subtest</i>	The Stanford 10 Vocabulary subtest is a multiple-choice assessment that assesses concepts such as synonyms, multiple-meaning words, and use of context clues to decipher a word's meaning. An abbreviated battery is available, in addition to the full battery (as cited in Fitzgerald & Hartry, 2008; Kim et al., 2011; and Meisch et al., 2011). The abbreviated battery was used in Kim et al. (2011).
<b>Alphabetics</b>	
<i>Stanford 10 Spelling subtest</i>	The Stanford 10 Spelling subtest is a multiple-choice assessment. This assessment is norm-referenced and vertically scaled (as cited in Fitzgerald & Hartry, 2008 and Kim et al., 2011).
<i>Test of Word Reading Efficiency (TOWRE) Total Score</i>	The TOWRE is designed to assess word reading accuracy and fluency. It is an individually-administered assessment that tests students' ability to recognize familiar words ("sight words") and their ability to "sound out" pseudo-words. Alternate form reliability is reported to exceed .90. The TOWRE Sight Word Reading and TOWRE Phonetic Decoding subtests are presented as supplemental findings since they are components of the TOWRE composite score (as cited in Kim et al., 2010).



Appendix C.1: Findings included in the rating for the comprehension domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Fitzgerald &amp; Hartry (2008)<sup>a</sup></b>								
<i>Stanford Achievement Test, Tenth Edition (Stanford 10) Reading Comprehension</i>	Cohort 1, First year	4 schools/ 296 students	635.41 (32.34)	625.75 (28.17)	9.66	0.32	+12	< .01
<i>Stanford 10 Vocabulary</i>	Cohort 1, First year	4 schools/ 296 students	639.11 (35.74)	630.68 (36.18)	8.43	0.23	+9	< .05
<i>Stanford 10 Reading Comprehension</i>	Cohort 2, First year	4 schools/ 187 students	nr	nr	-0.25	-0.01	0	.95
<i>Stanford 10 Vocabulary</i>	Cohort 2, First year	4 schools/ 187 students	nr	nr	0.78	0.02	+1	.87
<b>Domain average for comprehension (Fitzgerald &amp; Hartry, 2008)</b>						<b>0.14</b>	<b>+6</b>	<b>Statistically significant</b>
<b>Kim et al. (2010)<sup>b</sup></b>								
<i>Group Reading Assessment and Diagnostic Evaluation (GRADE) Total Score</i>	Full sample	3 schools/ 264 students	92.70 (13.22)	92.09 (12.09)	0.61	0.05	+2	> .05
<b>Domain average for comprehension (Kim et al., 2010)</b>						<b>0.05</b>	<b>+2</b>	<b>Not statistically significant</b>
<b>Interactive, Inc. (2002)<sup>c</sup></b>								
<i>Stanford Achievement Test, Ninth Edition (Stanford 9) Total Reading</i>	Boston, Houston, Dallas, grades 6–8	13 schools/ 710 students	648.48 (25.98)	642.42 (31.36)	6.06	0.21	+8	< .01
<i>Stanford 9 Reading Comprehension</i>	Columbus, grades 6–7	5 schools/ 171 students	621.52 (28.18)	602.25 (39.76)	19.27	0.60	+22	< .05
<b>Domain average for comprehension (Interactive, Inc., 2002)</b>						<b>0.40</b>	<b>+16</b>	<b>Statistically significant</b>
<b>Meisch et al. (2011)<sup>d</sup></b>								
<i>Stanford 10 Reading Comprehension</i>	3 years of exposure	19 schools/ 1,023 students	641.74 (22.83)	640.33 (23.91)	1.41	0.06	+2	.40
<i>Stanford 10 Vocabulary</i>	3 years of exposure	19 schools/ 1,023 students	642.91 (25.95)	641.47 (28.21)	1.44	0.05	+2	.51
<b>Domain average for comprehension (Meisch et al., 2011)</b>						<b>0.06</b>	<b>+2</b>	<b>Not statistically significant</b>

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>White et al. (2005)<sup>e</sup></b>								
<i>CTB/McGraw Hill Reading</i>	Grade 6, Level 2	16 schools/ 471 students	642.00 (21.00)	639.00 (19.00)	3.00	0.16	+6	nr
<i>New York State end-of-year test in ELA</i>	Grade 8, Level 2	16 schools/ 425 students	689.00 (18.00)	686.00 (14.00)	3.00	0.21	+8	nr
<i>New York State end-of-year test in ELA</i>	Grade 8, Level 3	16 schools/ 201 students	718.00 (21.00)	707.00 (16.00)	11.00	0.67	+25	nr
<b>Domain average for comprehension (White et al., 2005)</b>						<b>0.35</b>	<b>+14</b>	<b>Not statistically significant</b>
<b>Yurchak (2013)<sup>f</sup></b>								
<i>New Jersey High School Proficiency Assessment Analyzing Text Cluster Score</i>	Full sample	1 school/ 134 students	38.51 (10.60)	39.30 (10.60)	-0.79	-0.07	-3	nr
<i>New Jersey High School Proficiency Assessment Reading Cluster score</i>	Full sample	1 school/ 134 students	41.31 (10.90)	42.70 (11.00)	-1.39	-0.12	-5	nr
<b>Domain average for comprehension (Yurchak, 2013)</b>						<b>-0.10</b>	<b>-4</b>	<b>Not statistically significant</b>
<b>Domain average for comprehension across all studies</b>						<b>0.15</b>	<b>+6</b>	<b>na</b>

