



WWC Intervention Report

A summary of findings from a systematic review of the evidence



Adolescent Literacy

November 2017

Prentice Hall/Pearson Literature[©] (2007–15)

Intervention Description¹

Prentice Hall/Pearson Literature[©] (2007–15) is an English language arts curriculum designed for students in grades 6–12 that focuses on building reading, vocabulary, literary analysis, and writing skills. It uses passages from fiction and nonfiction texts, poetry, and contemporary digital media. The curriculum is based on a textbook. The publisher also provides online components and other materials that enable teachers to provide personalized assignments, monitor students' progress, and score writing assignments, enrich instruction, or provide additional practice to supplement the textbook.

Research²

The What Works Clearinghouse (WWC) identified three studies of *Prentice Hall/Pearson Literature*[©] (2007–15) that fall within the scope of the Adolescent Literacy topic area and meet WWC group design standards. One study meets WWC group design standards without reservations, and two meet WWC group design standards with reservations. Together, these studies included 4,149 adolescent readers in grades 7–10 in 20 schools in the United States.

According to the WWC review, the extent of evidence for *Prentice Hall/Pearson Literature*[©] (2007–15) on the achievement of adolescent readers was medium to large for two outcome domains—general literacy achievement and comprehension. No studies meet WWC group design standards in the alphabetic or reading fluency domains, so this intervention report does not report on the effectiveness of *Prentice Hall/Pearson Literature*[©] (2007–15) for those domains.³ (See the Effectiveness Summary on p. 5 for more details of effectiveness by domain.)

Effectiveness

Prentice Hall/Pearson Literature[©] (2007–15) had no discernible effects on general literacy achievement and comprehension for adolescent readers.

Table 1. Summary of findings⁴

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
General literacy achievement	No discernible effects	+2	+2 to +2	2	2,558	Medium to large
Comprehension	No discernible effects	-3	-4 to -2	3	4,149	Medium to large

Report Contents

Overview	p. 1
Intervention Information	p. 2
Research Summary	p. 4
Effectiveness Summary	p. 5
References	p. 7
Research Details for Each Study	p. 8
Outcome Measures for Each Domain	p. 14
Findings Included in the Rating for Each Outcome Domain	p. 15
Supplemental Findings for Each Outcome Domain	p. 17
Endnotes	p. 19
Rating Criteria	p. 20
Glossary of Terms	p. 21

This intervention report presents findings from a systematic review of *Prentice Hall/Pearson Literature*[©] (2007–15) conducted using the WWC Procedures and Standards Handbook (version 3.0) and the Adolescent Literacy review protocol (version 3.0).

Intervention Information

Background

Prentice Hall/Pearson Literature® is an English language arts curriculum designed for students in grades 6–12. It is available in multiple editions, including *Prentice Hall Literature*® (1989), *Prentice Hall Literature: Timeless Voices, Timeless Themes*® (2000, 2002, 2005), *Prentice Hall Literature: Penguin Edition*® (2007), *Prentice Hall Literature: Language and Literacy*® (2010), *Prentice Hall Literature: Common Core Edition*® (2012), and *Pearson Literature*® (2015). This report focuses on the latter editions—*Prentice Hall Literature: Penguin Edition*® (2007), *Prentice Hall Literature: Language and Literacy*® (2010), *Prentice Hall Literature: Common Core Edition*® (2012), and *Pearson Literature*® (2015).⁵ The WWC refers to each of these four editions as *Prentice Hall/Pearson Literature*® (2007–15) in this intervention report, unless the edition was noted in the original study.

Address: Pearson Prentice Hall, One Lake Street, Upper Saddle River, NJ 07458. Telephone: 800-848-9500. Web: <http://www.pearsonschool.com>

Intervention details

The *Prentice Hall/Pearson Literature*® (2007–15) curriculum uses passages from fiction and nonfiction texts, poetry, and contemporary digital media to help students develop literacy skills. Lessons are guided by what the developer refers to as the “Big Questions” (for example, “What is the best way to find the truth?”), which each lesson revisits via discussion and writing assignments. Teachers can differentiate instruction within classrooms by selecting texts and resources designated for students with varying levels of reading ability (below-level, on-level, and advanced students) as well as for English learners and students with special needs.

The curriculum for grades 6–10 focuses on building reading, vocabulary, and writing skills, whereas the curriculum for grades 11 and 12 focuses on literary analysis. Vocabulary-building exercises start by introducing new vocabulary words at the beginning of a lesson. Students acquire new vocabulary words by using worksheets and games that present vocabulary words in context and provide information on word origins, idioms, cognates, and multiple meanings of words. The curriculum includes writing exercises with each leveled reading selection. Pre-writing assignments help students develop and organize their content, and guided writing exercises help students develop their ideas into full-length compositions. Students can practice for assessments like the PSAT, the ACT, and the SAT by taking timed reading assignments, which are scored automatically online. Literary analysis instruction focuses on comparing literary works and practicing reading. Students analyze and interpret texts to develop skills such as figuring out the meaning of new words, interpreting texts, citing evidence, organizing information, synthesizing information across texts, evaluating the accuracy of information, strengthening reading comprehension, and writing in a variety of genres (such as narratives, poetry, and reflective essays).

Prentice Hall/Pearson Literature® (2007–15) provides several supplementary resources, such as *Readers’ Notebooks* and *Reality Central*. *Readers’ Notebooks* provide reading support and additional skills practice. They are available in different editions tailored for on-level students, English learners, below-level students, and Spanish speakers. *Reality Central* provides additional reading passages thematically linked to the Big Questions in each lesson.

Prentice Hall/Pearson Literature® (2007–15) makes available a teacher’s edition, with print and online editions, that provide detailed lesson plans and additional lessons to extend learning. It includes components that enable teachers to test and individualize assignments, as well as assess improvement. Diagnostic tests are available to administer at the beginning of grades 6–10 and after each reading selection. Online resources, such as *PHLitOnline* (for *Prentice Hall Literature*® editions) or *Pearson Realize* (for *Pearson Literature*®), enable teachers to personalize assignments for their students, provide lesson activities, and monitor students’ progress.

