# What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES



# READ 180

## Program Description[1]

*READ 180* is a reading program designed for students in elementary through high school whose reading achievement is below the proficient level. The goal of *READ 180* is to address gaps in students' skills through the use of a computer program, literature, and direct instruction in reading skills. The software component of the program aims to track and adapt to each student's progress. In addition to the computer program, the *READ 180* program includes workbooks designed to address reading comprehension skills, paperback books for independent reading, and audiobooks with corresponding CDs for modeled reading.

## Research

No studies of *READ 180* meet What Works Clearinghouse (WWC) evidence standards,[2] but seven studies meet WWC evidence standards with reservations. The seven studies included 10,638 students, ranging from grade 4 to grade 9, who attended elementary, middle, and high schools in Arizona, California, Florida, New York, Ohio, Texas, and Virginia.[3]

Based on these seven studies, the WWC considers the extent of evidence for *READ 180* to be medium to large for comprehension and general literacy achievement. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *READ 180* in the alphabetics or reading fluency domains.

## Effectiveness

*READ 180* was found to have potentially positive effects on comprehension and general literacy achievement.

| | *Alphabetics* | *Reading fluency* | *Comprehension* | *General literacy achievement* |
|---|---|---|---|---|
| Rating of effectiveness | na | na | Potentially positive effects | Potentially positive effects |
| Improvement index[4] | na | na | Average: +4 percentile points | Average: +12 percentile points |
| | na | na | Range: –22 to +25 percentile points | Range: +3 to +17 percentile points |

na = not applicable

1. The descriptive information for this program was obtained from a publicly available source: the program's website (http://teacher.scholastic.com/products/read180/overview/, downloaded July 2009). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. This review refers to studies of *READ 180* in fourth through twelfth grade. Studies of *READ 180* conducted in kindergarten through third grade were out of the scope of the Adolescent Literacy protocol.
2. The studies included in this report were reviewed using WWC Evidence Standards, Version 1.0 (see the WWC Standards).
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
4. These numbers show the average and range of student-level improvement indices for all findings across the studies.

## Additional program information

### Developer and contact

Developed by the Cognition and Technology Group at Vanderbilt University, the Orange County Literacy Project in Florida, and the development staff at Scholastic Inc., *READ 180* is distributed by Scholastic Inc. In 1985, research by Dr. Ted Hasselbring led to the creation of the software prototype for *READ 180*.

Address: 577 Broadway, New York, NY 10012. Email: readcustserv@scholastic.com (Mary Eckelkamp, Supervisor, *READ 180*). Web: http://teacher.scholastic.com/products/read180/. Telephone: (800) 724-6527.

### Scope of use

Scholastic published the *READ 180* program in 1999. Since then, more than 10,000 classrooms in all 50 states have used *READ 180*.

### Teaching

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 90-minute instructional model begins with 20 minutes of whole-group direct instruction, in which the teacher provides instruction in reading, writing, and vocabulary 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 uses resource books and works closely with individual students
- students' independent use of the *READ 180* computer program to practice reading skills
- modeled and independent reading, in which students use *READ 180* paperbacks or audiobooks

Finally, the session 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.

### Cost

*READ 180* pricing depends on implementation.

## Research

One hundred one studies reviewed by the WWC investigated the effects of *READ 180*. Seven studies (Haslam, White, & Klinge, 2006; Interactive Inc., 2002; Lang, Torgesen, Petscher, Vogel, Chanter, & Lefsky, 2008; Scholastic Research, 2008; White, Haslam, & Hewes, 2006; White, Williams, & Haslem, 2006; Woods, 2007), one of which is a randomized controlled trial and six of which are quasi-experimental designs, meet WWC evidence standards with reservations.[5] The remaining 94 studies do not meet either WWC evidence standards or eligibility screens.