**Table Notes:** For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable. nr = not reported.

<sup>a</sup> For Fitzgerald and Hartry (2008), a correction for multiple comparisons was needed and resulted in a WWC-computed critical p-value of .025 for the Cohort 1 Stanford 10 Vocabulary outcome; therefore, the WWC does not find this result to be statistically significant. The p-values presented here were reported in the original study. The WWC calculated the intervention group mean for Cohort 1 by adding the regression coefficient (presented in the mean difference column) to the unadjusted comparison group posttest mean. The intervention and comparison group means and standard deviations for Cohort 2 were not reported in the original study, but author-reported effect sizes matched the WWC's calculations. The mean difference reflects the regression coefficient for the impact estimate. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>b</sup> For Kim et al. (2010), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-value presented here was reported in the original study. The intervention and comparison group means reported in this table are analysis of covariance (ANCOVA)-adjusted, as reported by the authors in response to a query from the WWC. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>c</sup> For Interactive, Inc. (2002), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. The WWC did not need to make corrections for clustering or to adjust for baseline differences. The p-value presented for the combined sample from Boston, Houston, and Dallas (grades 6–8) was reported in the original study. The exact p-value for Columbus (grades 6–7) was not reported in the study, but the WWC-computed p-value of < .01 indicated that this result was statistically significant. The intervention and comparison group means reported in this table are ANCOVA-adjusted, as reported by the authors in the original report. This study is characterized as having a statistically significant positive effect because the mean effect reported was positive and statistically significant. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>d</sup> For Meisch et al. (2011), the WWC did not need to make corrections for multiple comparisons. Baseline data were provided by the authors, and all baseline measures were both within the adjustment range and included in the study's impact models. A correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The p-values presented here were reported in the original study. The intervention and comparison group means reported in this table are regression-adjusted, as reported by the authors in the original report. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>e</sup> For White et al. (2005), the study's full sample received a rating of does not meet WWC group design standards, but the results for the subgroups noted in this table (grade 6, level 2; grade 8, level 2; grade 8, level 3) received a rating of meets WWC group design standards with reservations. Means and standard deviations for these subgroup analyses were provided in response to an author query; the author query response did not include *p*-values. The WWC-computed *p*-values were not statistically significant for the grade 6, level 2 and grade 8, level 2 subgroups, but a *p*-value of .04 was found for the grade 8 level 3 outcome. A correction for multiple comparisons was needed and resulted in a WWC-computed critical *p*-value of .02 for the grade 8, level 3 New York State ELA outcome; therefore, the WWC does not find this result to be statistically significant. The WWC did not need to make corrections for clustering, and adjustments for baseline differences were unnecessary since all three outcomes had baseline differences of zero between intervention and comparison groups. This study is characterized as having a substantively important positive effect because the mean effect reported is positive and not statistically significant but is substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>f</sup> For Yurchak (2013), the WWC did not need to make corrections for clustering or multiple comparisons. The WWC calculated the program group mean using a difference-in-differences approach by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. The author did not report *p*-values in the original study, but the WWC-computed *p*-values were not statistically significant. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

## Appendix C.2: Findings included in the rating for the general literacy achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			<i>p</i> -value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Fitzgerald &amp; Hartry (2008)<sup>a</sup></b>								
<i>Stanford Achievement Test, Tenth Edition (Stanford 10) Total Reading</i>	Cohort 2, First year	4 schools/ 185 students	nr	nr	0.5	0.01	0	.87
<b>Domain average for general literacy achievement (Fitzgerald &amp; Hartry, 2008)</b>						<b>0.01</b>	<b>0</b>	<b>Not statistically significant</b>
<b>Kim et al. (2010)<sup>b</sup></b>								
<i>Massachusetts Comprehensive Assessment System English Language Arts (ELA) Assessment</i>	Full sample	3 schools/ 264 students	232.65 (11.78)	232.17 (11.28)	0.48	0.04	+2	.29
<b>Domain average for general literacy achievement (Kim et al., 2010)</b>						<b>0.04</b>	<b>+2</b>	<b>Not statistically significant</b>
<b>Swanlund et al. (2012)<sup>c</sup></b>								
<i>Measures of Academic Progress</i>	Intent-to-treat sample	5 schools/ 619 students	nr	nr	1.78	0.14	+6	< .05
<b>Domain average for general literacy achievement (Swanlund et al., 2012)</b>						<b>0.14</b>	<b>+6</b>	<b>Statistically significant</b>
<b>Meisch et al. (2011)<sup>d</sup></b>								
<i>Stanford 10 Language Arts</i>	3 years of exposure	19 schools/ 1,023 students	623.15 (24.11)	621.48 (22.63)	1.67	0.07	+3	.32
<b>Domain average for general literacy achievement (Meisch et al., 2011)</b>						<b>0.07</b>	<b>+3</b>	<b>Not statistically significant</b>
<b>Sprague et al. (2012)<sup>e</sup></b>								
<i>Stanford Diagnostic Reading Test, Fourth Edition (SDRT-4)</i>	Cohorts 1–5	5 schools/ 456 students	24.14 (13.37)	21.75 (13.38)	2.39	0.18	+7	.03
<b>Domain average for general literacy achievement (Sprague et al., 2012)</b>						<b>0.18</b>	<b>+7</b>	<b>Statistically significant</b>