Cost

As of January 2017, the publisher sold *Pearson Literature*® (2015) and three editions of *Prentice Hall Literature*®—*Prentice Hall Literature: Timeless Voices, Timeless Themes*® (2005); *Prentice Hall Literature*® (2010), and *Prentice Hall Literature: Common Core Edition*® (2012)—on its website. The student editions of these curricula cost \$82 to \$95 per book. The annotated teacher’s editions cost \$134 to \$145. Depending on the edition, Pearson also provides a variety of supplementary resources for these curricula, ranging from instructional CD-ROMs (\$150–\$345) and eText for iPad® or Android™ devices (\$60 for a 6-year student license, \$91 for a 6-year teacher license) to vocabulary cards (\$97), transparencies (\$31), teaching guides (\$22), and *Readers’ Notebooks* (\$14). Additional cost information is available from the publisher.

Research Summary

The WWC identified three eligible studies that investigated the effects of *Prentice Hall/Pearson Literature*® (2007–15) on the reading achievement of adolescent readers. Citations for all three studies are in the References section, which begins on p. 7.

The WWC reviewed three eligible studies against group design standards. One study is a randomized controlled trial that meets WWC group design standards without reservations, and two studies are randomized controlled trials with jeopardized random assignment that meet WWC group design standards with reservations (see the Glossary of Terms in this document for a definition of commonly used research terms). This report summarizes these three studies.

Table 2. Scope of reviewed research

Grades	7–10
Delivery method	Whole class
Intervention type	Curriculum

Summary of study meeting WWC group design standards without reservations

Resendez and Azin (2015) conducted a cluster, or group-based, randomized controlled trial examining the effects of *Pearson Literature*® (2015) on ninth-grade students in five high schools in California, Illinois, Michigan, and Washington states. Within each school, the authors randomly assigned teachers to intervention or comparison (business-as-usual) conditions. During the 2014–15 school year, nine intervention teachers implemented *Pearson Literature*® in 25 English language arts classrooms, and 10 comparison teachers implemented their schools' standard curriculum in 23 classrooms. Comparison teachers could design their own curriculum or supplement the available curriculum as they saw fit, following their school's policy. The WWC based its effectiveness rating on findings from the sample of 1,004 ninth-grade students in five schools; 530 students were in the *Pearson Literature*® group, and 474 students were in the comparison group.

Summary of studies meeting WWC group design standards with reservations

Berry et al. (2007) conducted a cluster randomized controlled trial examining the effects of *Prentice Hall Literature: Penguin Edition*® (2007) on students in grades 7 and 9 in four middle schools and three high schools in California, Colorado, and Illinois. Within each school, the authors randomly assigned teachers to the intervention or comparison (business-as-usual) groups. Because the study is a cluster randomized controlled trial that might have analyzed outcomes for students who were not present at the time of random assignment, the integrity of the study's random assignment was jeopardized. However, the authors demonstrated equivalence of the analytic intervention and comparison groups at baseline.⁶ During the 2006–07 school year, 15 intervention teachers taught English language arts to 1,016 students using *Prentice Hall Literature: Penguin Edition*®, and 16 comparison teachers taught 906 students using their schools' standard curricula. The WWC based its effectiveness rating on findings from the sample of 726 seventh-grade students from four middle schools and 901 ninth-grade students from three high schools: overall, 890 students were in the intervention group, and 737 students were in the comparison group.

Eddy et al. (2010) conducted a cluster randomized controlled trial examining the effects of *Prentice Hall Literature*® (2010) on students in grades 7, 8, and 10 in eight schools in Arizona, California, Ohio, and Oregon. Within each school, the authors randomly assigned teachers to the intervention or comparison groups. Because the study is a cluster randomized controlled trial that might have analyzed outcomes for students who were not present at the time of random assignment, the integrity of the study's random assignment was jeopardized. However, the authors demonstrated equivalence of the analytic intervention and comparison group at baseline.⁷ During the 2009–10 school year, 16 intervention teachers used *Prentice Hall Literature*®, and 13 comparison teachers implemented their schools' standard English language arts curriculum. The WWC based its effectiveness rating on findings from the combined sample of 1,518 students in grades 7, 8, and 10 in eight schools; 744 students were in the intervention group, and 774 students were in the comparison group.

Effectiveness Summary

The WWC review of *Prentice Hall/Pearson Literature*® (2007–15) for the Adolescent Literacy topic area includes outcomes in four domains: alphabetics, reading fluency, comprehension, and general literacy achievement. The three studies of *Prentice Hall/Pearson Literature*® (2007–15) that met WWC group design standards reported findings in two of the four domains: general literacy achievement and comprehension. The following findings present the authors’ and WWC-calculated estimates of the size and statistical significance of the effects of *Prentice Hall/Pearson Literature*® (2007–15) on adolescent readers. Additional comparisons are available as supplemental findings in Appendix D. The supplemental findings do not factor into the intervention’s rating of effectiveness. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 20.

Summary of effectiveness for the general literacy achievement domain

Table 3. Rating of effectiveness and extent of evidence for the general literacy achievement domain

Rating of effectiveness	Criteria met
No discernible effects <i>No affirmative evidence of effects.</i>	In the two studies that reported findings, the estimated impact of the intervention on outcomes in the <i>general literacy achievement</i> domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	Criteria met
Medium to large	Two studies that included 2,558 students in 12 schools reported evidence of effectiveness in the <i>general literacy achievement</i> domain.

Two studies that met WWC group design standards with or without reservations reported findings in the general literacy domain.

Resendez and Azin (2015) reported, and the WWC confirmed, no statistically significant effects of *Pearson Literature*® for students in grade 9 on the overall English language arts score of the Iowa Assessment, Form E (Iowa Form E). The effect size was not large enough to be considered substantively important according to WWC criteria (that is, an effect size of at least 0.25). The WWC characterizes this study finding as an indeterminate effect.