### Meets evidence standards with reservations

- Haslam, White, and Klinge (2006) conducted a quasi-experiment that examined the effects of *READ 180* on struggling readers in grades 7 and 8 in the Austin Independent School District of Texas. Intervention students were matched with comparison-group students using a one-to-one propensity score matching method. The WWC based its effectiveness ratings on findings from comparisons of the 307 students that received *READ 180* and the 307 comparison-group students that received the standard district curriculum. The

5. During the period covered by the studies in this report, two versions of *READ 180* (version 1.6 and the Enterprise Edition) were available. The WWC was unable to obtain information about which version of *READ 180* was used in the studies included in this report.

study reported student outcomes after one year of program implementation.

- Interactive Inc. (2002) conducted a quasi-experiment that examined the effects of *READ 180* on students in grades 6, 7, and 8 in three urban school districts in Texas and Ohio.[6] The WWC based its effectiveness ratings on findings from comparisons of the 52 control students and 119 students enrolled in *READ 180* in Columbus, the 142 control students and 101 students enrolled in *READ 180* in Dallas, and the 36 control students and 59 students enrolled in *READ 180* in Houston.[7] The study reported student outcomes after one year of program implementation.

- Lang et al. (2008) conducted a randomized controlled trial of 1,265 struggling readers in seven high schools in Florida. Ninth-grade students who scored in the high-risk or moderate-risk categories on the Florida Comprehensive Assessment Test (FCAT) were randomly assigned to one of four groups (Appendix A1.3 provides more details about these groups). The WWC based its effectiveness ratings on findings from comparisons of the 100 high-risk students who received *READ 180* and 90 high-risk students who were in the school-designed "business as usual" control group, as well as comparisons between 207 moderate-risk students who received *READ 180* and 202 moderate-risk students who were in the control group. The inability to determine whether differential attrition occurred and the use of multiple imputation of missing data led to the study's rating of "meets standards with reservations." The study reported student outcomes after one year of program implementation.

- Scholastic Research (2008) conducted a quasi-experiment that examined the effects of *READ 180* on students in grades 6, 7, and 9 in California. *READ 180* students who scored at the below-basic or basic performance level on the prior-year state reading test were matched to comparison-group students on the basis of reading scores and demographic characteristics. The WWC based its effectiveness ratings on findings from comparisons of the 285 students that received *READ 180* and the 285 comparison-group students that received either the Holt Literature and Language Arts curriculum (grades 6 and 7) or the Prentice Hall Literature curriculum (grade 9). The study reported student outcomes after one year of program implementation.

- White, Haslam, and Hewes (2006) conducted a quasi-experiment that examined the effects of *READ 180* on three cohorts of students in 12 schools in Arizona. *READ 180* students in grade 9 who were reading one or more grade levels below their assigned grade level were matched to comparison-group students with similar reading levels and demographic characteristics. The comparison-group students received the standard district curriculum. The WWC based its effectiveness ratings on findings from three cohorts of students. Cohort 1 was formed in the 2003–04 school year and consisted of 826 students in grade 9 who received *READ 180* and 826 comparison-group students in grade 9 who did not receive *READ 180*. Cohort 2 was formed in the 2004–05 school year and consisted of 815 students in grade 9 who received *READ 180* and 815 comparison-group students in grade 9 who did not receive *READ 180*. Cohort 3 was formed in the 2005–06 school year and consisted of 1,029 students in grade 9 who received *READ 180* and 1,029 comparison students who did not receive *READ 180*. The study reported student outcomes after the first year of program implementation.

- White, Williams, and Haslem (2005) conducted a quasi-experiment that examined the effects of *READ 180* on students in grades 4–8 in 16 schools in New York City. *READ 180* students' test outcomes were compared to outcomes of their

6. The study was originally designed as a randomized controlled trial, but none of the districts ultimately followed through with the research design.
7. The intervention and comparison groups in the grade 7 sample in Houston were not shown to be equivalent at baseline, so they were excluded from the review.

## Research *(continued)*

peers attending the same schools who did not receive *READ 180* but instead received the standard reading instruction in the 2001–02 school year. Comparisons were made between students with the same proficiency levels (1, 2, and 3)[8] within each grade.[9] The WWC based its effectiveness ratings on findings from comparisons between 362 students in the treatment group and 2,528 students in the control group, across grades 4, 6, and 8.[10] The study reported student outcomes after one year of program implementation.