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>White et al. (2006)<sup>f</sup></b>								
<i>TerraNova Reading Test</i>	Cohort 2	1,630 students	41.20 (8.90)	38.30 (12.20)	2.90	0.27	+11	< .05
<i>TerraNova Reading Test</i>	Cohort 3	2,058 students	39.00 (9.80)	38.10 (12.30)	0.90	0.08	+3	< .05
<b>Domain average for general literacy achievement (White et al., 2006)</b>						<b>0.18</b>	<b>+7</b>	<b>Statistically significant</b>
<b>Domain average for general literacy achievement across all studies</b>						<b>0.10</b>	<b>+4</b>	<b>na</b>

**Table Notes:** For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable. nr = not reported.

<sup>a</sup> For Fitzgerald and Hartry (2008), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-value presented here was reported in the original study. The intervention and comparison group means and standard deviations were not reported in the original study, but author-reported effect sizes matched the WWC's calculations. The mean difference reflects the regression coefficient for the impact estimate. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>b</sup> For Kim et al. (2010), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-value presented here was reported in the original study. The intervention and comparison group means reported in this table are analysis of covariance (ANCOVA)-adjusted, as reported by the authors in response to a query from the WWC. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>c</sup> For Swanlund et al. (2012), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-value and effect size presented here were reported in the original study. This study is characterized as having a statistically significant positive effect because the mean effect reported was positive and statistically significant. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>d</sup> For Meisch et al. (2011), the WWC did not need to make corrections for multiple comparisons. Baseline data were provided by the authors, and all baseline measures were both within the adjustment range and included in the study's impact models. A correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-value presented here was reported in the original study. The intervention and comparison group means reported in this table are regression-adjusted, as reported by the authors in the original report. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>e</sup> For Sprague et al. (2012), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-value presented here was reported in the original study. The intervention and comparison group means reported in this table are ANCOVA-adjusted, as reported by the authors in the original study. Standard deviations are also covariate-adjusted, which will not yield effect size calculations comparable to other findings reported in this table since the WWC computes effect sizes using unadjusted standard deviations. This study is characterized as having a statistically significant positive effect because the mean effect reported was positive and statistically significant. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>f</sup> For White et al. (2006), the *p*-values presented here were reported in the original study. A correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. Although a difference-in-differences adjustment was needed, it was not applied for Cohort 2 and Cohort 3 because baseline differences were zero. This study is characterized as having a statistically significant positive effect because the mean effect reported was positive and statistically significant. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix C.3: Findings included in the rating for the reading fluency domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Fitzgerald &amp; Hartry (2008)<sup>a</sup></b>								
<i>Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency</i>	Cohort 1, First year	4 schools/ 297 students	106.27 (27.01)	103.73 (24.48)	2.54	0.10	+4	> .05
<b>Domain average for reading fluency (Fitzgerald &amp; Hartry, 2008)</b>						<b>0.10</b>	<b>+4</b>	<b>Not statistically significant</b>
<b>Kim et al. (2010)<sup>b</sup></b>								
<i>DIBELS Oral Reading Fluency</i>	Full sample	3 schools/ 264 students	111.00 (35.52)	107.27 (36.94)	3.73	0.10	+4	.04
<b>Domain average for reading fluency (Kim et al., 2010)</b>						<b>0.10</b>	<b>+4</b>	<b>Statistically significant</b>
<b>Domain average for reading fluency across all studies</b>						<b>0.10</b>	<b>+4</b>	<b>na</b>

**Table Notes:** For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

<sup>a</sup> For Fitzgerald and Hartry (2008), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-value presented here was reported in the original study. The WWC calculated the intervention group mean by adding the regression coefficient (presented in the mean difference column) to the unadjusted comparison group posttest mean. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>b</sup> For Kim et al. (2010), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-value presented here was reported in the original study. The intervention and comparison group means reported in this table are analysis of covariance (ANCOVA)-adjusted, as reported by the authors in response to a query from the WWC. This study is characterized as having a statistically significant positive effect because the mean effect reported was positive and statistically significant. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix C.4: Findings included in the rating for the alphabetics domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Fitzgerald &amp; Hartry (2008)<sup>a</sup></b>								
<i>Stanford Achievement Test, Tenth Edition (Stanford 10) Spelling</i>	Cohort 1, First year	4 schools/ 295 students	630.82 (31.28)	625.88 (37.85)	4.94	0.14	+6	> .05
<i>Stanford 10 Spelling</i>	Cohort 2, First year	4 schools/ 187 students	nr	nr	-1.72	-0.04	-2	.68
<b>Domain average for alphabetics (Fitzgerald &amp; Hartry, 2008)</b>						<b>0.05</b>	<b>+2</b>	<b>Not statistically significant</b>
<b>Kim et al. (2010)<sup>b</sup></b>								
<i>Test of Word Reading Efficiency Total Score</i>	Full sample	3 schools/ 264 students	96.46 (13.70)	96.88 (14.34)	-0.42	-0.03	-1	> .05
<b>Domain average for alphabetics (Kim et al., 2010)</b>						<b>-0.03</b>	<b>-1</b>	<b>Not statistically significant</b>
<b>Domain average for alphabetics across all studies</b>						<b>0.01</b>	<b>0</b>	<b>na</b>

**Table Notes:** For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable. nr = not reported.

<sup>a</sup> For Fitzgerald and Hartry (2008), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-value presented here was reported in the original study. The WWC calculated the intervention group mean for Cohort 1 by adding the regression coefficient (presented in the mean difference column) to the unadjusted comparison group posttest mean. The intervention and comparison group means and standard deviations for Cohort 2 were not reported in the original study, but author-reported effect sizes matched the WWC's calculations. The mean difference reflects the regression coefficient for the impact estimate. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