Berry et al. (2007) examined students’ scores on the language subtest of the Iowa Test of Basic Skills (ITBS) and reported, and the WWC confirmed, no statistically significant effects of *Prentice Hall: Penguin Edition*® on the scores of grade 7 students. The authors also reported, and the WWC confirmed, no statistically significant effects of *Prentice Hall Literature: Penguin Edition*® on the spelling subtest of the Iowa Test of Educational Development (ITED) for students in grade 9. The average effect size across the two outcomes was not large enough to be substantively important. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the general literacy achievement domain, two studies showed an indeterminate effect. This results in a rating of no discernible effects, with a medium to large extent of evidence.

Summary of effectiveness for the comprehension domain

Table 4. Rating of effectiveness and extent of evidence for the comprehension domain

Rating of effectiveness	Criteria met
No discernible effects <i>No affirmative evidence of effects.</i>	In the three studies that reported findings, the estimated impact of the intervention on outcomes in the <i>comprehension</i> domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	Criteria met
Medium to large	Three studies that included 4,149 students in 20 schools reported evidence of effectiveness in the <i>comprehension</i> domain.

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

Resendez and Azin (2015) examined scores on the two subtests of the Iowa Form E—Reading and Vocabulary—for students in grade 9. The authors reported, and the WWC confirmed, no statistically significant effects of *Pearson Literature*® on either of the subtests. The average effect size across the two outcomes was not large enough to be substantively important. The WWC characterizes these study findings as an indeterminate effect.

Berry et al. (2007) examined scores on the reading comprehension subtest of the ITBS for grade 7 students and the reading subtest of the ITED for grade 9 students. The authors reported, and the WWC confirmed, no statistically significant effects of *Prentice Hall: Penguin Edition*® for students on either of the subtests. The average effect size across the two outcomes was not large enough to be substantively important. The WWC characterizes these study findings as an indeterminate effect.

Eddy et al. (2010) examined scores on the Gates-MacGinitie Reading Test (GMRT). The authors reported, and the WWC confirmed, no statistically significant effects of *Prentice Hall Literature*® on the GMRT scores for the combined sample of students in grades 7, 8, and 10. The effect was not large enough to be substantively important. The WWC characterizes this study finding as an indeterminate effect.

Thus, for the comprehension domain, three studies showed an indeterminate effect. This results in a rating of no discernible effects, with a medium to large extent of evidence.

References

Study that meets WWC group design standards without reservations

Resendez, M., & Azin, M. (2015). *A report on the effects of the Pearson Literature Program on student language arts skills*. Jackson, WY: PRES Associates, Inc.

Studies that meet WWC group design standards with reservations

Berry, T., Eddy, R. M., Fleischer, D., Asgarian, M., & Malek, Y. (2007). *The effects of Prentice Hall Literature (Penguin Edition) curriculum on student performance: Randomized control trial final report*. Claremont, CA: Claremont Graduate University.

Eddy, R. M., Ruitman, H. T., Hanken, N., & Sloper, M. (2010). *The effects of Pearson Prentice Hall Literature (2010) on student performance: Efficacy study*. La Verne, CA: Cobblestone Applied Research and Evaluation, Inc.

Studies that do not meet WWC group design standards

None.

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

None.

Appendix A.1: Research details for Resendez et al. (2015)

Resendez, M., & Azin, M. (2015). *A report on the effects of the Pearson Literature Program on student language arts skills*. Jackson, WY: PRES Associates, Inc.

Table A1. Summary of findings

Meets WWC group design standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General literacy achievement	932 students/5 schools	+2	No
Comprehension	1,004 students/5 schools	-4	No

Setting The study took place in five public high schools located in urban and suburban areas in California, Illinois, Michigan, and Washington states. The size of the schools range from medium (below 1,000) to large (over 2,000). The study was implemented in 48 ninth-grade classrooms (25 intervention, 23 comparison) taught by 19 teachers.

Study sample The study used a cluster randomized controlled trial design. Within each school, teachers were randomly assigned to intervention or comparison groups. To be eligible to participate in the study, schools had to meet the following criteria: (1) school staff had to be willing to participate in the study and support implementation of the intervention, (2) schools were also required to have no other major English/language arts initiatives taking place, and (3) schools had to have low student mobility rates (less than 20% student attrition during a school year).

During fall 2014, the study was implemented in 48 classrooms (25 intervention, 23 comparison) taught by 19 teachers. The analysis sample included 1,004 ninth-grade students: 530 students were in the *Pearson Literature* group, and 474 students were in the comparison group. Across the sample, there were 75.8% White students, 12.1% Hispanic students, 6.3% African-American students, and the remaining students were from other racial or ethnic groups. Other subpopulation breakdowns included: 3.7% special education status, 7.9% limited English proficiency, 21.3% eligible for free or reduced-price lunch, and reading levels ranged from low (28.1%) to mid-range (35.6%) to high (36.3%). For the 19 teachers in the analysis sample, 85% were female and 95% were White; 74% held a Master’s degree, 21% held a Bachelor’s degree, and 5% held a Ph.D. On average, the teachers had 5 years of experience.

Intervention group Intervention teachers implemented the *Pearson Literature*® (2015) curriculum in their English/ language arts classrooms. The *Pearson Literature* (2015) program consisted of five units with four topics per unit. The topics include (1) setting expectations, (2) reading complex texts while providing support and guidance, (3) removing the level of support to “provide a more authentic reading environment for students,” and (4) provide students independence to respond to a range of works. Occasionally, the teachers had to incorporate other resources to meet district requirements.

Comparison group

Comparison teachers were allowed to design their own curriculum or supplement their schools' available curriculum as they saw fit, following their schools' policies. They were also encouraged to use teacher- and district-created resources available online to all teachers. In general, the comparison curriculum consisted of 12 chapters with the following features: (1) reading skills and strategies, (2) making meanings (critical thinking about texts), (3) writer's notebook (writing notes about text), and (4) grammar handbook (practicing grammar skills).

