

Appendix

Appendix A1.1 Study characteristics: Haslam, White, & Klinge, 2006 (quasi-experimental design)

Characteristic	Description
Study citation	Haslam, M. B., White, R. N., & Klinge, A. (2006). <i>Improving student literacy: READ 180 in the Austin Independent School District, 2004–05</i> . Washington, DC: Policy Studies Associates.
Participants	From the initial pool of 409 <i>READ 180</i> students in grades 7 and 8 who scored at least one reading level below grade level, 307 students were matched to 307 comparison students using a one-to-one propensity score matching method. ¹ Although the percentage of Limited English Proficiency students turned out to be significantly higher in the <i>READ 180</i> group (89%) than in the comparison group (73%), the groups were equivalent on the pretest achievement measure. In all, 307 students in the <i>READ 180</i> group and 307 students in the comparison group were included in the analysis sample.
Setting	The study took place in seventh- and eighth-grade classrooms in the Austin Independent School District (AISD) in Texas.
Intervention	Data on students' exposure to the <i>READ 180</i> software were not provided in this study. The study reported student outcomes after one year of program implementation.
Comparison	The comparison group received the standard instruction provided in the regular school curriculum.
Primary outcomes and measurement	For both the pretest and posttest, students took the English-language version of the 2004 Texas Assessment of Knowledge and Skills (TAKS) Reading Test. For a more detailed description of this outcome measure, see Appendix A2.1.
Staff/teacher training	No information on training for teachers or staff was provided in this study.

1. One hundred two *READ 180* students were not included in the matching procedure because data were missing for one or more of the categories used for matching.

Appendix A1.2 Study characteristics: Interactive Inc., 2002 (quasi-experimental design)

Characteristic	Description
Study citation	Interactive Inc. (2002). <i>An efficacy study of READ 180, a print and electronic adaptive intervention program, grades 4 and above</i> . New York, NY: Scholastic Inc.
Participants	The study took place in seven districts across the United States. ¹ Each district agreed to recruit two middle schools for the study. Each middle school was to establish two <i>READ 180</i> classrooms and two comparison classrooms. In addition, each school was to rank its students by reading ability. The lowest-ranking 320 students were to be randomly assigned to a <i>READ 180</i> class, the comparison group, or a backup group. None of the districts ultimately followed through with this research design, ² but pretest equivalence on a reading measure was established for the analysis sample in three school districts. The analysis sample consisted of 52 comparison students and 119 students enrolled in <i>READ 180</i> in Columbus, 142 comparison students and 101 students enrolled in <i>READ 180</i> in Dallas, and 36 comparison students and 59 students enrolled in <i>READ 180</i> in Houston.
Setting	The analysis sample was located in three districts: five schools in Columbus, Ohio; four schools in Dallas, Texas; and two schools in Houston, Texas.
Intervention	The intervention group received the <i>READ 180</i> intervention during a 90-minute literacy block. During that block, small classes of 15–18 students spent the first 10 minutes together with the teacher doing language-arts instruction. Over the next hour, the class broke into three smaller groups and cycled through three 20-minute rotations as follows: small-group instruction, independent reading, and direct instruction. There was some deviation from the intervention design across schools. The study reported students' outcomes after one year of program implementation.
Comparison	The comparison group received the standard instruction provided in the regular school curriculum. ³
Primary outcomes and measurement	In Houston and Dallas, the Total Reading score from the Stanford Achievement Test (SAT-9) was used as the pretest and posttest measure. In Columbus, the Reading Comprehension subtest score from the SAT-9 was used as the pretest and posttest measure. For a more detailed description of these outcome measures, see Appendix A2.1.
Staff/teacher training	No information on training for teachers or staff was provided in this study.

1. Test scores were not available for Atlanta and San Francisco, and the Miami-Dade district did not provide the correct form of test scores. Therefore, the findings from these locations are not included in this report. In addition, the intervention and comparison groups in Boston and in the grade 7 sample in Houston were not shown to be equivalent at baseline, so they were excluded from the review.
2. The random assignment of students was violated. A number of schools decided that there were students for whom assignment to the *READ 180* program would be most beneficial. Only after these students were assigned to *READ 180* was a comparison group identified. Individual parents or caregivers were allowed to request inclusion or exclusion from the program. Students were allowed to decline participation in *READ 180*. No students with a reading grade equivalent lower than grade 1.5 were allowed to be placed in *READ 180* classes.
3. All students assigned to the comparison group did not experience the same literacy instruction. Comparison students within the same district or same school were often exposed to different curricula. The authors acknowledge that the realities of local control confounded their ability to completely understand the curricular and instructional practices to which the comparison groups were exposed.

