

What Works Clearinghouse™



SpellRead™

Program Description¹

SpellRead™, formerly known as *SpellRead Phonological Auditory Training®*, is a small-group literacy program for struggling readers in grades 2–12. *SpellRead™* integrates the auditory and visual aspects of the reading process and emphasizes specific skill mastery through systematic and explicit instruction. Students are taught to recognize and manipulate English sounds; to practice, apply, and transfer their skills using texts at their reading level; and to write about their reading.

Research²

Two studies of *SpellRead™* that fall within the scope of the Adolescent Literacy review protocol meet What Works Clearinghouse (WWC) evidence standards without reservations.³ The two studies included 137 adolescent readers in grades 5 and 6 in Pennsylvania and Newfoundland, Canada. Based on these two studies, the WWC considers the extent of evidence for *SpellRead™* on the reading performance of adolescent readers to be small for alphabetics, reading fluency, and comprehension. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *SpellRead™* in the general literacy achievement domain. (See the Effectiveness Summary on p. 4 for further description of all domains.)

Effectiveness

SpellRead™ was found to have potentially positive effects on alphabetics, reading fluency, 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			
Alphabetics	Potentially positive effects	+21	–9 to +49	2	137	Small
Reading fluency	Potentially positive effects	+14	+3 to +32	2	137	Small
Comprehension	Potentially positive effects	+11	–2 to +24	2	137	Small

Report Contents	
Overview	p. 1
Program Information	p. 2
Research Summary	p. 3
Effectiveness Summary	p. 4
References	p. 7
Research Details for Each Study	p. 9
Outcome Measures for Each Domain	p. 13
Findings Included in the Rating for Each Outcome Domain	p. 14
Supplemental Findings for Each Outcome Domain	p. 18
Endnotes	p. 21
Rating Criteria	p. 22
Glossary of Terms	p. 23

Program Information

Background

SpellRead[™] is distributed through PCI Education. Address: 4560 Lockhill Selma Rd., Ste. 100, San Antonio, TX, 78249-2075. Web: <http://www.pcieducation.com/spellread/default.aspx>. Telephone: (800) 594-4263.

Program details

SpellRead[™] consists of 105 lessons implemented in three distinct phases that interweave phonemics, phonetics, and instruction in language-based reading and writing. The program takes five to nine months to complete and can be implemented at any grade from 2 to 12. *SpellRead*[™] Libraries, which contain accompanying readers and trade books, are tailored to each grade level. Phase A has 50 lessons designed to train the auditory process function of the brain to hear and manipulate the 44 sounds of the English language. Phase B, which has 30 lessons, focuses on secondary vowel spelling, consonant blends, and decoding two-syllable words. Phase C has 25 lessons and concentrates on how to decode words of three or more syllables, as well as clusters and verb forms. The *SpellRead*[™] program is used with small groups of five students and one instructor in 60–90 minute classes. The daily instructional cycle includes linguistic foundations, active reading, and writing connections to develop reading comprehension, vocabulary, and fluency skills. Linguistic foundation activities focus on phonics and phonemic awareness, active reading emphasizes oral-reading practice using texts at students' reading levels, and writing connection activities focus on links between oral and written language.

SpellRead[™] includes professional development and ongoing support for educators as they implement the program, including five days of initial workshops, two follow-up workshops, and regular onsite coaching visits. A web-based instructor support system allows educators to closely monitor student progress.

Cost

The cost of implementing *SpellRead*[™] varies based on the number of participating students and the number of teachers or schools participating in the program. The cost for a complete set of materials for five participating students is \$999.95. One complete set of teacher materials costs \$1,495.95. Additional information can be found on the distributor's website.

Research Summary

The WWC identified 14 studies on the effects of *SpellRead*[™] on the reading achievement of adolescent readers. Two studies (Rashotte, MacPhee, & Torgesen, 2001; Torgesen et al., 2006) are randomized controlled trials that meet WWC evidence standards without reservations. These two studies are summarized in this report. The remaining 12 studies do not meet either WWC eligibility screens or evidence standards. (See references beginning on p. 7 for citations for all 14 studies.)