<sup>b</sup> For Kim et al. (2010), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-value presented here was reported in the original study. The intervention and comparison group means reported in this table are analysis of covariance-adjusted, as reported by the authors in response to a query from the WWC. This study is characterized as having an indeterminate effect because the mean effect reported is neither statistically significant nor substantively important, after correcting for multiple comparisons. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix D.1: Description of supplemental findings for the comprehension domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Fitzgerald &amp; Hartry (2008)<sup>a</sup></b>								
<i>Stanford 10 Reading Comprehension</i>	Cohort 1, Grade 4	108 students	622.32 (28.09)	623.20 (28.02)	-0.88	-0.03	-1	> .05
<i>Stanford 10 Vocabulary</i>	Cohort 1, Grade 4	108 students	620.15 (31.20)	621.24 (38.14)	-1.09	-0.03	-1	> .05
<i>Stanford 10 Reading Comprehension</i>	Cohort 1, Grade 5	132 students	644.34 (29.99)	627.20 (29.86)	17.14	0.57	+22	> .05
<i>Stanford 10 Vocabulary</i>	Cohort 1, Grade 5	132 students	651.04 (34.57)	634.91 (36.47)	16.13	0.45	+17	> .05
<i>Stanford 10 Reading Comprehension</i>	Cohorts 1 & 2, Year 2	294 students	nr	nr	1.58	0.04	+2	.60
<i>Stanford 10 Vocabulary</i>	Cohorts 1 & 2, Year 2	293 students	nr	nr	-0.56	-0.01	0	.88
<b>Kim et al. (2010)<sup>b</sup></b>								
<i>Group Reading Assessment and Diagnostic Evaluation (GRADE) Comprehension</i>	Full sample	264 students	92.95 (13.61)	92.06 (12.29)	0.89	0.07	+3	> .05
<i>GRADE Vocabulary</i>	Full sample	264 students	92.89 (13.20)	92.77 (13.33)	0.12	0.01	0	> .05
<b>Interactive, Inc. (2002)<sup>c</sup></b>								
<i>Stanford Achievement Test, Ninth Edition (Stanford 9) Total Reading</i>	Dallas, grade 8	243 students	648.27 (21.69)	641.40 (33.05)	6.87	0.24	+9	< .01
<b>Meisch et al. (2011)<sup>d</sup></b>								
<b>1 Year of Exposure</b>								
<i>Stanford 10 Reading Comprehension</i>	Full sample	2,555 students	610.24 (27.95)	609.11 (27.98)	1.13	0.04	+2	.34
<i>Stanford 10 Vocabulary</i>	Full sample	2,555 students	614.76 (29.98)	613.37 (31.65)	1.39	0.05	+2	.32
<i>Stanford 10 Reading Comprehension</i>	African-American students	1,445 students	610.26 (27.86)	607.77 (27.23)	2.49	0.09	+4	.29
<i>Stanford 10 Vocabulary</i>	African-American students	1,445 students	615.52 (30.16)	614.22 (32.42)	1.30	0.04	+2	.49
<i>Stanford 10 Reading Comprehension</i>	Hispanic students	1,061 students	612.64 (28.20)	611.53 (29.13)	1.11	0.04	+2	.51
<i>Stanford 10 Vocabulary</i>	Hispanic students	1,061 students	615.33 (29.71)	612.51 (30.82)	2.82	0.09	+4	.15
<i>Stanford 10 Reading Comprehension</i>	Male students	1,479 students	607.93 (27.56)	606.83 (29.30)	1.10	0.04	+2	.46

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<i>Stanford 10 Vocabulary</i>	Male students	1,479 students	615.91 (29.59)	613.60 (32.13)	2.31	0.07	+3	.09
<i>Stanford 10 Reading Comprehension</i>	Female students	1,075 students	614.00 (28.19)	612.05 (25.99)	1.95	0.07	+3	.19
<i>Stanford 10 Vocabulary</i>	Female students	1,075 students	613.77 (30.53)	612.44 (30.94)	1.33	0.04	+2	.61
<b>2 Years of Exposure</b>								
<i>Stanford 10 Reading Comprehension</i>	Full sample	1,520 students	624.44 (25.33)	620.85 (26.24)	3.59	0.14	+6	.02
<i>Stanford 10 Vocabulary</i>	Full sample	1,520 students	629.83 (26.82)	628.20 (27.16)	1.63	0.06	+2	.18
<i>Stanford 10 Reading Comprehension</i>	African-American students	827 students	625.28 (24.43)	621.30 (26.59)	3.98	0.16	+6	.05
<i>Stanford 10 Vocabulary</i>	African-American students	827 students	631.07 (26.35)	629.77 (26.86)	1.30	0.05	+2	.33
<i>Stanford 10 Reading Comprehension</i>	Hispanic students	657 students	623.43 (25.83)	621.54 (25.92)	1.89	0.07	+3	.34
<i>Stanford 10 Vocabulary</i>	Hispanic students	657 students	630.89 (27.48)	625.89 (27.85)	5.00	0.18	+7	.22
<i>Stanford 10 Reading Comprehension</i>	Male students	854 students	622.40 (25.39)	617.19 (25.09)	5.21	0.21	+8	< .01
<i>Stanford 10 Vocabulary</i>	Male students	854 students	629.57 (28.94)	626.69 (28.09)	2.88	0.10	+4	.19
<i>Stanford 10 Reading Comprehension</i>	Female students	665 students	626.81 (25.07)	625.73 (26.95)	1.08	0.04	+2	.47
<i>Stanford 10 Vocabulary</i>	Female students	665 students	630.63 (23.55)	630.00 (26.04)	0.63	0.03	+1	.47
<b>3 Years of Exposure</b>								
<i>Stanford 10 Reading Comprehension</i>	African-American students	550 students	640.80 (24.06)	638.14 (25.10)	2.66	0.11	+4	.28
<i>Stanford 10 Vocabulary</i>	African-American students	550 students	641.95 (25.09)	640.49 (29.60)	1.46	0.05	+2	.59
<i>Stanford 10 Reading Comprehension</i>	Hispanic students	447 students	644.80 (21.78)	643.60 (22.07)	1.20	0.05	+2	.63
<i>Stanford 10 Vocabulary</i>	Hispanic students	447 students	645.86 (27.21)	646.60 (26.22)	-0.74	-0.03	-1	.89
<i>Stanford 10 Reading Comprehension</i>	Male students	587 students	641.26 (22.37)	638.07 (24.51)	3.19	0.14	+5	.13



Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<i>Stanford 10 Vocabulary</i>	Male students	587 students	643.75 (27.19)	641.01 (31.47)	2.74	0.09	+4	.34
<i>Stanford 10 Reading Comprehension</i>	Female students	436 students	642.36 (23.47)	643.80 (23.08)	-1.44	-0.06	-2	.50
<i>Stanford 10 Vocabulary</i>	Female students	436 students	642.11 (23.99)	641.87 (23.81)	0.24	0.01	0	.92
<b>Yurchak (2013)<sup>e</sup></b>								
<i>New Jersey High School Proficiency Assessment (HSPA) Analyzing Text Cluster Score</i>	African-American students	23 students	37.00 (9.60)	37.80 (14.20)	-0.80	-0.07	-3	nr
<i>HSPA Analyzing Text Cluster Score</i>	Female students	61 students	40.15 (11.20)	39.80 (10.80)	0.35	0.04	+1	nr
<i>HSPA Analyzing Text Cluster Score</i>	Male students	73 students	37.82 (10.30)	38.90 (10.50)	-1.08	-0.11	-4	nr
<i>HSPA Reading Cluster Score</i>	African-American students	23 students	40.10 (9.70)	40.30 (13.70)	-0.20	-0.02	-1	nr
<i>HSPA Reading Cluster Score</i>	Female students	61 students	41.05 (10.50)	42.70 (11.40)	-1.65	-0.14	-6	nr
<i>HSPA Reading Cluster Score</i>	Male students	73 students	41.32 (11.30)	42.70 (10.70)	-1.38	-0.13	-5	nr

**Table Notes:** The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported.

<sup>a</sup> For Fitzgerald and Hartry (2008), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-values presented here were reported in the original study. The WWC calculated the intervention group mean for Cohort 1 subgroup analyses (grades 4 and 5) by adding the regression coefficient (presented in the mean difference column) to the unadjusted comparison group posttest mean. The intervention and comparison group means and standard deviations for Cohorts 1 & 2, year 2 were not reported in the original study, but author-reported effect sizes matched the WWC's calculations. The mean difference reflects the regression coefficient for the impact estimate.

<sup>b</sup> For Kim et al. (2010), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-values presented here were reported in the original study. The intervention and comparison group means reported in this table are analysis of covariance (ANCOVA)-adjusted, as reported by the authors in response to a query from the WWC.

<sup>c</sup> For Interactive, Inc. (2002), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-value presented here was reported in the original study. The intervention and comparison group means reported in this table are ANCOVA-adjusted.

<sup>d</sup> For Meisch et al. (2011), corrections for clustering and multiple comparisons were needed and resulted in a WWC-computed critical *p*-value of .01 for Stanford 10 Reading Comprehension for all students with 2 years of exposure; therefore, the WWC does not find the result to be statistically significant. These corrections also resulted in a WWC-computed critical *p*-value of .02 for Stanford 10 Reading Comprehension for African-American students with 2 years of exposure; therefore, the WWC does not find the result to be statistically significant as well. The *p*-values presented here were reported in the original study. The intervention and comparison group means reported in this table are regression-adjusted, as reported by the authors in the original report.

<sup>e</sup> For Yurchak (2013), the WWC did not need to make corrections for clustering or multiple comparisons. The WWC calculated the program group mean using a difference-in-differences approach by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. The author did not report *p*-values in the original study, and the WWC-computed *p*-values for all outcomes were not statistically significant.