Appendix A1.3 Study characteristics: Lang et al., 2008 (randomized controlled trial where differential attrition could not be ruled out)

Characteristic	Description
Study citation	Lang, L. H., Torgesen, J. K., Petscher, Y., Vogel, W., Chanter, C., & Lefsky, E. (2008, March). <i>Exploring the relative effectiveness of reading interventions for high school students</i> . Paper presented at the annual research conference of the Society for Research on Educational Effectiveness, Crystal City, VA.
Participants	A total of 1,265 ninth-grade students in 87 classrooms were identified as struggling readers (at high or moderate risk) based on prior-year reading performance on the Florida Comprehensive Assessment Test (FCAT). Students scoring in the high-risk or moderate-risk categories were randomly assigned to one of three treatment conditions— <i>REACH</i> , <i>RISE</i> , or <i>READ 180</i> —or to a control condition— <i>School Offered Accelerated Reading (SOAR)</i> . After multiple imputation and removing 68 outliers, the analysis sample across all conditions was reduced to 1,197 participants. For this review, the analysis sample consisted of 100 high-risk students who received <i>READ 180</i> and 90 high-risk students in the comparison group, as well as 207 moderate-risk students who received <i>READ 180</i> and 202 moderate-risk students in the comparison group.
Setting	The study included seven comprehensive high schools in a large Florida school district.
Intervention	The intervention group received an intensive reading program for 90 minutes per day. The program, which is a combination of instructional, modeled, and independent reading components, begins with 20 minutes of teacher-led, whole-group instruction followed by three 20-minute rotations. The rotations last for a total of 60 minutes and include small-group direct instruction, use of <i>READ 180</i> software, and independent and modeled reading. Once all rotations are complete, the class convenes for 10 minutes of whole-group wrap-up. The study reported students' outcomes after one year of program implementation.
Comparison	Students in the comparison group received the district's standard curriculum: <i>SOAR</i> . The implementation of <i>SOAR</i> involved the following materials: the <i>Reading and Writing Sourcebook</i> by Great Source, the <i>Reader's Handbook</i> by Great Source, <i>Reading Nonfiction</i> by Jamestown, and the <i>Daybook of Critical Reading and Writing</i> by Great Source. The <i>SOAR</i> classes typically included FCAT-preparatory activities aligned with the Sunshine State Standards and Benchmarks that were available to all students through a software program called FCAT Explorer. This type of practice provided students opportunities to answer questions based on the types of text (70% informational and 30% literary) and length of passages (range of words, 300–1400; average number of words, 800) that they would encounter on the ninth-grade test (Florida Department of Education, 2007). ¹
Primary outcomes and measurement	For both the pretest and the posttest, students took the Florida Comprehensive Assessment Test–Sunshine State Standards (FCAT-SSS). For a more detailed description of this outcome measure, see Appendix A2.1.
Staff/teacher training	School leaders identified teachers to deliver the <i>READ 180</i> and <i>SOAR</i> interventions. Both <i>READ 180</i> teachers and <i>SOAR</i> teachers received coaching and feedback related to fidelity and quality of implementation from two sources: the project coordinator and the school-level reading coach assigned to each school. Professional development continued throughout the year for both <i>READ 180</i> and <i>SOAR</i> teachers, and intervention-specific monthly support meetings were held to address concerns. The publisher of the <i>READ 180</i> intervention was asked to participate in the provision of materials, the conduct of professional development for <i>READ 180</i> teachers and school leaders, and the development of fidelity of implementation checklists.

1. Florida Department of Education. (2007). FCAT Explorer. Retrieved January 6, 2007, from <http://www.fcatexplorer.com/>.