- Woods (2007) conducted a quasi-experiment that examined the effects of *READ 180* on three annual cohorts of students in an urban middle school in Virginia. Based on pretest scores and teacher recommendations, the school's guidance counselor assigned students in grades 6–8 who needed additional literacy support to either the *READ 180* treatment group or the comparison group that would participate in the school's traditional reading-remediation program. The WWC based its effectiveness ratings on findings from comparisons of 58 middle-school students who received *READ 180* during the 2003–04 school year and 58 matched students who received the school's traditional reading-remediation program.[11] The study reported student outcomes after the first year of program implementation.

### Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.[12]

The WWC considers the extent of evidence for *READ 180* to be medium to large for comprehension and medium to large for general literacy achievement. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *READ 180* in the alphabetics or reading fluency domains.

## Effectiveness

### Findings

The WWC review of interventions for *READ 180* addresses student outcomes in four domains: alphabetics, reading fluency, comprehension, and general literacy achievement. The studies included in this report cover two domains: comprehension and general literacy achievement. Comprehension includes two constructs: reading comprehension and vocabulary development. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *READ 180* on students.[13]

8. There were only two treatment students in proficiency level 4 across grades 4–8; therefore, proficiency level 4 was excluded from the review.
9. The sample sizes, means, and standard deviations that were used to calculate the intervention group values by grade and proficiency level were not reported in White, Williams, and Haslem (2005), but were provided to the WWC by the author.
10. The intervention and comparison groups at grade 5 (proficiency levels 1, 2, and 3) were not shown to be equivalent at baseline and were excluded from the review. In addition, there were no treatment students in the grade 7 analysis sample, so grade 7 students were excluded from the review.
11. The 2004–05 and 2005–06 student cohorts, though included in the study, 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.
12. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept—external validity, such as the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *READ 180* is in Appendix A6.
13. 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 and for multiple comparisons. For an explanation, 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 for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the cases of Haslam, White, and Klinge (2006); Scholastic Research (2008); White, Haslam, and Hewes (2006); and Woods (2007), no corrections for clustering or multiple comparisons were needed. In the cases of Lang et al. (2008) and White, Williams, and Haslem (2005), corrections for multiple comparisons were needed, so the significance levels may differ from those reported in the original study. In the case of Interactive Inc. (2002), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.

## Effectiveness *(continued)*

*Comprehension.* Six studies reviewed findings in the comprehension domain. Haslam, White, and Klinge (2006) reported a statistically significant effect of *READ 180* on the Texas Assessment of Knowledge and Skills Reading Test. The WWC found that the effect was not statistically significant or large enough to be considered substantively important according to WWC criteria. Interactive Inc. (2002) reported statistically significant effects of *READ 180* on the Stanford Achievement Test for the Columbus and Dallas school districts but did not report a statistically significant effect of *READ 180* in the Houston school district. According to WWC calculations, the Columbus and Dallas effects were not statistically significant, but the average effect size across the three sites was large enough to be considered substantively important (that is, an effect size of at least 0.25).[14] Lang et al. (2008) reported, and the WWC confirmed, a statistically significant effect of *READ 180* on the Florida Comprehensive Assessment Test for moderate-risk students. For high-risk students, the effect was not statistically significant or large enough to be considered substantively important according to WWC criteria. White, Haslam, and Hewes (2006) found, and the WWC confirmed, a statistically significant effect of *READ 180* on the Reading Comprehension subtest of the Stanford Achievement Test. White, Williams, and Haslem (2005) did not find a statistically significant effect of *READ 180* on the New York State English Language Arts Test for grades 4 or 8 or on the CTB/McGraw-Hill Reading Test for grade 6. The WWC-calculated average effect across these three grades and two tests was not statistically significant or large enough to be considered substantively important according to WWC criteria. Woods (2007) did not find a statistically significant effect of *READ 180* on the Degrees of Reading Power test. The effect also was not statistically significant or large enough to be considered substantively important according to WWC criteria.