Outcomes and measurement

Outcomes were measured in spring 2015, and the pretest was administered in the fall of 2014. The WWC reviewed the Iowa Form E Reading and Vocabulary subtests under the comprehension domain. An outcome in the general reading achievement domain was measured using the Iowa Form E Overall English/Language Arts (ELA) score. For a more detailed description of these outcome measures, see Appendix B.

Subscale scores on usage and grammar, sentence structure, and mechanics are reported in Appendix D and do not factor into the intervention's rating of effectiveness. Supplemental findings are also presented for subgroups of non-White students and students at a below-average achievement level.

The authors also reported findings for Iowa Form E reading subscales that did not meet WWC reliability requirements and written expression subscales that were not eligible under the Adolescent Literacy review protocol, version 3.0.

Support for implementation

At the beginning of the 2014–15 school year, teachers were provided with about 6 hours of training by a professional trainer in the use of the *Pearson Literature*® curriculum materials. The training consisted of an overview of all program components, including the technology component, *Pearson Realize*. In addition, they were provided with detailed implementation guidelines. Researchers from Pearson used a classroom observation form to measure how faithfully they were following the program. Additional trainings were held in November and January of the same school year to cover more specific details on upcoming units.

Appendix A.2: Research details for Berry et al. (2007)

Berry, T., Eddy, R. M., Fleischer, D., Asgarian, M., & Malek, Y. (2007). *The effects of Prentice Hall Literature (Penguin Edition) curriculum on student performance: Randomized control trial final report*. Claremont, CA: Claremont Graduate University.

Table A2. Summary of findings

Meets WWC group design standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General literacy achievement	1,626 students/7 schools	+2	No
Comprehension	1,627 students/7 schools	-2	No

Setting

The study took place in three high schools and four middle schools in California, Colorado, and Illinois in the 2006–07 school year.

Study sample

The authors used a cluster randomized controlled trial design to study the effects of the *Prentice Hall Literature: Penguin Edition*® curriculum on English language arts achievement for students in grades 7 and 9. Researchers contacted schools to participate in the study that were socioeconomically diverse, had low student mobility rates, were willing to randomly assign teachers to study groups, and had enrollments of at least 750 students or had at least four teachers with multiple sections of college-preparatory English language arts. From the contacted schools, the study recruited seven schools that agreed to participate. In summer 2006, the authors randomly assigned within schools 31 teachers of 1,922 students to conditions, with 15 teachers in the intervention group and 16 teachers in the comparison group. Of those randomly assigned, 13 teachers (six intervention and seven comparison) and 867 students (463 intervention and 404 comparison) were in grade 7, and 18 teachers (nine per condition) and 1,055 students (553 intervention and 502 comparison) were in grade 9.

The WWC considers random assignment jeopardized, however, for two reasons. First, the analysis excluded data on students who changed study conditions after random assignment (from intervention to comparison or vice versa). Second, the analysis included data on students who were added to study classroom rosters after random assignment (by the beginning of fall 2006).

Across the seventh- and ninth-grade analytic samples, students were 47% male, 45% non-Latino White, 24% Latino, 10% African American, 4% Asian, and 16% multi-ethnic or other race or ethnicity. About 92% of students spoke English as their primary language, while the remainder spoke another language. The study did not report characteristics of each of the analytic samples. The analytic sample included 1,627 students: 890 students were in the *Pearson Literature* group (414 students in grade 7 and 476 students in grade 9), and 737 students were in the comparison group (312 students in grade 7 and 425 students in grade 9).

Intervention group

Students in intervention classrooms received English language arts instruction using *Prentice Hall Literature: Penguin Edition*® during the 2006–07 school year. The study provided intervention classrooms with teacher and student textbooks and ancillary materials, such as student notebooks and workbooks, strategy kits, and teaching resource books.

The intervention was delivered over, on average, a 32-week school year, but not all weeks could be used to implement the curriculum because of standardized testing and other school events. On average, seventh-grade teachers covered nine out of the 12 possible sections within a unit and approximately five of the six possible units. Unlike seventh-grade teachers, ninth-grade teachers varied substantially in intervention implementation. In grade 9, California teachers covered, on average, nearly six sections within a unit, while teachers from the other states covered nine sections within a unit. Similar to grade 7, grade 9 teachers covered, on average, nearly five of the six possible units.

Comparison group

Teachers in the comparison group used six different curricula across the seven schools: two schools used an earlier (2002) edition of a *Prentice Hall Literature* textbook, while five schools used other textbooks published between 1985 and 2002. All but one of these textbooks provided additional elements beyond reading selections and related activities (e.g., discussion questions or vocabulary), such as writing exercises and standardized test practice. The oldest textbook (published in 1985) was unique in its predominant focus on reading selections and related activities with almost no other elements.

All comparison teachers supplemented their textbooks with outside reading selections, such as novels, poetries, biographies, or articles and handouts, with two comparison teachers using more outside reading selections than textbook selections. Teachers in the comparison condition differentiated instruction through various techniques such as adapting assignments or through mixed-ability group work, but only sometimes using the textbook to differentiate instruction.

Outcomes and measurement

Outcomes were measured in May 2007, and the pretest was administered in September 2006. For grade 7 students, the authors used the Iowa Test of Basic Skills (ITBS) and analyzed separately two subtests, the ITBS Reading Comprehension subtest (comprehension domain) and the Language subtest (general literacy achievement domain). For grade 9 students, the authors used the Iowa Test of Educational Development (ITED). They administered and analyzed separately the ITED Reading Comprehension subtest (comprehension domain) and the Spelling subtest (general literacy achievement domain). For a more detailed description of these outcome measures, see Appendix B.

The authors also administered a student writing assessment (a timed persuasive essay using a prompt from the Iowa Writing Assessment) and a student survey on attitudes toward English language arts, parent involvement, the classroom environment, and background characteristics. These outcomes are not eligible under the Adolescent Literacy review protocol (version 3.0).