Appendix A1.4 Study characteristics: Scholastic Research, 2008 (quasi-experimental design)

Characteristic	Description
Study citation	Scholastic Research. (2008). <i>Desert Sands Unified School District, CA</i> . New York, NY: Scholastic Inc.
Participants	Two hundred eighty-five students in grades 6, 7, and 9 who scored at the below-basic or basic performance level on the Spring 2006 California Standards Test, English Language Arts (CST-ELA), and who were identified as struggling readers received the <i>READ 180</i> intervention. More than half of the students (58%) were classified as English language learners (ELL). Within each grade level, a one-to-one matching procedure based on pretest reading scores was used to select students for the comparison group. In all, 285 students in the <i>READ 180</i> group and 285 students in the comparison group were included in the analysis sample. ¹
Setting	The study was conducted in the Desert Sands Unified School District in California.
Intervention	The intervention group used <i>READ 180</i> as a core English Language Arts curriculum replacement for two periods, which was a total of 90 minutes per day. The study reported students' outcomes after the first year of program implementation.
Comparison	The comparison group received the regular reading curriculum. Students in grades 6 and 7 used the Holt Literature and Language Arts curriculum. Students in grade 9 used the Prentice Hall Literature curriculum. No comparison-group students received any additional reading-comprehension instruction other than what a teacher would choose to use in the publisher's materials.
Primary outcomes and measurement	For both the pretest and the posttest, students took the California Standards Test, English Language Arts (CST-ELA). For a more detailed description of this outcome measure, see Appendix A2.2.
Staff/teacher training	No information on training for teachers or staff was provided in this study.

1. Results from a subset of ELL students were also reported but were not included in this report because the population of ELL students was outside the scope of the Adolescent Literacy review.

Appendix A1.5 Study characteristics: White, Haslam, & Hewes, 2006 (quasi-experimental design)

Characteristic	Description
Study citation	White, R. N., Haslam, M. B., & Hewes, G. M. (2006). <i>Improving student literacy: READ 180 in the Phoenix Union High School District, 2003–04 and 2004–05</i> . Washington, DC: Policy Studies Associates.
Participants	Three cohorts of ninth-grade students who were reading one or more years below grade level participated in <i>READ 180</i> in 12 schools. For cohort 1, a propensity score matching procedure was used to identify the subset of nonparticipants whose reading level and English language learner (ELL) eligibility were similar to those of students in the treatment group. For cohorts 2 and 3, a propensity score matching procedure was conducted to identify the comparison group; it was based on eighth-grade reading proficiency, ELL status, special-education eligibility, gender, and ethnicity. The cohort 1 analysis sample included 826 intervention students who received <i>READ 180</i> in 2003–04 and 826 matched nonparticipants. The cohort 2 analysis sample consisted of 815 students who received <i>READ 180</i> in 2004–05 and 815 matched nonparticipants. The cohort 3 analysis sample consisted of 1,029 students who received <i>READ 180</i> in 2005–06 and 1,029 matched nonparticipants. The study reported students' outcomes for all three cohorts after one year of program implementation; these findings can be found in Appendices A3.2 (cohort 1) and A3.3 (cohort 2 and cohort 3). Additional findings reflecting cohort 1 students' outcomes two years after the start of the implementation of the intervention can be found in Appendix A4.2 (for at least some students, these findings reflect an additional semester of exposure to the intervention).
Setting	The study took place in an urban school district in Phoenix, Arizona.
Intervention	The intervention group received <i>READ 180</i> , stage C, version 1.6. The study reported students' outcomes after one year of program implementation.
Comparison	The comparison group received the standard instruction provided in the regular school curriculum.
Primary outcomes and measurement	For cohort 1 and cohort 2, the authors used the Reading Comprehension subtest from the Stanford Achievement Test (SAT-9) as the pretest measure. For cohort 3, TerraNova reading scores were used as the pretest measure. For cohort 1, the SAT-9 Reading Comprehension subtest was used as the posttest, and the Reading Score on the Arizona Instrument to Measure Standards (AIMS) was used as the second-year posttest. For cohort 2 and cohort 3, TerraNova reading scores were used as the posttest measure. For a more detailed description of these outcome measures, see Appendices A2.1–A2.2.
Staff/teacher training	No information on training for teachers or staff was provided in this study.

Appendix A1.6 Study characteristics: White, Williams, & Haslem, 2005 (quasi-experimental design)