For the comprehension domain, two studies showed statistically significant positive effects, one study showed substantively important positive effects, and three studies showed indeterminate effects.

*General literacy achievement.* Two studies reviewed findings in the general literacy achievement domain. Scholastic Research (2008) reported, and the WWC confirmed, a statistically significant effect of *READ 180* on the English Language Arts subtest of the California Standards Test. White, Haslam, and Hewes (2006) reported a statistically significant positive effect of *READ 180* on the TerraNova Reading Test among cohort 2 and cohort 3 students. The WWC confirmed the statistically significant positive effect for cohort 2 students but found that the effect for cohort 3 students was not statistically significant or large enough to be considered substantively important according to the WWC criteria.

For the general literacy achievement domain, two studies showed statistically significant positive effects.

### Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings, the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Procedures and Standards Handbook, Appendix E).

**The WWC found *READ 180* to have potentially positive effects for comprehension and general literacy achievement**

### Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see WWC Procedures and Standards Handbook, Appendix F). The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the

14. The WWC computes an average effect size as a simple average of the effect sizes across all individual findings within the study domain.

**The WWC found *READ 180* to have potentially positive effects for comprehension and general literacy achievement** *(continued)*

analysis. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group. The average improvement index for comprehension is +4 percentile points across the six studies, with a range of –22 to +25 percentile points across findings. The average improvement index for general literacy achievement is +12 percentile points across the two studies, with a range of +3 to +17 percentile points across findings.

### Summary

The WWC reviewed 101 studies on *READ 180*. Seven of these studies meet WWC evidence standards with reservations; the remaining 94 studies do not meet either WWC evidence standards or eligibility screens. Based on the seven studies, the WWC found potentially positive effects in comprehension and general literacy achievement. The conclusions presented in this report may change as new research emerges.

## References

### Meets WWC evidence standards with reservations


Haslam, M. B., White, R. N., & Klinge, A. (2006). *Improving student literacy:* READ 180 *in the Austin Independent School District, 2004–05*. Washington, DC: Policy Studies Associates.

Interactive Inc. (2002). *An efficacy study of* READ 180, *a print and electronic adaptive intervention program, grades 4 and above.* New York, NY: Scholastic Inc.

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.* Paper presented at the annual research conference of the Society for Research on Educational Effectiveness, Crystal City, VA.

Scholastic Research. (2008). *Desert Sands Unified School District,* CA. New York, NY: Scholastic Inc.

White, R. N., Haslam, M. B., & Hewes, G. M. (2006). *Improving student literacy:* READ 180 *in the Phoenix Union High School District, 2003–04 and 2004–05*. Washington, DC: Policy Studies Associates.

***Additional source:***

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

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

Woods, D. E. (2007). *An investigation of the effects of a middle school reading intervention on school dropout rates.* Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg.


### Studies that fall outside the Adolescent Literacy review protocol or do not meet WWC evidence standards


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 a comparison group.

Aguhob, M. (2006). READ 180 *in Seminole County, Florida*. New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

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 a comparison group.

Alvermann, D. E., & Rush, L. S. (2004). Literacy intervention programs at the middle and high school levels. In T. L. Jetton & J. A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 210–227). New York: Guilford Press. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Banasik, B. (2002). *The effectiveness of Scholastic's* READ 180 *curriculum in improving reading comprehension in middle school aged students having a specific learning disability.* Unpublished master's thesis, Cardinal Stritch University, Milwaukee, WI. The study is ineligible for review because it does not use a comparison group.

## References *(continued)*


Barbato, P. F. (2007). A preliminary evaluation of the *READ 180* program (Doctoral dissertation, Fairleigh Dickinson University, 2006). *Dissertation Abstracts International, 67*(11A), 46–4130. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

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, Texas A&M University–Kingsville, 2007). *Dissertation Abstracts International, 68*(7A), 104–2877. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Benavidez-Candelaria, M. R. (2006). *An investigation of the program* READ 180 *and the effect it has on students' reading scores and students' grades.* Unpublished master's thesis, New Mexico Highlands University, Las Vegas. The study is ineligible for review because it does not use a comparison group.