Support for implementation

Intervention teachers participated in a 3–4 hour training administered by a Prentice Hall consultant and or the study authors. The consultant/study authors reviewed the curriculum, implementation guidelines, and all ancillary materials. After teachers began using the products, a consultant returned to each site to hold a question and answer session.

Appendix A.3: Research details for Eddy et al. (2010)

Eddy, R. M., Ruitman, H. T., Hanken, N., & Sloper, M. (2010). *The effects of Pearson Prentice Hall Literature (2010) on student performance: Efficacy study*. La Verne, CA: Cobblestone Applied Research and Evaluation, Inc.

Table A3. Summary of findings

Meets WWC group design standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Comprehension	1,518 students/8 schools	-3	No

Setting

The study took place in eight schools in Arizona, California, Ohio, and Oregon. Three grade levels (7, 8, 10) were included in the study.

Study sample

The study was conducted in the 2009–10 school year. Both the developer and the evaluator nominated potential study schools, beginning in February 2009 and continuing through summer 2009. Recruitment was targeted to schools with diverse student ethnicity and lower socioeconomic status that had at least two teachers with multiple sections of language arts or English classes. In the summer of 2009, within each of the eight participating schools, teachers were randomly assigned either to implement *Prentice Hall Literature*® (2010) or to implement the regular curriculum (the comparison group). Altogether, there were 16 teachers randomly assigned to the intervention group and 13 teachers assigned to the comparison group. The WWC determined that the study’s randomized controlled trial design was jeopardized because the analytic sample included students who moved into the study classrooms after random assignment. The analysis sample included 1,518 students: 744 students were in the *Prentice Hall Literature* group, and 774 students were in the comparison group.

The pretest for the study was conducted in August or September 2009, depending on the start date of each school. The intervention began in August 2009 at seven schools and in September 2009 at one school. The eight schools included six suburban schools with at least 1,200 students in each school, and two rural schools with at least 700 students in each school. Seven of the eight schools had at least 35% of students who were eligible for free or reduced-price lunch. Most communities had median household incomes between \$30,000 and \$60,000. Students were 52% male and 48% female; 55% Hispanic, 22% White, 15% African American, 3% American Indian, 1% Asian, and 3% multiracial; and 86% spoke English as their primary language.

Intervention group

The intervention teachers implemented *Prentice Hall Literature*® (2010). The intervention generally includes six units focused on a specific genre for each grade level (e.g., nonfiction, fiction, poetry, etc.). Instruction is organized by a “Big Question” which is introduced at the beginning of each unit and revisited throughout the unit to reinforce concepts. *Prentice Hall Literature*® (2010) includes paired reading selections of differing difficulty so instruction can be tailored to students’ ability level. Ancillary materials are available to teachers to further enhance instruction of students of different ability levels. The *Reality Central* textbook and accompanying writing journal, for example, provide students with additional reading practice below grade level. In the present study, participating teachers were instructed to implement Units 1–6 throughout the school year, and the *Reality Central* textbook and other supplementary materials (e.g., study workbooks) were made available.

Comparison group

Comparison teachers implemented their normal language arts curriculum. Most (10 out of 13) comparison teachers used a textbook to guide instruction, and followed district pacing guidelines so specific material would be covered ahead of state testing. Many of these 10 teachers also supplemented textbook instruction with their own writing and vocabulary activities. The remaining three comparison teachers did not use a textbook to guide instruction, and instead read novels and short stories followed by activities that the teachers either created themselves or found on the Internet.

Outcomes and measurement

Outcomes were measured in the spring of 2010, and the pretest was administered in the fall of 2009. The authors reported a total reading score on the Gates MacGinitie Reading Test (GMRT) for the combined sample of students in grades 7, 8, and 10. The outcome was reviewed in the comprehension domain. For a more detailed description of this outcome measure, see Appendix B.

Supplemental findings are presented for students in grade 10. The supplemental findings are reported in Appendix D and do not factor into the intervention's rating of effectiveness.

The authors also examined scores on the Metropolitan Achievement Test, 8th Edition (an assessment of writing), administered a survey on student attitudes toward reading and learning, and examined teacher implementation measures. These outcomes are not eligible under the Adolescent Literacy review protocol (version 3.0).

Support for implementation

All participating schools received training prior to the start of the study, in August or September 2009, depending on the timing of implementation. All participating teachers in the intervention group received a 2-day training on *Prentice Hall Literature*® (2010) prior to the start of the school year to review program components and learn about online features of the program. A follow-up training was also held a few weeks into the school year. All teachers in the intervention group received the teacher's edition textbook, student textbooks, and all available ancillary materials.