Characteristic	Description
Study citation	White, R. N., Williams, I. J., & Haslem, M. B. (2005). <i>Performance of District 23 students participating in Scholastic READ 180</i> . Washington, DC: Policy Studies Associates.
Participants	The authors compared English Language Arts test outcomes for <i>READ 180</i> students in 16 schools to outcomes of their peers attending the same schools who did not participate in <i>READ 180</i> . For the overall sample of students in grades 4–8, the profile of the students selected to participate in <i>READ 180</i> was similar to that of comparison students. About 85% of students were African-American, and 90% were eligible for free or reduced-price lunch. <i>READ 180</i> students were somewhat less likely to be eligible for special-education services than nonparticipating students (6% versus 11%). For grades 4, 6, and 8, the students in the two groups were similar on the reading pretest. The analysis sample consisted of 362 students in the <i>READ 180</i> group and 2,528 students in the comparison group across grades 4, 6, and 8. Comparisons were made between students with the same proficiency levels (1, 2, and 3) within each grade. ¹
Setting	The study was conducted in 16 public schools in central Brooklyn in New York City.
Intervention	The intervention group received <i>READ 180</i> during the 2001–02 academic year. The study reported students' outcomes after one year of program implementation.
Comparison	The comparison group received the standard instruction provided in the regular school curriculum.
Primary outcomes and measurement	For the pretest, students took a reading test developed by CTB/McGraw-Hill for the city of New York. This CTB/McGraw-Hill Reading 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 grades 4 and 8 took the New York State Department of Education End-of-Year Test in English Language Arts (NYSDE/ELA), and students in grade 6 took the CTB/McGraw-Hill Reading Test developed for the city of New York. For a more detailed description of these outcome measures, see Appendix A2.1.
Staff/teacher training	No information on training for teachers or staff was provided in this study.

1. There were no treatment students in the grade 7 analysis sample; therefore, grade 7 students were excluded from the review. There were only two treatment students in the proficiency level 4 across grades 4–8; therefore, proficiency level 4 was excluded from the review. The intervention and comparison groups in grade 5 (proficiency levels 1, 2, and 3) were not shown to be equivalent at baseline and were excluded from the review.

Appendix A1.7 Study characteristics: Woods, 2007 (quasi-experimental design)

Characteristic	Description
Study citation	Woods, D. E. (2007). <i>An investigation of the effects of a middle school reading intervention on school dropout rates</i> . Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg.
Participants	Three annual cohorts of middle-school students participated in <i>READ 180</i> from 2003 to 2006. ¹ Based on reading pretest scores and teacher recommendations, the school guidance counselor assigned students in grades 6, 7, and 8 to either the computer-based <i>READ 180</i> program or the school's traditional reading-remediation program. ² In total, the 2003–04 school year analysis sample included 58 students who participated in <i>READ 180</i> and 58 students who were in the comparison group. Additional findings reflecting students' outcomes by grade and ethnicity can be found in Appendix A4.1.
Setting	This study took place in an urban middle school in southeastern Virginia.
Intervention	The intervention group participated in <i>READ 180</i> every other day for 90 minutes for the entire school year, in addition to a daily 55-minute language-arts class and 20 minutes of sustained silent reading. Because of technical problems during the first year, the fidelity of <i>READ 180</i> program implementation was downgraded from Level One (the highest level of fidelity) to Level Two, according to the <i>READ 180</i> Research Protocol and Tools (Scholastic, Inc., 2004). ³ All implementation indicators were met, with the exception of a daily class schedule of 90-minute blocks five days a week. The study reported students' outcomes after the first year of program implementation.
Comparison	The comparison group received 90 minutes of remedial reading every other day for one quarter of the school year. The traditional reading remediation program provided focused, skill-based instruction and opportunities to integrate writing and thinking skills. In addition, comparison students participated in 20 minutes of sustained silent reading and 55 minutes of daily language-arts instruction.
Primary outcomes and measurement	For both pretests and posttests, the author used the Degrees of Reading Power (DRP) test. For a more detailed description of this outcome measure, see Appendix A2.1. The Standardized Test for Assessment of Reading (STAR) and the Scholastic Reading Inventory were also used in the study for the 2004–05 and 2005–06 cohorts of students that were not included in this report. ¹
Staff/teacher training	<i>READ 180</i> teachers, all of whom were licensed reading specialists, received comprehensive instructional materials, professional development support, and training in best teaching practices. Comparison-group teachers, all of whom were licensed reading specialists, received a limited professional-development component. No additional details on the professional development provided to comparison group teachers were provided.

1. The 2004–05 and 2005–06 student cohorts do not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one *READ 180* teacher in the treatment condition in both cohorts. This information was not reported in Woods (2007), but was provided to the WWC by the author.
2. The grade 8 cohort does not meet WWC evidence standards because the intervention and comparison groups were not shown to be equivalent at baseline.
3. Scholastic Inc. (Ed.). (2004). *READ 180 research protocol and tools*. New York, NY: Scholastic Inc.