Appendix D.2: Description of supplemental findings for the general literacy achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Fitzgerald &amp; Hartry (2008)<sup>a</sup></b>								
<i>Stanford Achievement Test, Tenth Edition (Stanford 10) Total Reading</i>	Cohorts 1 & 2, Year 2	291 students	nr	nr	0.39	0.01	0	.87
<b>Swanlund et al. (2012)<sup>b</sup></b>								
<i>Measures of Academic Progress (MAP)</i>	Treatment-on-the-treated (TOT) sample	617 students	nr	nr	2.38	0.18	+7	< .05
<b>Meisch et al. (2011)<sup>c</sup></b>								
<b>1 Year of Exposure</b>								
<i>Stanford 10 Language Arts</i>	Full sample	2,555 students	599.10 (24.91)	598.40 (26.58)	0.70	0.03	+1	.40
<i>Stanford 10 Language Arts</i>	African-American students	1,445 students	599.35 (25.03)	597.63 (25.60)	1.72	0.07	+3	.16
<i>Stanford 10 Language Arts</i>	Hispanic students	1,061 students	599.36 (24.94)	599.61 (28.11)	-0.25	-0.01	0	.83
<i>Stanford 10 Language Arts</i>	Male students	1,479 students	595.12 (24.22)	594.96 (26.31)	0.16	0.01	0	.90
<i>Stanford 10 Language Arts</i>	Female students	1,075 students	605.11 (24.88)	603.00 (26.20)	2.11	0.08	+3	.14
<b>2 Years of Exposure</b>								
<i>Stanford 10 Language Arts</i>	Full sample	1,520 students	611.23 (24.64)	609.12 (25.66)	2.11	0.08	+3	.30
<i>Stanford 10 Language Arts</i>	African-American students	827 students	611.09 (23.19)	608.82 (25.01)	2.27	0.09	+4	.33
<i>Stanford 10 Language Arts</i>	Hispanic students	657 students	612.77 (26.38)	609.28 (26.38)	3.49	0.13	+5	.06
<i>Stanford 10 Language Arts</i>	Male students	854 students	607.02 (23.38)	604.59 (24.44)	2.43	0.10	+4	.33
<i>Stanford 10 Language Arts</i>	Female students	665 students	616.60 (25.35)	616.60 (25.96)	0	0	0	.33
<b>3 Years of Exposure</b>								
<i>Stanford 10 Language Arts</i>	African-American students	550 students	623.17 (24.24)	620.64 (21.78)	2.53	0.11	+4	.17
<i>Stanford 10 Language Arts</i>	Hispanic students	447 students	626.09 (23.92)	625.32 (23.50)	0.77	0.03	+1	.66
<i>Stanford 10 Language Arts</i>	Male students	587 students	619.88 (22.17)	617.57 (22.40)	2.31	0.10	+4	.19

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<i>Stanford 10 Language Arts</i>	Female students	436 students	627.32 (25.52)	626.67 (22.10)	0.65	0.03	+1	.79
<b>Sprague et al. (2012)<sup>d</sup></b>								
<i>Stanford Diagnostic Reading Test, Fourth Edition (SDRT-4)</i>	Cohorts 1–4	364 students	665.41 (48.85)	660.12 (48.16)	5.29	0.11	+4	.03
<i>SDRT-4</i>	Cohorts 1–3	334 students	665.27 (54.50)	659.99 (52.58)	5.28	0.10	+4	.03
<i>SDRT-4</i>	Cohorts 1–2	241 students	664.78 (27.80)	661.94 (25.74)	2.84	0.11	+4	.31
<b>White et al. (2006)<sup>e</sup></b>								
<i>TerraNova Reading Test</i>	Scored below 40 normal curve equivalent (NCE) on pretest	1,268 students	39.80 (8.40)	36.20 (12.20)	3.60	0.34	+13	< .05
<i>TerraNova Reading Test</i>	Scored above 40 NCE on pretest	362 students	46.10 (8.40)	45.60 (9.20)	0.50	0.06	+2	> .05

**Table Notes:** The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual’s percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported.

<sup>a</sup> For Fitzgerald and Hartry (2008), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-value and effect size presented here were reported in the original study. The intervention and comparison group means and standard deviations were not reported in the original study, but author-reported effect sizes matched the WWC’s calculations. The mean difference reflects the regression coefficient for the impact estimate.

<sup>b</sup> For Swanlund et al. (2012), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-value and effect size presented here were reported in the original study. The study findings reflect the TOT sample, defined as students who were assigned to the intervention group who attended *READ 180*® classes. This study met the WWC’s CACE standards, which are available on the WWC’s website. The intent-to-treat (ITT) findings are prioritized over the TOT findings because the ITT analysis addresses the type of research question most commonly posed in this report (i.e., the effects of being assigned to *READ 180*®).

<sup>c</sup> For Meisch et al. (2011), corrections for clustering and multiple comparisons were needed but did not affect whether any of the contrasts were found to be statistically significant. The p-values presented here were reported in the original study. The intervention and comparison group means reported in this table are regression-adjusted, as reported by the authors in the original report.

<sup>d</sup> For Sprague et al. (2012), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-values presented here were reported in the original study. The intervention and comparison group means reported in this table, which are standardized scale scores, are analysis of covariance-adjusted and reported by the authors in the original study. The standard deviations reported in this table for Year 4 data are covariate adjusted. Unadjusted standard deviations, which are used in WWC effect size and statistical significance calculations, were not available.

<sup>e</sup> For White et al. (2006), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-values presented here were reported in the original study.

Appendix D.3: Description of supplemental findings for the reading fluency domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Fitzgerald &amp; Hartry (2008)<sup>a</sup></b>								
<i>Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency</i>	Cohort 1, Grade 4	109 students	105.21 (25.51)	101.13 (25.70)	4.08	0.16	+6	> .05
<i>DIBELS Oral Reading Fluency</i>	Cohort 1, Grade 5	132 students	110.76 (27.55)	108.67 (20.40)	2.09	0.09	+3	> .05
<b>Kim et al. (2010)<sup>b</sup></b>								
<i>DIBELS Oral Reading Fluency</i>	Grade 4	93 students	88.41 (33.35)	77.68 (28.30)	10.73	0.35	+14	< .01
<i>DIBELS Oral Reading Fluency</i>	Grade 5	100 students	113.85 (25.48)	118.51 (32.67)	-4.66	-0.16	-6	> .05
<i>DIBELS Oral Reading Fluency</i>	Grade 6	71 students	133.48 (32.01)	129.50 (29.51)	3.98	0.13	+5	> .05

**Table Notes:** The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

<sup>a</sup> For Fitzgerald and Hartry (2008), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-values presented here were reported in the original study. The WWC calculated the intervention group mean by adding the regression coefficient (presented in the mean difference column) to the unadjusted comparison group posttest mean.