Table 2. Scope of reviewed research

Grade	5, 6
Delivery method	Small group
Program type	Curriculum
Studies reviewed	14
Group design studies that meet WWC evidence standards	
• without reservations	2 studies
• with reservations	0 studies

Summary of studies meeting WWC evidence standards without reservations

Rashotte, MacPhee, & Torgesen (2001) randomly assigned a total of 33 fifth- and sixth-grade students from one school in Newfoundland, Canada to the intervention and comparison groups.⁵ Students in the intervention group received the *SpellRead*[™] program. Students in the comparison group received the regular literacy-based reading program at their school. The study reported student outcomes after two months (eight weeks) of program implementation.

Torgesen et al. (2006) randomly assigned 32 school units⁶ in Allegheny County, Pennsylvania to one of four interventions: *SpellRead*[™],⁷ *Corrective Reading*, *Failure Free Reading*[™], and *Wilson Reading System*[®]. Within each school, eligible students were randomly assigned either to the treatment group that would receive the intervention assigned to that school or to the comparison group that would not receive any of the four interventions. Students were eligible for participation if their teacher identified them as struggling readers and if they scored at or below the 30th percentile on a word-level reading test and at or above the 5th percentile on a vocabulary test. The WWC based its effectiveness ratings on findings from comparisons of the 45 fifth-grade students who received the standard district curriculum and the 59 fifth-grade students who received *SpellRead*[™]. The study reported student outcomes after six months of program implementation.⁸

Summary of studies meeting WWC evidence standards with reservations

No studies of *SpellRead*[™] meet WWC evidence standards with reservations.

Effectiveness Summary

The WWC review of interventions for Adolescent Literacy addresses student outcomes in four domains: alphabets, reading fluency, comprehension, and general literacy achievement. The two studies that contribute to the effectiveness rating in this report cover three domains: alphabets, reading fluency, and comprehension. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *SpellRead*TM on adolescent readers for each domain. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 22.

Summary of effectiveness for the alphabets domain

Two studies reported findings in the alphabets domain.

Rashotte, MacPhee, & Torgesen (2001) examined the following eight outcomes in the alphabets domain: Woodcock Reading Mastery Test–Revised (WRMT-R) Word Identification and Word Attack subtests; Test of Word Reading Efficiency (TOWRE) Phonetic Decoding Efficiency and Sight Word Efficiency subtests; Comprehensive Test of Phonological Processing (CTOPP) Elision, Blending Words, and Segmenting Words subtests; and the Schonell Spelling test.

The authors reported statistically significant positive effects on fifth and sixth graders' scores on seven of eight measured outcomes, the exception being the TOWRE Sight Word Efficiency subtest. The WWC analysis accounted for multiple comparisons and confirmed statistically significant differences only for these four outcomes: WRMT-R Word Attack subtest, TOWRE Phonetic Decoding Efficiency subtest, and CTOPP Blending Words and Segmenting Words subtests. The WWC characterizes these study findings as a statistically significant positive effect, because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant.

Torgesen et al. (2006) examined four outcomes in the phonics construct of the alphabets domain: the WRMT-R Word Identification and Word Attack subtests and the TOWRE Phonemic Decoding Efficiency and Sight Word Efficiency subtests. The authors reported statistically significant effects of *SpellRead*TM on fifth graders' scores on two of these outcomes: the WRMT-R Word Attack subtest and the TOWRE Phonemic Decoding Efficiency subtest. The WWC-calculated estimates of program effects were not statistically significant. The average effect across the four outcomes was not large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25).⁹ The WWC characterizes these study findings as an indeterminate effect.

Thus, for the alphabets domain, one study showed statistically significant positive effects and one study showed indeterminate effects. This results in a domain rating of potentially positive effects, with a small extent of evidence.

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

Rating of effectiveness	Criteria met
Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i>	The review of <i>SpellRead</i> TM in the alphabets domain had one study showing statistically significant positive effects and one study showing indeterminate effects.
Extent of evidence	Criteria met
Small	The review of <i>SpellRead</i> TM in the alphabets domain was based on two studies that included at least nine schools and 137 students. The number of schools was not reported in one of the studies, so the exact number of schools cannot be determined.

Summary of effectiveness for the reading fluency domain

Two studies reported findings in the reading fluency domain.