Appendix A2.1 Outcome measures for the comprehension domain

Outcome measure	Description
Reading comprehension construct	
Arizona Instrument to Measure Standards (AIMS) Reading Test	This standardized test assesses students' ability to understand, interpret, and analyze what they have read. The test consists of approximately 60 multiple-choice items (as cited in White, Haslam, & Hewes, 2006; http://www.ade.state.az.us/standards/AIMS/AIMSSTGuides/).
CTB/McGraw-Hill Reading Test	This standardized reading test was developed by CTB/McGraw-Hill for the city of New York. The test produces scores that can be aligned with and compared to the New York State Department of Education End-of-Year Tests, which are also published by CTB/McGraw-Hill (as cited in White, Williams, & Haslem, 2005) and are summarized in this table (see below).
Degrees of Reading Power (DRP) test	The Degrees of Reading Power (DRP) test is a criterion-referenced test to assess how well messages within text are understood. The primary concept of the test is to measure current levels of reading achievement. The reading paragraphs in the test contain a sentence with a blank space. Four or five single-word options are available for students to select to complete the sentence (as cited in Woods, 2007).
Florida Comprehensive Assessment Test–Sunshine State Standards (FCAT-SSS)	The reading portion of this standardized test is a group-administered, criterion-referenced test consisting of six to eight informational and literary reading passages (Florida Department of Education, 2005). ¹ In grades 3 through 10, students respond to between six and eleven multiple-choice items for each passage and are assessed across four content clusters: (1) reading comprehension in the areas of words and phrases in context, (2) main idea, (3) comparison/cause and effect, and (4) reference and research. In grades 4, 8, and 10, open-ended questions are included (as cited in Lang et al., 2008; Schatschneider et al., 2004). ²
New York State Department of Education End-of-Year Test in English Language Arts (NYSDE/ELA)	This standardized test is published by McGraw-Hill and contains multiple-choice questions and performance-assessment items. The multiple-choice questions are based on brief reading passages. For the performance assessment, students listen to and read passages and write responses to open-ended questions based on the passages. The reading and listening selections may be stories, articles, or poems. Three subtests are embedded within the ELA test: information and understanding; literacy response; and expression and critical analysis (as cited in White, Williams, & Haslem, 2005; http://schools.nyc.gov/Accountability/YearlyTesting/TestInformation/English+Language+Arts+(ELA).htm).
Stanford Achievement Test (SAT-9), Reading Comprehension subtest	This standardized subtest is composed of multiple-choice questions that measure reading comprehension (as cited in Interactive Inc., 2002). The Reading Comprehension subtest is composed of a scale of questions that range from interpreting simple sentences to understanding more complex paragraphs. The complex paragraphs ask the student to recognize directly stated details or relationships as well as implicit information and relationships that demand integration of what is provided in the text (as cited in Interactive Inc., 2002; Naglieri, Booth, & Winsler, 2004). ³
Stanford Achievement Test (SAT-9), Total Reading score	In this standardized test, students answer multiple-choice questions on two reading subtests (Reading Vocabulary and Reading Comprehension). The scores from these two subtests were aggregated into a single Total Reading score (as cited in Interactive Inc., 2002).
Texas Assessment of Knowledge and Skills (TAKS) Reading Test	This standardized test is designed to measure the extent to which a student has learned and is able to apply the defined knowledge and skills at each tested grade level. The reading test consists of multiple-choice and short answer items that assess basic understanding, ability to apply literary elements, ability to use strategies to analyze, and ability to apply critical thinking skills (as cited in Haslam, White, & Klinge, 2006; http://www.tea.state.tx.us/index3.aspx?id=3272&menu_id3=793).

1. Florida Department of Education. (2005, September). Florida Comprehensive Assessment Test Summary of Tests and Design. Retrieved August 21, 2008, from <http://fcats.fldoe.org/pdf/fc05designsummary.pdf>.
2. Schatschneider, C., Buck, J., Torgesen, J. K., Wagner, R. K., Hassler, L., Hecht, S., & Powell-Smith, K. (2004). *A multivariate study of factors that contribute to individual differences in performance on the Florida Comprehensive Reading Assessment Test*. (Technical Report No. 5). Tallahassee: Florida Center for Reading Research.
3. Naglieri, J. A., Booth, A. L., & Winsler, A. (2004). Comparison of Hispanic children with and without limited English proficiency on the Naglieri Nonverbal Ability Test. *Psychological Assessment*, 16(1), 81–84.