<sup>b</sup> For Kim et al. (2010), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. The WWC did not need to make corrections for clustering or to adjust for baseline differences. The p-values presented here were reported in the original study. The intervention and comparison group means reported in this table are analysis of covariance-adjusted, as reported by the authors in response to a query from the WWC.

Appendix D.4: Description of supplemental findings for the alphabetic domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Fitzgerald &amp; Hartry (2008)<sup>a</sup></b>								
<i>Stanford Achievement Test, Tenth Edition (Stanford 10) Spelling</i>	Cohort 1, Grade 4	107 students	619.81 (32.59)	613.45 (42.85)	6.36	0.17	+7	> .05
<i>Stanford 10 Spelling</i>	Cohort 1, Grade 5	132 students	637.20 (29.63)	634.14 (35.61)	3.06	0.09	+4	> .05
<i>Stanford 10 Spelling</i>	Cohorts 1 & 2, Year 2	292 students	nr	nr	-0.33	-0.01	0	.92
<b>Kim et al. (2010)<sup>b</sup></b>								
<i>Test of Word Reading Efficiency (TOWRE) Sight Word Reading</i>	Full sample	264 students	96.62 (10.62)	97.40 (11.25)	-0.78	-0.07	-3	.17
<i>TOWRE Phonetic Decoding</i>	Full sample	264 students	96.48 (14.08)	97.38 (14.62)	-0.90	-0.06	-2	> .05

**Table Notes:** The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported.

<sup>a</sup> For Fitzgerald and Hartry (2008), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-values presented here were reported in the original study. The WWC calculated the intervention group mean for Cohort 1 subgroup analyses (grades 4 and 5) by adding the regression coefficient (presented in the mean difference column) to the unadjusted comparison group posttest mean. The intervention and comparison group means and standard deviations for Cohorts 1 & 2, Year 2 were not reported in the original study, but author-reported effect sizes matched the WWC's calculations.

<sup>b</sup> For Kim et al. (2010), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-values presented here were reported in the original study. The intervention and comparison group means reported in this table are analysis of covariance-adjusted, as reported by the authors in response to a query from the WWC.

### Endnotes

<sup>1</sup> The descriptive information for this program was obtained from a publicly available source: the program's website (<http://www.hmhco.com/products/read-180/>; accessed September 22, 2016). The WWC requests distributors review the program description sections for accuracy from their perspective. The program description was provided to the distributor in September 2014, and the WWC incorporated feedback from the distributor. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.

<sup>2</sup> The literature search reflects documents publicly available by November 2015. This report has been updated to include reviews of 71 studies that were not included in the previous intervention report that was released in 2009. Of the additional studies, 49 were not within the scope of the review protocol for the Adolescent Literacy topic area, and 16 were within the scope of the review protocol for the Adolescent Literacy topic area but did not meet WWC group design standards. A complete list and disposition of all studies reviewed are provided in the references. This report includes reviews of all previous studies that met WWC group design standards with or without reservations and resulted in a revised disposition of four studies:

- (1) Haslam, White, & Klinge (2006) received a disposition in this report of *ineligible for review*, where it had previously received the rating of *meets WWC evidence standards with reservations*: the study was previously reviewed under the Adolescent Literacy protocol (version 1.0), and is currently reviewed using the Adolescent Literacy protocol (version 3.0), which identifies studies in which the majority of the study sample was identified as English learners as ineligible for review;
- (2) Lang, Torgesen, Petscher, Vogel, Chanter, & Lefsky (2008) received a disposition in this report of *does not meet WWC group design standards*, where it had previously received the rating of *meets WWC evidence standards with reservations*: the study was previously reviewed using version 1.0 standards, and is currently reviewed using version 3.0 standards which include a clarification in guidance that imputed data cannot be used to demonstrate equivalence of the analytic sample—the author did not respond to the WWC's request for data that could be used to demonstrate equivalence, so it is now rated *does not meet WWC group design standards*;
- (3) Scholastic Research (2008) received a disposition in this report of *ineligible for review*, where it had previously received the rating of *meets WWC evidence standards with reservations*: the study was previously reviewed under the Adolescent Literacy protocol (version 1.0), and is currently reviewed using the Adolescent Literacy protocol (version 3.0), which identifies studies in which the majority of the study sample was identified as English learners as ineligible for review; and
- (4) Woods (2007) received a disposition in this report of *does not meet WWC group design standards*, where it had previously received the rating of *meets WWC evidence standards with reservations*: the study was previously reviewed using version 1.0 standards, and is currently reviewed using version 3.0 standards which include a clarification in guidance that baseline differences of more than .05 SD require a statistical adjustment for pretest differences—the author did not adjust for pretest differences for the 2003–04 cohort so it is now rated *does not meet WWC group design standards*. Both the 2004–05 and 2005–06 cohorts received a disposition of *does not meet WWC group design standards* in the previous and current report because the study included one teacher in the *READ 180*<sup>®</sup> group in each cohort, which is a confounding factor because it is not possible to tell whether the *READ 180*<sup>®</sup> intervention or the teacher is responsible for the difference in outcomes.

The studies in this report were reviewed using the standards from the WWC Procedures and Standards Handbook (version 3.0) and the Adolescent Literacy review protocol (version 3.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

<sup>3</sup> For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 56. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the studies.

<sup>4</sup> The studies reviewed by the WWC do not include evaluations of the two most recent versions of the intervention: *READ 180*<sup>®</sup> *Next Generation* (2011) and *READ 180*<sup>®</sup> *Universal* (2016).