Bishop-Temple, C. (2008). The effects of interactive read-alouds on the reading achievement of middle grade reading students in a core remedial program. *Dissertation Abstracts International, 68*(10A), 4179. The study is ineligible for review because it does not examine the effectiveness of an intervention.

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 does not examine the effectiveness of an 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. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Brown, L. (2006). *The impact of self-perception, reading attitude, and the use of the* READ 180 *program on reading achievement.* Unpublished master's thesis, Gwynedd-Mercy College, Gwynedd Valley, PA. The study is ineligible for review because it does not use a comparison group.

Brown, S. H. (2006). The effectiveness of the *READ 180* intervention for struggling readers in grades 6–8 (Doctoral dissertation, Union University, 2006). *Dissertation Abstracts International, 67*(8A), 246–2922. The study is ineligible for review because it does not use a comparison group.

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, College of William and Mary, 2007). *Dissertation Abstracts International, 68*(4A), 1383. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Campbell, Y. C. (2006). Effects of an integrated learning system on the reading achievement of middle school students (Doctoral dissertation, University of Miami, 2006). *Dissertation Abstracts International, 67*(8A), 100–2923. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

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 a comparison group.

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, University of Delaware, 2004). *Dissertation Abstracts International, 65*(5A), 1717. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

## References *(continued)*


Deshler, D. D., Palincsar, A. S., Biancarosa, G., & Nair, M. (2007). *Informed choices for struggling adolescent readers: A research-based guide to instructional programs and practices.* Newark, DE: International Reading Association. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Dunn, C. A. (2002). *An investigation of the effects of computer assisted reading instruction versus traditional reading instruction on selected high school freshmen.* Unpublished doctoral dissertation, Loyola University, Chicago, IL. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—the intervention was combined with another intervention.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort.* Washington, DC: U.S. Department of Education, Institute of Education Sciences. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—the effects are not reported separately for the intervention.

Felty, R. L. (2008). *READ 180* implementation: Reading achievement and motivation to read within an alternative education middle school program (Doctoral dissertation, Immaculata College, 2008). *Dissertation Abstracts International, 69*(1A), 182–161. The study is ineligible for review because it does not use a comparison group.

Ferguson, J. M. (2005). *The implementation of technology in reading classrooms and the impact of technology integration and student perceptions on reading achievement.* Unpublished doctoral dissertation, Texas A&M University–Commerce. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Gentry, L. (2006). An evaluation of *READ 180* in an urban secondary school (Doctoral dissertation, American University, 2006). *Dissertation Abstracts International, 67*(9A), 151–3346. The study does not meet WWC evidence standards because the overall attrition rate exceeds WWC standards for this area.

Goodloe-Johnson, M. L., McGinley, N. J., Rose, J. S., & 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 does not examine the effectiveness of an intervention.

Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading & Writing Quarterly, 20*(2), 123–144. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

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 is not a primary analysis of the effectiveness of an intervention.

Hewes, G. M., Palmer, N., Haslam, M. B., & Mielke, M. B. (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 a comparison group.

Holyoke Public Schools. (2006). READ 180 *overview.* Holyoke, MA: Author. The study is ineligible for review because it does not use a comparison group.

Iwamiya, C., & Pritchard, R. (2005). *Application and effects of Scholastic's* READ 180 *as an after-school reading intervention program for English language learners.* Unpublished master's thesis, California State University, Sacramento. The study is ineligible for review because it does not use a comparison group.

Knutson, K. A. (2006). *Because you can't wait until spring: Using the SRI to improve reading performance.* New York, NY:

## References *(continued)*


Scholastic Inc. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Kratofil, M. D. (2006). A comparison of the effect of Scholastic *READ 180* and traditional reading interventions on the reading achievement of middle school low-level readers (Master's thesis, University of Central Missouri, 2006). *Master's Abstracts International, 44*(6), 52–2531. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

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 does not examine the effectiveness of an intervention.