Appendix A2.2 Outcome measures for the general literacy achievement domain

Outcome measure	Description
California Standards Test, English Language Arts (CST-ELA)	This standardized achievement test is a component of the STAR (State Testing and Reporting) program, which is aligned with California's state standards for each grade level. The test addresses reading, writing, written and oral English language conventions, and listening and speaking. For grades 4–11, the test consists of 75 multiple-choice questions with an additional six field-test questions. At grades 4 and 7, the CST-ELA also includes a writing component, the California Writing Standards Test, which addresses a writing-applications standard selected for testing each year (as cited in Scholastic Research, 2008; http://www.cde.ca.gov/ta/tg/sr/elapreface.asp).
TerraNova Reading Test	This assessment is published by CTB/McGraw-Hill and combines multiple-choice items with open-ended questions that allow students to produce short and extended responses. The Reading Composite score is the average of the Reading Comprehension and Vocabulary subtest scores (as cited in White, Haslam, & Hewes, 2006; CTB/McGraw-Hill, 1996). ¹

1. CTB/McGraw-Hill. (1996). *TerraNova prepublication technical bulletin*. Monterey, CA: Author.

Appendix A3.1 Summary of study findings of all domains¹

Meets WWC evidence standards with reservations	Domain	
	Comprehension	General literacy achievement
Haslam, White, & Klinge (2006)	ind	nr
Interactive Inc. (2002)	(+)	nr
Lang et al. (2008)	+	nr
Scholastic Research (2008)	nr	+
White, Haslam, & Hewes (2006)	+	+
White, Williams, & Haslem (2005)	ind	nr
Woods (2007)	ind	nr
Rating of effectiveness	potentially positive effects	potentially positive effects

nr = no reported outcomes under this domain

+ = study finding was positive and statistically significant

(+) = study finding was positive and substantively important, but not statistically significant

ind = study finding was indeterminate; that is, neither substantively important nor statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices in each domain. More detailed information on findings for all measures within the domains and the constructs that factor into the domains can be found in Appendices A3.2–A3.3.

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain¹

Outcome measure	Study sample	Sample size (clusters/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation) ²		Mean difference ³ (READ 180– comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			READ 180 group	Comparison group				
Haslam, White, & Klinge, 2006 (quasi-experimental design)⁷								
Texas Assessment of Knowledge and Skills (TAKS) Reading Test ⁸	Grades 7 and 8	614	23.90 (12.0)	22.10 (14.40)	1.80	0.14	ns	+5
Average for comprehension (Haslam, White, & Klinge, 2006)						0.14	ns	+5
Interactive Inc., 2002 (quasi-experimental design)⁷								
Stanford Achievement Test (SAT-9), Reading Comprehension subtest ⁹	Grades 6 and 7, Columbus	5/171	621.52 (28.18)	602.25 (39.76)	19.27	0.60	ns	+23
Stanford Achievement Test (SAT-9), Total Reading score ⁹	Grade 8, Dallas	4/243	648.27 (21.69)	641.40 (33.05)	6.87	0.24	ns	+9
Stanford Achievement Test (SAT-9), Total Reading score ⁹	Grade 8, Houston	2/95	666.66 (22.09)	662.89 (32.25)	3.77	0.14	ns	+6
Average for comprehension (Interactive Inc., 2002)						0.33	na	+13
Lang et al., 2008 (randomized controlled trial where differential attrition could not be ruled out)⁷								
Florida Comprehensive Assessment Test–Sunshine State Standards (FCAT-SSS) ⁸	Grade 9, high risk	190	1,682.89 (196.92)	1,729.21 (236.27)	–46.32	–0.21	ns	–8
Florida Comprehensive Assessment Test–Sunshine State Standards (FCAT-SSS) ⁸	Grade 9, moderate risk	409	1,904.77 (134.15)	1,870.09 (130.09)	34.68	0.26	Statistically significant	+10
Average for comprehension (Lang et al., 2008)						0.02	na	+1
White, Haslam, & Hewes, 2006 (quasi-experimental design)⁷								
Stanford Achievement Test (SAT-9), Reading Comprehension subtest ⁸	Grade 9, cohort 1	1652	31.40 (9.30)	30.10 (11.30)	1.30	0.13	Statistically significant	+5
Average for comprehension (White, Haslam, & Hewes, 2006)						0.13	Statistically significant	+5

(continued)

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain¹ (continued)