Appendix B: Outcome measures for each domain

General literacy achievement	
<i>Iowa Assessment Form E (Iowa Form E) - Overall ELA</i>	This standardized assessment combines results from reading, vocabulary, and written expression Iowa Form E subtests. It includes 40 reading items, 54 written expression items, and 40 vocabulary items (as cited in Resendez et al., 2015).
<i>Iowa Form E Skill: Mechanics</i>	This standardized subtest evaluates student understanding of writing conventions and syntax. Percent correct scores are provided for Iowa skill domain areas (as cited in Resendez et al., 2015). This outcome is only reported as a supplemental finding.
<i>Iowa Form E Skill: Sentence Structure</i>	This standardized subtest evaluates student understanding of sentence formation, including complexity, and the associated grammatical rules. Percent correct scores are provided for Iowa skill domain areas (as cited in Resendez et al., 2015). This outcome is only reported as a supplemental finding.
<i>Iowa Form E Skill: Usage & Grammar</i>	This standardized test evaluates student understanding of appropriate grammatical rules within text. Percent correct scores are provided for Iowa skill domain areas (as cited in Resendez et al., 2015). This outcome is only reported as a supplemental finding.
<i>Iowa Test of Basic Skills (ITBS), Language</i>	This standardized outcome is an English language arts assessment, which provides a measure of the basic language skills of grade 7 students (as cited in Berry et al., 2007).
<i>Iowa Test of Educational Development (ITED) Spelling</i>	This standardized test is an English language arts assessment, which measures the spelling ability of grade 9 students. The test presents students with groups of words; students must indicate which word is misspelled or whether they are all spelled correctly (as cited in Berry et al., 2007).
Comprehension	
<i>Gates-MacGinitie Reading Tests (GMRT): Total</i>	The outcome is a norm-referenced assessment that combines two subtests, Vocabulary Knowledge (45 items) and Reading Comprehension (48 items), to form a Total Reading score (as cited in Eddy et al., 2010).
<i>ITED, Reading</i>	The outcome is a standardized English language arts assessment, which provides a measure of the reading skills of grade 9 students. Although assessed separately in the ITED, scores from the Reading Comprehension and Vocabulary Skills were combined to assess an overall rating of reading in the study (as cited in Berry et al., 2007).
Reading Comprehension	
<i>Iowa Form E - Reading</i>	This standardized test of 40 items assesses students' comprehension skills in different contexts and genres, such as reading magazine and newspaper articles, and evaluating ideas from a variety of sources for research projects (as cited in Resendez et al., 2015).
<i>ITBS, Reading Comprehension</i>	This outcome is a standardized English language arts assessment, which provides a measure of the reading comprehension skills of grade 7 students (as cited in Berry et al., 2007).
Vocabulary Development	
<i>Iowa Form E - Vocabulary</i>	This standardized test of 40 items assesses general vocabulary development and students' ability to recognize words with similar meanings (as cited in Resendez et al., 2015).

Appendix C.1: Findings included in the rating for the general literacy achievement domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Resendez & Azin (2015)^a								
<i>Iowa Form E - Overall ELA</i>	Grade 9	19 teachers/ 932 students	267.31 (37.25)	265.63 (37.71)	1.68	0.05	+2	.72
Domain average for general literacy achievement (Resendez & Azin, 2015)						0.05	+2	Not statistically significant
Berry et al. (2007)^b								
<i>Iowa Test of Basic Skills, Language</i>	Grade 7	13 teachers/ 722 students	43.74 (27.26)	42.48 (25.48)	1.26	0.05	+2	.46
<i>Iowa Test of Educational Development, Spelling</i>	Grade 9	18 teachers/ 904 students	51.38 (26.12)	50.00 (25.63)	1.38	0.05	+2	.53
Domain average for general literacy achievement (Berry et al., 2007)						0.05	+2	Not statistically significant
Domain average for general literacy achievement across all studies						0.05	+2	na

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

^a For Resendez and Azin (2015), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-value presented here was reported in the original study. The WWC calculated the intervention group mean by adding the impact of the intervention (the hierarchical linear modeling [HLM] level-2 coefficient) to the unadjusted comparison group posttest means. This study is characterized as having an indeterminate effect because the effect for the measure in this domain was neither statistically significant nor large enough to be substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

^b For Berry et al. (2007), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-values presented here were reported in the original study. The unadjusted posttest means and standard deviations in the table were obtained through an author query, and the author query confirmed that the numbers were for the same sample presented in the HLM analysis in the study. The reported intervention group means are calculated as the comparison group means plus the HLM level-2 coefficient. This study is characterized as having an indeterminate effect because the effect for the one measure in this domain was neither statistically significant nor large enough to be substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix C.2: Findings included in the rating for the comprehension domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Resendez & Azin (2015)^a								
<i>Iowa Form E - Reading</i>	Grade 9	19 teachers/ 1,002 students	263.87 (43.39)	265.41 (40.09)	-1.54	-0.04	-1	.78
<i>Iowa Form E - Vocabulary</i>	Grade 9	19 teachers/ 1,004 students	263.77 (31.19)	268.52 (29.58)	-4.75	-0.16	-6	.16
Domain average for comprehension (Resendez & Azin, 2015)						-0.10	-4	Not statistically significant
Berry et al. (2007)^b								
<i>Iowa Test of Basic Skills, Reading Comprehension</i>	Grade 7	13 teachers/ 726 students	47.21 (28.54)	48.19 (24.57)	-0.98	-0.04	-1	.55
<i>Iowa Test of Educational Development, Reading</i>	Grade 9	18 teachers/ 901 students	45.45 (24.91)	47.58 (25.58)	-2.13	-0.08	-3	.35
Domain average for comprehension (Berry et al., 2007)						-0.06	-2	Not statistically significant
Eddy et al. (2010)^c								
<i>Gates-MacGinitie Reading Tests: Total</i>	Grades 7, 8, 10	29 teachers/ 1,518 students	532.49 (29.22)	534.85 (28.16)	-2.36	-0.08	-3	.64
Domain average for comprehension (Eddy et al., 2010)						-0.08	-3	Not statistically significant
Domain average for comprehension across all studies						-0.08	-3	na

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

^a For Resendez and Azin (2015), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-values presented here were reported in the original study. The WWC calculated the intervention group mean by adding the impact of the intervention (the HLM level-2 coefficient) to the unadjusted comparison group posttest means. This study is characterized as having an indeterminate effect because the effect for the measures in this domain was neither statistically significant nor large enough to be substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

^b For Berry et al. (2007), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The p-values presented here were reported in the original study. The unadjusted posttest means and standard deviations in the table were obtained through an author query, and the author query confirmed that the numbers were for the same sample presented in the HLM analysis in the study. The reported intervention group means are calculated as the comparison group means plus the HLM level-2 coefficient. This study is characterized as having an indeterminate effect because the effect for the two measures in this domain was neither statistically significant nor large enough to be substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