Moore, S. A. (2007). *Impact of* READ 180 *on reader's lexile scores in grades three through five.* Unpublished master's thesis, Troy University, Troy, AL. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

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, East Tennessee State University, 2007). *Dissertation Abstracts International, 68*(6A), 2265. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Palmer, N. (2003). READ 180 *middle-school study: Des Moines, Iowa, 2000–2002: An evaluation of* READ 180 *with special education students.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

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 evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

***Additional source:***

Papalewis, R. (2004). Struggling middle school readers: Successful, accelerating intervention. *Reading Improvement, 41*(1), 24–37.

Papalewis, R. (2003). *A study of* READ 180 *in middle schools in Clark County School District, Las Vegas, Nevada.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

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 a comparison group.

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 a comparison group.

Scholastic Inc. (2005). READ 180 *implementation year study.* New York, NY: Author. The study is ineligible for review because it does not use a comparison group.

Scholastic Inc. (n.d.). *Platinum performers: Blackhawk Middle School.* New York, NY: Author. The study is ineligible for review because it does not use a comparison group.

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 a comparison group.

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 a comparison group.

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 a comparison group.

Scholastic Inc. (n.d.). *Platinum performers: Madison Middle School.* New York, NY: Author. The study is ineligible for review because it does not use a comparison group.

Scholastic Inc. (n.d.). *Platinum performers: Martin Luther King Elementary School.* New York, NY: Author. The study is

## References *(continued)*


ineligible for review because it does not use a comparison group.

Scholastic Inc. (n.d.). *Platinum performers: Sebastopol Attendance Center.* New York, NY: Author. The study is ineligible for review because it does not use a comparison group.

Scholastic Inc. (n.d.). *Platinum performers: Socorro Independent School District, TX.* New York, NY: Author. The study is ineligible for review because it does not use a comparison group.

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 a comparison group.

Scholastic Research and Validation. (2003). *A study of* READ 180 *stages A and B in upper elementary and middle schools in Iredell-Statesville Schools, North Carolina.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2003). *Calera Elementary School, Calera, Alabama*. New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2003). *Carter Junior High School, Arlington, Texas.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and 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 a comparison group.

Scholastic Research and Validation. (2003). *Highview Elementary School, Nanuet, New York*. New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2003). *Holmes Middle School, Alexandria, Virginia.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2003). *Lufkin High School, Lufkin, Texas.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2003). *Marshall Middle School, Beaumont, Texas*. New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and 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 a comparison group.

Scholastic Research and Validation. (2003). *Raba Elementary School, San Antonio, Texas.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2003). *READ 180*: Compilation of school studies: Impact study. *READ 180 reports guide: Using data to drive instruction* (pp. 80). New York, NY: Scholastic Inc. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Scholastic Research and Validation. (2003). *Rochester Elementary School, Rochester, Pennsylvania.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2003). *Rogers Middle School, Boston, Massachusetts.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and 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 a comparison group.

Scholastic Research and 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 a comparison group.

## References *(continued)*


Scholastic Research and Validation. (2003). *Two Eagle River School, Pablo, Montana.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2003). *Washington Middle School, Kenosha, Wisconsin*. New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and 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 a comparison group.

Scholastic Research and Validation. (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: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2006). *Compendium of* READ 180 *research.* New York, NY: Scholastic Inc. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Scholastic Research and Validation. (2006). READ 180: *A heritage of research.* New York, NY: Scholastic Inc. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Scholastic Research and Validation. (2007). *Department of Defense Education Activity Schools (DoDEA).* New York, NY: Scholastic Inc. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

Scholastic Research and Validation. (2007). *Phoenix College, AZ.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a sample within the age or grade range specified in the review protocol.

Scholastic Research and Validation. (2008). *Compendium of* READ 180 *research*. New York, NY: Scholastic Inc. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Scholastic Research and Validation. (2008). *Martin County Public Schools, FL*. New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2008). *Peoria Public School District 150, IL*. New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2008). *Socorro Independent School District, TX.* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and Validation. (2008). *Traverse City Area Public Schools, MI*. New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Scholastic Research and 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 a comparison group.

Scholastic Research and Validation. (n.d.). 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 a comparison group.