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (READ 180-comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			READ 180 group	Comparison group				
White, Williams, & Haslem, 2005 (quasi-experimental design)⁷								
New York State Department of Education End-of-Year Test in English Language Arts (NYSDE/ELA) ¹⁰	Grade 4, proficiency level 1	229	606.8 (19.0)	609.0 (22.0)	-2.20	-0.10	ns	-4
New York State Department of Education End-of-Year Test in English Language Arts (NYSDE/ELA) ¹⁰	Grade 4, proficiency level 2	482	637.6 (20.0)	633.0 (24.0)	4.60	0.20	ns	+8
New York State Department of Education End-of-Year Test in English Language Arts (NYSDE/ELA) ¹⁰	Grade 4, proficiency level 3	319	665.0 (30.0)	671.0 (34.0)	-6.00	-0.18	ns	-7
CTB/McGraw-Hill Reading Test ¹⁰	Grade 6, proficiency level 1	215	606.7 (18.0)	619.0 (21.0)	-12.30	-0.59	ns	-22
CTB/McGraw-Hill Reading Test ¹⁰	Grade 6, proficiency level 2	471	642.1 (21.0)	639.0 (19.0)	3.10	0.16	ns	+6
CTB/McGraw-Hill Reading Test ¹⁰	Grade 6, proficiency level 3	274	674.1 (21.0)	667.0 (21.0)	7.10	0.34	ns	+13
New York State Department of Education End-of-Year Test in English Language Arts (NYSDE/ELA) ¹⁰	Grade 8, proficiency level 1	274	664.90 (16.0)	667.0 (12.0)	2.10	-0.17	ns	-7
New York State Department of Education End-of-Year Test in English Language Arts (NYSDE/ELA) ¹⁰	Grade 8, proficiency level 2	425	689.0 (18.0)	686.0 (14.0)	3.00	0.21	ns	+8

(continued)

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain¹ (continued)

Outcome measure	Study sample	Sample size (clusters/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (READ 180-comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			READ 180 group	Comparison group				
New York State Department of Education End-of-Year Test in English Language Arts (NYSDE/ELA) ¹⁰	Grade 8, proficiency level 3	201	717.90 (21.0)	707.0 (16.0)	10.90	0.67	ns	+25
Average for comprehension (White, Williams, & Haslem, 2005)¹¹						0.08	ns	+3
Woods, 2007 (quasi-experimental design)⁷								
Degrees of Reading Power (DRP) test ⁸	Grades 6, 7, and 8	116	44.81 (11.70)	45.21 (12.55)	-0.40	-0.03	ns	-1
Average for comprehension (Woods, 2007)						-0.03	ns	-1
Domain average for comprehension across all studies¹²						0.11	na	+4

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain. Subgroup findings from Woods (2007) are not included in these ratings but are reported in Appendix A4.1. Longitudinal findings from White, Haslam, and Hewes (2006) are not included in these ratings but are reported in Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the cases of Haslam, White, and Klinge (2006); White, Haslam, and Hewes (2006); and Woods (2007), no corrections for clustering or multiple comparisons were needed. In the cases of Interactive Inc. (2002); Lang et al. (2008); and White, Williams, and Haslem (2005), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
8. The intervention group values are the comparison group means plus the difference in mean gains between the intervention and comparison groups.
9. The intervention and control group means are ANCOVA-adjusted posttest scores reported by the authors in the article.
10. The intervention group values reported for White, Williams, and Haslem (2005) are the comparison group means plus the difference in mean gains between the intervention and comparison groups. The pretest and posttest means that were used to calculate the intervention group values were not reported in White, Williams, and Haslem (2005) but were provided to the WWC by the author. Because the NYSDE/ELA test was not vertically integrated across grades, the WWC calculated the effect size as the difference between the effect size for the posttest and the standardized pretest mean difference.
11. The average effect size is based on effect sizes that have been weighted by the sample size for each proficiency level within grade for White, Williams, and Haslem (2005).
12. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A3.3 Summary of study findings included in the rating for the general literacy achievement domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²	Comparison group	Mean difference ³ (READ 180-comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
Scholastic Research, 2008 (randomized controlled trial)⁷								
California Standards Test, English Language Arts (CST-ELA) ⁸	Grades 6, 7, and 9	570	293.05 (29.74)	280.16 (27.75)	12.89	0.45	Statistically significant	+17
Average for general literacy achievement (Scholastic Research, 2008)						0.45	Statistically significant	+17
White, Haslam, & Hewes, 2006 (quasi-experimental design)⁷								
TerraNova Reading Test ⁹	Grade 9, cohort 2	1630	41.20 (8.90)	38.30 (12.20)	2.90	0.27	Statistically significant	+11
TerraNova Reading Test ¹⁰	Grade 9, cohort 3	2058	39.00 (9.80)	38.10 (12.30)	0.90	0.08	ns	+3
Average for general literacy achievement (White, Haslam, & Hewes, 2006)						0.18	na	+7
Domain average for general literacy achievement across all studies¹¹						0.31	na	+12