^c For Eddy et al. (2010), the p-value presented here was calculated by the WWC. A correction for clustering was needed but did not affect whether the contrast was found to be statistically significant. The unadjusted posttest means and standard deviations in the table were obtained through an author query. The WWC calculated the intervention group mean using a difference-in-differences approach by adding the impact of the intervention (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. This study is characterized as having an indeterminate effect because the effect for the measure in this domain was neither statistically significant nor large enough to be substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix D.1: Description of supplemental findings for the general literacy achievement domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Resendez & Azin (2015)^a								
<i>Iowa Form E - Overall ELA</i>	Non-White	19 teachers/ 216 students	252.13 (31.83)	252.40 (36.65)	-0.27	-0.01	0	.07
<i>Iowa Form E - Skill: Mechanics</i>	Grade 9	19 teachers/ 1,001 students	64.29 (24.75)	59.64 (27.21)	4.65	0.18	+7	.03
<i>Iowa Form E - Skill: Sentence Structure</i>	Grade 9	19 teachers/ 1,001 students	59.30 (24.82)	54.80 (25.24)	4.50	0.18	+7	.00
<i>Iowa Form E - Skill: Usage and Grammar</i>	Grade 9	19 teachers/ 1,001 students	52.98 (21.49)	50.65 (23.80)	2.33	0.10	+4	.22

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

^a For Resendez and Azin (2015), the p-values presented here were reported in the original study. Corrections for clustering were needed and resulted in a WWC-computed p-value of .45 for the Iowa Form E - Skill: Mechanics measure and .49 for the Iowa Form E - Skill: Sentence Structure measure; therefore, the WWC does not find the results for the Iowa Form E - Skill: Mechanics and the Iowa Form E - Skill: Sentence Structure outcomes to be statistically significant. For the Iowa Form E - Overall ELA measure, the unadjusted posttest mean and standard deviation in the table for the comparison group were obtained through an author query. The WWC calculated the intervention group mean using a difference-in-differences approach by adding the impact of the intervention (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0), p. 23 for more information.

Appendix D.2: Description of supplemental findings for the comprehension domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Resendez & Azin (2015)^a								
<i>Iowa Form E - Reading</i>	Non-White	19 teachers/ 233 students	249.32 (38.36)	253.58 (41.33)	-4.26	-0.11	-4	.16
<i>Iowa Form E - Vocabulary</i>	Below-average baseline reading/writing levels	19 teachers/ 274 students	240.65 (25.77)	245.20 (27.73)	-4.55	-0.17	-7	.05
Eddy et al. (2010)^b								
<i>Gates-MacGinitie Reading Tests: Total</i>	Grade 10	13 teachers/ 591 students	550.66 (22.12)	549.46 (23.03)	1.20	.05	2	.84

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

^a For Resendez and Azin (2015), the p -values presented here were reported in the original study. A correction for clustering was needed and resulted in a WWC-computed p -value of .48 for the Iowa Form E - Vocabulary: Below-average baseline reading/writing levels; therefore, the WWC does not find the result to be statistically significant. The unadjusted posttest means and standard deviations in the table for the comparison group were obtained through an author query. The WWC calculated the intervention group mean using a difference-in-differences approach by adding the impact of the intervention (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0), p. 23 for more information.

^b For Eddy et al. (2010), the p -value presented here was calculated by the WWC. A correction for clustering was needed but did not affect whether the contrast was found to be statistically significant. The unadjusted posttest means and standard deviations in the table were obtained through an author query. The WWC calculated the intervention group mean using a difference-in-differences approach by adding the impact of the intervention (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0), p. 23 for more information.

Endnotes

¹ The descriptive information for this intervention comes from publically available sources: Pearson's website for *Pearson Literature*® (2015) and the video overview on the *Prentice Hall Literature*® (2010) site, which was accessed online (January 2017): http://www.pearsonschool.com/live/customer_central/microsite/connectedsampling/overview/nat/lit/player.html. The What Works Clearinghouse (WWC) requests developers review the intervention description sections for accuracy from their perspective. The WWC provided the developer with the intervention description in January 2017, and the WWC incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this intervention is beyond the scope of this review.

² The literature search reflects documents publicly available by October 2016. Reviews of the studies in this report used the standards from the WWC Procedures and Standards Handbook (version 3.0) and the Adolescent Literacy review protocol (version 3.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

³ Please see the Adolescent Literacy review protocol (version 3.0) for a list of all outcome domains.

⁴ For criteria used to determine the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 20. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the studies.

⁵ The WWC published a separate intervention report called *Prentice Hall Literature*® (1989–2005), which covers earlier versions of this intervention: <https://ies.ed.gov/ncee/wwc/InterventionReport/680>. This separate report was generated because the developer noted that the intervention fundamentally changed with the introduction of *Prentice Hall Literature: Penguin Edition*® (2007). The WWC confirmed the decision to separate evidence on *Prentice Hall Literature*® into two intervention reports with the WWC literacy content expert. The companion report covers *Prentice Hall Literature*® (1989) and *Prentice Hall Literature: Timeless Voices, Timeless Themes*® (2000, 2002, 2005).

⁶ The WWC Reviewer Guidance, for use with the WWC Procedures and Standards Handbook (version 3.0), indicates that if the authors of a cluster randomized controlled trial study characterize the intervention as having effects on student scores (rather than only on cluster-level scores), and some students enter clusters after random assignment, then the study must demonstrate equivalence of the analytic intervention and comparison groups at baseline.

⁷ The WWC Reviewer Guidance, for use with the WWC Procedures and Standards Handbook (version 3.0), indicates that if the authors of a cluster randomized controlled trial study characterize the intervention as having effects on student scores (rather than only on cluster-level scores), and some students enter clusters after random assignment, then the study must demonstrate equivalence of the analytic intervention and comparison groups at baseline. Note that separate impact analyses for students in grades 7 and 8 did not meet WWC group design standards because the study (Eddy et al., 2010) did not establish baseline equivalence for the intervention and comparison groups.