Scholastic Research and Validation & Indian River School District. (2006). *Special education students, Selbyville Middle and Sussex Central Middle Schools, Indian River School District (Delaware).* New York, NY: Scholastic Inc. The study is ineligible for review because it does not use a comparison group.

Shanahan, C. (2005). *Adolescent literacy intervention programs: Chart and program review guide.* Naperville, IL: Learning Point Associates/North Central Regional Educational Laboratory. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

## References *(continued)*


Slavin, R. E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly, 43*(3), 290–322. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Soto, J. (2004). *The effects of* READ 180 *on the reading fluency of fifth grade students.* Unpublished master's thesis, California State University, San Marcos. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Sousa, D. A. (2005). *How the brain learns to read.* Thousand Oaks, CA: Corwin Press. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Thomas, D. M. (2005). Examining the academic and motivational outcomes of students participating in the *READ 180* program. (Doctoral dissertation, University of Kentucky, 2005). *Dissertation Abstracts International, 67*(1A), 96–112. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Thomas, J. (2003). *Reading program evaluation:* READ 180*, grades 4–8, November, 2003.* Kirkwood, MO: Kirkwood School District. The study is ineligible for review because it does not use a comparison group.

Thompson, L. T. (2005). READ 180: *An English language learner intervention strategy with grades 6–8*. Unpublished master's thesis, Alabama A&M University, Normal. The study is ineligible for review because it does not use a sample aligned with the review protocol.

Thorpe, P. K. (2003). *Escher's intersecting worlds: Evaluation as a reflection of the evaluation, the evaluator being reflected in the evaluation.* Wichita, KS: Wichita Public Schools. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Visher, M., & Hartry, A. (2007). *Can after-school programs boost academic achievement? An impact evaluation of a reading intervention in an after-school program.* Washington, DC: MPR Associates, Inc. The study does not meet WWC evidence standards because it does not provide adequate information to determine whether it uses an outcome that is valid or reliable.

Wahl, M. (2008). READ 180 *Enterprise Edition.* Tallahassee, FL: Florida Center for Reading Research. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

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, University of Delaware, 2007). *Dissertation Abstracts International, 69*(1A), 63–117. The study is ineligible for review because it does not use a comparison group.

Witkowski, P. M. (2004). A comparison study of two intervention programs for reading-delayed high school students (Doctoral dissertation, University of Missouri–Saint Louis, 2004). *Dissertation Abstracts International, 65*(6A), 163–2142. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Zapata, J. M. (2007). *If I can read, I can learn: Examining the effects of* READ 180 *on the reading comprehension of English language learners.* Unpublished master's thesis, California State University, San Marcos. The study is ineligible for review because it does not use a sample aligned with the review 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 a comparison group.