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the general literacy achievement domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the cases of Scholastic Research (2008) and White, Haslam, and Hewes (2006), no corrections for clustering or multiple comparisons were needed.
8. The intervention and control group means are ANCOVA-adjusted posttest scores provided by the authors.
9. The intervention and control group means are posttest scores reported by the authors in the article.
10. The intervention group values are the comparison group means plus the difference in mean gains between the intervention and comparison groups. The intervention and control group standard deviations were not reported in White, Haslam, and Hewes (2006) or Scholastic Research (2008), but were provided to the WWC by the authors.
11. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4.1 Summary of subgroup findings for the comprehension domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (READ 180-comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			READ 180 group	Comparison group				
Woods, 2007 (quasi-experimental design)⁷								
Degrees of Reading Power (DRP) test ⁸	Grade 6	42	41.0 (10.98)	44.05 (16.08)	-3.05	-0.22	ns	-9
Degrees of Reading Power (DRP) test ⁸	Grade 7	36	46.56 (10.19)	44.83 (11.78)	1.72	0.15	ns	+6
Degrees of Reading Power (DRP) test ⁸	Grades 6, 7, and 8, African-American students	72	42.55 (12.39)	43.51 (11.19)	-0.96	-0.08	ns	-3

ns = not statistically significant

1. This appendix presents subgroup findings for measures that fall in the comprehension domain. The grade 8 cohort is not included because the intervention and comparison groups were not shown to be equivalent at baseline. Total group scores were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the case of Woods (2007), no correction for clustering was needed.
8. The intervention group values are the comparison group means plus the difference in mean gains between the intervention and comparison groups.

Appendix A4.2 Summary of later findings from longitudinal studies for the comprehension domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (READ 180–comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			READ 180 group	Comparison group				
White, Haslam, & Hewes, 2006 (quasi-experimental design)⁷								
<i>Two years after the start of the implementation of the intervention</i>								
Arizona Instrument to Measure Standards (AIMS) Reading Test ⁸	Grade 10, cohort 1	1448	664.10 (28.50)	664.20 (31.90)	–0.10	0.00	ns	0

ns = not statistically significant

1. This appendix presents later longitudinal findings for measures that fall in the comprehension domain. Data that reflected students' initial exposure to one year of the intervention were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the case of White, Haslam, and Hewes (2006), no correction for clustering was needed.
8. The intervention and control group means are posttest scores reported by the authors in the article.

Appendix A5.1 *READ 180* rating for the comprehension domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of comprehension, the WWC rated *READ 180* as having potentially positive effects for adolescent learners. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, or negative effects) were not considered because *READ 180* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. Two studies showed statistically significant positive effects and one study showed a substantively important positive effect on comprehension.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No study showed a statistically significant or substantively important negative effect, and three studies showed indeterminate effects on comprehension.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

Not met. Two studies showed statistically significant positive effects, but no studies met WWC evidence standards for a strong design.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. No study showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A5.2 READ 180 rating for the general literacy achievement domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of general literacy achievement, the WWC rated *READ 180* as having potentially positive effects for adolescent learners. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, or negative effects) were not considered because *READ 180* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. Two studies showed statistically significant positive effects on general literacy achievement.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies showed a statistically significant or substantively important negative effect, and no studies showed indeterminate effects on general literacy achievement.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

Not met. Two studies showed statistically significant positive effects, but no studies met WWC evidence standards for a strong design.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. No study showed statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Alphabetics	na	na	na	na
Reading fluency	na	na	na	na
Comprehension ²	6	>47	6380	Medium to large
General literacy achievement ³	2	>12	4258	Medium to large

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.”
2. One study (Haslam, White, & Klinge, 2006) did not report number of schools represented in the sample.
3. One study (Scholastic Research, 2008) did not report number of schools represented in the sample.