Recommended Citation

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, November). *Adolescent Literacy intervention report: Prentice Hall/Pearson Literature*® (2007–15). Retrieved from <https://whatworks.ed.gov>

WWC Rating Criteria

Criteria used to determine the rating of a study

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

Criteria used to determine the rating of effectiveness for an intervention

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

Criteria used to determine the extent of evidence for an intervention

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

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all subjects initially assigned to the intervention and comparison groups. If a randomized controlled trial (RCT) or regression discontinuity design (RDD) study has high levels of attrition, the validity of the study results can be called into question. An RCT with high attrition cannot receive the highest rating of *meets WWC group design standards without reservations*, but can receive a rating of *meets WWC group design standards with reservations* if it establishes baseline equivalence of the analytic sample. Similarly, the highest rating an RDD with high attrition can receive is *meets WWC RDD standards with reservations*.

For single-case design research, attrition occurs when an individual fails to complete all required phases or data points in an experiment, or when the case is a group and individuals leave the group. If a single-case design does not meet minimum requirements for phases and data points within phases, the study cannot receive the highest rating of *meets WWC pilot single-case design standards without reservations*.

Baseline A point in time before the intervention was implemented in group design research and in regression discontinuity design studies. When a study is required to satisfy the baseline equivalence requirement, it must be done with characteristics of the analytic sample at baseline. In a single-case design experiment, the baseline condition is a period during which participants are not receiving the intervention.

Clustering adjustment An adjustment to the statistical significance of a finding when the units of assignment and analysis differ. When random assignment is carried out at the cluster level, outcomes for individual units within the same clusters may be correlated. When the analysis is conducted at the individual level rather than the cluster level, there is a mismatch between the unit of assignment and the unit of analysis, and this correlation must be accounted for when assessing the statistical significance of an impact estimate. If the correlation is not accounted for in a mismatched analysis, the study may be too likely to report statistically significant findings. To fairly assess an intervention's effects, in cases where study authors have not corrected for the clustering, the WWC applies an adjustment for clustering when reporting statistical significance.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The method by which intervention and comparison groups are assigned (group design and regression discontinuity design) or the method by which an outcome measure is assessed repeatedly within and across different phases that are defined by the presence or absence of an intervention (single-case design). Designs eligible for WWC review are randomized controlled trials, quasi-experimental designs, regression discontinuity designs, and single-case designs.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analytic sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence from group design studies supports the findings in an intervention report. The extent of evidence categorization for intervention reports focuses on the number and sizes of studies of the intervention in order to give an indication of how broadly findings may be applied to different settings. There are two extent of evidence categories: small and medium to large.

- **small:** includes only one study, or one school, or findings based on a total sample size of less than 350 students and 14 classrooms (assuming 25 students in a class)
- **medium to large:** includes more than one study, more than one school, and findings based on a total sample of at least 350 students or 14 classrooms

Gain scores The result of subtracting the pretest from the posttest for each individual in the sample. Some studies analyze gain scores instead of the unadjusted outcome measure as a method of accounting for the baseline measure when estimating the effect of an intervention. The WWC reviews and reports findings from analyses of gain scores, but gain scores do not satisfy the WWC's requirement for a statistical adjustment under the baseline equivalence requirement. This means that a study that must satisfy the baseline equivalence requirement and has baseline differences between 0.05 and 0.25 standard deviations *does not meet WWC group design standards* if the study's only adjustment for the baseline measure was in the construction of the gain score.

Group design A study design in which outcomes for a group receiving an intervention are compared to those for a group not receiving the intervention. Comparison group designs eligible for WWC review are randomized controlled trials and quasi-experimental designs.

Improvement index Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from -50 to +50.

Intervention An educational program, product, practice, or policy aimed at improving student outcomes.

Intervention report A summary of the findings of the highest-quality research on a given program, product, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against design standards, and summarizes the findings of those that meet WWC design standards.

Multiple comparison adjustment An adjustment to the statistical significance of results to account for multiple comparisons in a group design study. The WWC uses the Benjamini-Hochberg (BH) correction to adjust the statistical significance of results within an outcome domain when study authors perform multiple hypothesis tests without adjusting the p -value. The BH correction is used in three types of situations: studies that tested multiple outcome measures in the same outcome domain with a single comparison group; studies that tested a given outcome measure with multiple comparison groups; and studies that tested multiple outcome measures in the same outcome domain with multiple comparison groups. Because repeated tests of highly correlated constructs will lead to a greater likelihood of mistakenly concluding that the impact was different from zero, in all three situations, the WWC uses the BH correction to reduce the possibility of making this error. The WWC makes separate adjustments for primary and secondary findings.

Outcome domain	A group of closely-related outcomes. A domain is the organizing construct for a set of related outcomes through which studies claim effectiveness.
Quasi-experimental design (QED)	A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.
Randomized controlled trial (RCT)	A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.
Rating of effectiveness	For group design research, the WWC rates the effectiveness of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. For single-case design research, the WWC rates the effectiveness of an intervention in each domain based on the quality of the research design and the consistency of demonstrated effects. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 20.
Regression discontinuity design (RDD)	A design in which groups are created using a continuous scoring rule. For example, students may be assigned to a summer school program if they score below a preset point on a standardized test, or schools may be awarded a grant based on their score on an application. A regression line or curve is estimated for the intervention group and similarly for the comparison group, and an effect occurs if there is a discontinuity in the two regression lines at the cutoff.
Single-case design	A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.
Standard deviation	The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.
Statistical significance	Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < .05$).
Study rating	The result of the WWC assessment of a study. The rating is based on the strength of the evidence of the effectiveness of the educational intervention. Studies are given a rating of <i>meets WWC design standards without reservations</i> , <i>meets WWC design standards with reservations</i> , or <i>does not meet WWC design standards</i> , based on the assessment of the study against the appropriate design standards. The WWC has design standards for group design, single-case design, and regression discontinuity design studies.
Substantively important	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.
Systematic review	A review of existing literature on a topic that is identified and reviewed using explicit methods. A WWC systematic review has five steps: 1) developing a review protocol; 2) searching the literature; 3) reviewing studies, including screening studies for eligibility, reviewing the methodological quality of each study, and reporting on high quality studies and their findings; 4) combining findings within and across studies; and, 5) summarizing the review.

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



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

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