Rashotte, MacPhee, & Torgesen (2001) examined two outcomes in the reading fluency domain, the Gray Oral Reading Tests, Third Edition (GORT-3) Accuracy and Rate subtests, and reported statistically significant positive effects for both outcomes. Accounting for multiple comparisons, the WWC confirmed a statistically significant difference only for the GORT-3 Rate subtest. The WWC characterizes these study findings as a statistically significant positive effect, because the effect for the GORT-3 Rate subtest is positive and statistically significant, and no effects are negative and statistically significant.

Torgesen et al. (2006) did not find statistically significant effects of *SpellRead*TM on fifth graders' scores on the Oral Reading Fluency test. The WWC-calculated effect was not large enough to be considered substantively important according to WWC criteria. The WWC characterizes this study finding as an indeterminate effect.

Thus, for the reading fluency domain, one study showed statistically significant positive effects and one study showed indeterminate effects. This results in a domain rating of potentially positive effects, with a small extent of evidence.

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

Rating of effectiveness	Criteria met
Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i>	The review of <i>SpellRead</i> TM in the reading fluency domain had one study showing statistically significant positive effects and one study showing indeterminate effects.
Extent of evidence	Criteria met
Small	The review of <i>SpellRead</i> TM in the reading fluency domain was based on two studies that included at least nine schools and 137 students. The number of schools was not reported in one of the studies, so the exact number of schools cannot be determined.

Summary of effectiveness for the comprehension domain

Two studies reported findings in the comprehension domain.

Rashotte, MacPhee, & Torgesen (2001) examined two outcomes in the comprehension domain, the Woodcock Diagnostic Reading Battery (WDRB) Passage Comprehension subtest and the GORT-3 Comprehension subtest. The authors reported statistically significant effects for both outcomes. Although the WWC could not confirm the statistical significance of the findings, the average effect size across the two outcomes was large enough to be considered substantively important. Thus, the WWC characterizes these study findings as a substantively important positive effect.

Torgesen et al. (2006) examined two outcomes in this domain: the WRMT-R Passage Comprehension subtest and the Group Reading Assessment and Diagnostic Evaluation (GRADE) Passage Comprehension subtest, and reported no statistically significant effects. The average effect size across the two outcomes was neither statistically significant nor large enough to be considered substantively important. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the comprehension domain, one study showed substantively important positive effects and one study showed indeterminate effects. This results in a domain rating of potentially positive effects, with a small extent of evidence.

Table 5. Rating of effectiveness and extent of evidence for the comprehension domain

Rating of effectiveness	Criteria met
Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i>	The review of <i>SpellRead</i> TM in the comprehension domain had one study showing substantively important positive effects and one study showing indeterminate effects.
Extent of evidence	Criteria met
Small	The review of <i>SpellRead</i> TM in the comprehension domain was based on two studies that included at least nine schools and 137 students. The number of schools was not reported in one of the studies, so the exact number of schools cannot be determined.

References

Studies that meet WWC evidence standards without reservations

- Rashotte, C. A., MacPhee, K., & Torgesen, J. K. (2001). The effectiveness of a group reading instruction program with poor readers in multiple grades. *Learning Disability Quarterly, 24*(2), 119–134.
- Torgesen, J., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., ...Haan, C. (2006). *National assessment of Title I. Interim report. Volume II: Closing the reading gap: First year findings from a randomized trial of four reading interventions for striving readers*. Washington, DC: National Center for Education Evaluation and Regional Assistance.
- Additional source:**
- Torgesen, J., Schirm, A., Castner, L., Vartivarian, S., Mansfield, W., Myers, D., ...Haan, C. (2007). *National assessment of Title I. Final report. Volume II: Closing the reading gap: Findings from a randomized trial of four reading interventions for striving readers* (NCEE 2008-4013). Washington, DC: National Center for Education Evaluation and Regional Assistance.