# 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). *Improving student literacy:* READ 180 *in the Austin Independent School District, 2004–05*. Washington, DC: Policy Studies Associates. |
| **Participants** | From the initial pool of 409 *READ 180* 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 *READ 180* group (89%) than in the comparison group (73%), the groups were equivalent on the pretest achievement measure. In all, 307 students in the *READ 180* 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 *READ 180* 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). *An efficacy study of* READ 180, *a print and electronic adaptive intervention program, grades 4 and above*. 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 *READ 180* 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 *READ 180* 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 *READ 180* in Columbus, 142 comparison students and 101 students enrolled in *READ 180* in Dallas, and 36 comparison students and 59 students enrolled in *READ 180* 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 *READ 180* 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). *Exploring the relative effectiveness of reading interventions for high school students*. 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—*REACH, RISE,* or *READ 180*—or to a control condition—*School Offered Accelerated Reading (SOAR).* 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 *READ 180* and 90 high-risk students in the comparison group, as well as 207 moderate-risk students who received *READ 180* 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 *READ 180* 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 comparsion group received the district's standard curriculum: *SOAR*. The implementation of *SOAR* involved the following materials: the *Reading and Writing Sourcebook* by Great Source, the *Reader's Handbook* by Great Source, *Reading Nonfiction* by Jamestown, and the *Daybook of Critical Reading and Writing* by Great Source. The *SOAR* 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 *READ 180* and *SOAR* interventions. Both *READ 180* teachers and *SOAR* 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 *READ 180* and *SOAR* teachers, and intervention-specific monthly support meetings were held to address concerns. The publisher of the *READ 180* intervention was asked to participate in the provision of materials, the conduct of professional development for *READ 180* 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). *Desert Sands Unified School District, CA.* 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 *READ 180* 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 *READ 180* 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 *READ 180* 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 that 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). *Improving student literacy:* READ 180 *in the Phoenix Union High School District, 2003–04 and 2004–05.* Washington, DC: Policy Studies Associates. |
| **Participants** | Three cohorts of ninth-grade students who were reading one or more years below grade level participated in *READ 180* 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 *READ 180* in 2003–04 and 826 matched nonparticipants. The cohort 2 analysis sample consisted of 815 students who received *READ 180* in 2004–05 and 815 matched nonparticipants. The cohort 3 analysis sample consisted of 1,029 students who received *READ 180* 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 *READ 180*, 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). *Performance of District 23 students participating in Scholastic* READ 180. Washington, DC: Policy Studies Associates. |
| **Participants** | The authors compared English Language Arts test outcomes for *READ 180* students in 16 schools to outcomes of their peers attending the same schools who did not participate in *READ 180*. For the overall sample of students in grades 4–8, the profile of the students selected to participate in *READ 180* was similar to that of comparison students. About 85% of students were African-American, and 90% were eligible for free or reduced-price lunch. *READ 180* 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 *READ 180* 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 *READ 180* 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). *An investigation of the effects of a middle school reading intervention on school dropout rates*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg. |
| **Participants** | Three annual cohorts of middle-school students participated in *READ 180* 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 *READ 180* 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 *READ 180* 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 *READ 180* 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 *READ 180* program implementation was downgraded from Level One (the highest level of fidelity) to Level Two, according to the *READ 180* 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** | *READ 180* 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://fcat.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 domains[1]

| | Domain | |
|---|---|---|
| **Meets WWC evidence standards with reservations** | **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 domain[1]

| Outcome measure | Study sample | Sample size (clusters/ students) | Authors' findings from the study: Mean outcome (standard deviation)[2] — *READ 180* group | Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group | WWC calculations: Mean difference[3] (*READ 180*– comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: Improvement index[6] |
|---|---|---|---|---|---|---|---|---|
| **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 domain[1] *(continued)*

| Outcome measure | Study sample | Sample size (clusters/ students) | Authors' findings from the study: Mean outcome (standard deviation)[2] — *READ 180* group | Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group | WWC calculations: Mean difference[3] (*READ 180*– comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: Improvement index[6] |
|---|---|---|---|---|---|---|---|---|
| **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 domain[1] *(continued)*

| Outcome measure | Study sample | Sample size (clusters/ students) | Authors' findings from the study: Mean outcome (standard deviation)[2] — *READ 180* group | Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group | WWC calculations: Mean difference[3] (*READ 180*– comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: Improvement index[6] |
|---|---|---|---|---|---|---|---|---|
| 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 studies[12]** | | | | | | **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, Willams, 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 domain[1]

| Outcome measure | Study sample | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation)[2] — *READ 180* group | Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group | WWC calculations: Mean difference[3] (*READ 180*– comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: 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 studies[11]** | | | | | | **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 domain[1]

| Outcome measure | Study sample | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation)[2] — *READ 180* group | Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group | WWC calculations: Mean difference[3] (*READ 180*– comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: Improvement index[6] |
|---|---|---|---|---|---|---|---|---|
| **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 domain[1]

| Outcome measure | Study sample | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation)[2] — *READ 180* group | Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group | WWC calculations: Mean difference[3] (*READ 180*– comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: Improvement index[6] |
|---|---|---|---|---|---|---|---|---|
| **White, Haslam, & Hewes, 2006 (quasi-experimental design)[7]** | | | | | | | | |
| ***Two years after the start of the implementation of the intervention*** | | | | | | | | |
| 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. 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. 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.