<sup>5</sup> In the previous intervention report, findings from the Dallas and Houston samples were presented separately for Stanford 9 Reading Comprehension measures. The study was previously reviewed using version 1.0 standards, and is currently reviewed using version 3.0 standards which include updated baseline equivalence standards. Findings from the Boston sample were excluded, since they did not meet the WWC's baseline equivalence standards in place at that time. In the present report, we combined the Boston, Dallas, and Houston subsamples, which pooled together, meet the WWC's baseline equivalence standards so these findings are now rated as *meets WWC group design standards with reservations*. When samples are assessed individually, however, the Boston and Houston samples do not meet WWC version 3.0 baseline equivalence standards, while the Dallas sample does.

<sup>6</sup> White et al. (2006) was previously reviewed under the Adolescent Literacy protocol (version 1.0), and is currently reviewed using the Adolescent Literacy protocol (version 3.0), which identifies studies in which the majority of the study sample was identified as English

learners as ineligible for review. In the previous intervention report, findings in the reading comprehension domain for Cohort 1 were presented; however, findings from this cohort have been determined by the WWC to be ineligible for review, since the sample includes 53% English learners.

<sup>7</sup> Some findings White et al. (2005) are not included in this intervention report, but were included in the previous report, although the study's disposition is unchanged. The study was previously reviewed using version 1.0 standards, and is currently reviewed using version 3.0 standards which include updated baseline equivalence standards. Findings reported in the study by combinations of grade and proficiency level did not demonstrate baseline equivalence under the version 3.0 standards, and so those findings are now rated *does not meet WWC group design standards*. However, three subgroup analyses in this study did demonstrate equivalence, and so the study receives the same rating of *meets WWC group design standards with reservations*.

<sup>8</sup> Kim et al. (2011) present treatment-on-the-treated (TOT) estimates of *READ 180*<sup>®</sup> impact on alphabets, comprehension, and reading fluency outcomes. While the underlying standardized outcomes meet WWC standards, the analysis that produced these estimates is not eligible for review under the WWC complier average causal effect (CACE) guidance. The authors used a two-stage least-squares estimation, using intervention receipt as the endogenous independent variable and assignment status as the instrumental variable. However, the authors used a continuous variable for intervention receipt: the number of days receiving *READ 180*<sup>®</sup>. The CACE guidance requires a dichotomous indicator for intervention receipt, so this analysis is not eligible for review.

### Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, November). *Adolescent Literacy intervention report: READ 180*<sup>®</sup>. Retrieved from <http://whatworks.ed.gov>

## WWC Rating Criteria

### Criteria used to determine the rating of a study

Study rating	Criteria
<b>Meets WWC group design standards without reservations</b>	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
<b>Meets WWC group design standards with reservations</b>	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

### Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
<b>Positive effects</b>	Two or more studies show statistically significant positive effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
<b>Potentially positive effects</b>	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
<b>Mixed effects</b>	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
<b>Potentially negative effects</b>	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
<b>Negative effects</b>	Two or more studies show statistically significant negative effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
<b>No discernible effects</b>	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

### Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
<b>Medium to large</b>	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
<b>Small</b>	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.



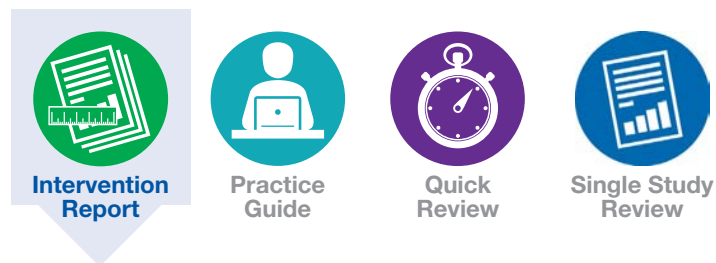
### Glossary of Terms

<b>Attrition</b>	Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.
<b>Clustering adjustment</b>	If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.
<b>Confounding factor</b>	A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.
<b>Design</b>	The design of a study is the method by which intervention and comparison groups were assigned.
<b>Domain</b>	A domain is a group of closely related outcomes.
<b>Effect size</b>	The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.
<b>Eligibility</b>	A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.
<b>Equivalence</b>	A demonstration that the analytic sample groups are similar on observed characteristics defined in the review area protocol.
<b>Extent of evidence</b>	An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 56.
<b>Improvement index</b>	Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from -50 to +50.
<b>Intervention</b>	An educational program, product, practice, or policy aimed at improving student outcomes.
<b>Intervention report</b>	A summary of the findings of the highest-quality research on a given program, product, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against design standards, and summarizes the findings of those that meet WWC design standards.
<b>Multiple comparison adjustment</b>	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
<b>Quasi-experimental design (QED)</b>	A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.
<b>Randomized controlled trial (RCT)</b>	A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.
<b>Rating of effectiveness</b>	The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 56.
<b>Single-case design</b>	A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

### Glossary of Terms

- Standard deviation** The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.
- Statistical significance** Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ( $p < .05$ ).
- Substantively important** A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.
- Systematic review** A review of existing literature on a topic that is identified and reviewed using explicit methods. A WWC systematic review has five steps: 1) developing a review protocol; 2) searching the literature; 3) reviewing studies, including screening studies for eligibility, reviewing the methodological quality of each study, and reporting on high quality studies and their findings; 4) combining findings within and across studies; and, 5) summarizing the review.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.



An **intervention report** summarizes the findings of high-quality research on a given program, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against evidence standards, and summarizes the findings of those that meet standards.

This intervention report was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.