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

- 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*. New York: Carnegie Corporation of New York. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Gray, E. S. (2008). Understanding dyslexia and its instructional implications: A case to support intense intervention. *Literacy Research & Instruction, 47*(2), 116–123. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Malmgren, K. W., & Trezek, B. J. (2009). Literacy instruction for secondary students with disabilities. *Focus on Exceptional Children, 41*(6), 1–12. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- PCI Education. (n.d.). *SpellRead research summary*. San Antonio, TX: Author. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Rashotte, C. A. (2001). *The effectiveness of a phonologically-based reading program on the reading skills of incarcerated youths at the Newfoundland and Labrador Youth Centre at Whitbourne*. Tallahassee: Florida State University. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- 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 a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research, 79*(4), 1391–1466. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2009). *Effective programs for struggling readers: A best evidence synthesis*. Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- SpellRead P.A.T. (2002). *Maryland middle school success*. Ellicott City, MD: Author. The study is ineligible for review because it does not provide enough information about its design to assess whether it meets standards.

- SpellRead P.A.T. (2004). *An effective reading intervention: Research evidence of SpellRead results*. Ellicott City, MD: Author. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Torgesen, J. K. (2005). Remedial interventions for students with dyslexia: National goals and current accomplishments. In S. Richardson & J. Gilger (Eds.), *Research-based education and intervention: What we need to know* (pp. 103–124). Boston, MA: International Dyslexia Association. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Torgesen, J. K., Rashotte, C. A., Alexander, A. W., Alexander, J., & MacPhee, K. (2003). Progress towards understanding the instructional conditions necessary for remediating reading difficulties in older children. In B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 275–298). Parkton, MD: York Press. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Appendix A.1: Research details for Rashotte, MacPhee, & Torgesen, 2001

Rashotte, C. A., MacPhee, K., & Torgesen, J. K. (2001). The effectiveness of a group reading instruction program with poor readers in multiple grades. *Learning Disability Quarterly, 24*(2), 119–134.

Table A1. Summary of findings

Meets WWC evidence standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Alphabetics	33 students	+35	Yes
Reading fluency	33 students	+24	Yes
Comprehension	33 students	+20	No

Setting The study took place in an elementary school in Newfoundland, Canada.

Study sample The study included 116 students from grades 1–6 with below-average phonetic decoding and word-level reading skills (as measured by the Word Attack and Word Identification subtests of the Woodcock Reading Mastery Test–Revised [WRMT-R]). This WWC report focuses on 33 fifth- and sixth-grade students. Students were matched on phonemic decoding and word-level skills at each grade level, with one of each pair randomly assigned to *SpellRead*TM, and the other assigned to the comparison condition. Most of the students in the sample were from low-income families, and all were White.

Intervention group *SpellRead*TM was implemented in small groups of three to five students outside of the regular classroom. The comparison group remained in class during this period receiving the regular reading program. The students received 31–35 hours of the program over eight weeks. Each lesson consisted of 30 minutes of phonemic activities, 15 minutes of shared reading, and 5–6 minutes of free reading. The phonemic activities included unscripted lessons with sound cards such as using single sounds (shown on two sound cards /sh/ and /oo/) to form the whole syllable (*shoo*). New phonemic and phonetic skills were practiced during shared reading, followed by a free writing time to write about what they read.

Comparison group Students in the comparison group participated in the school’s regular literacy-based reading program. The regular classroom teachers did not have training in phonetics. After the posttest assessment, the comparison group was given the *SpellRead*TM program, while the intervention group was given no further *SpellRead*TM instruction.

Outcomes and measurement

The primary outcomes in the alphabets domain were the Word Identification and Word Attack subtests of the WRMT-R; the Phonemic Decoding Efficiency and Sight Word Efficiency subtests of the Test of Word Reading Efficiency (TOWRE); the Elision, Blending Words, and Segmenting Words subtests of the Comprehensive Test of Phonological Processing (CTOPP); and the Schonell Spelling test. The primary outcomes in the reading fluency domain were the Word Accuracy and Rate subtests of the Gray Oral Reading Test, Third Edition (GORT-3). The primary outcomes in the comprehension domain were the Passage Comprehension subtest of the Woodcock Diagnostic Reading Battery (WDRB) and the Comprehension subtest of the GORT-3. The study reported student outcomes after two months (eight weeks) of program implementation. For a more detailed description of these outcome measures, see Appendix B. The study also used the Spelling test from the *SpellRead*TM test battery (pseudo-spelling), but this measure overlapped with the intervention and did not meet inclusion criteria as an outcome for the Adolescent Literacy review.

Support for implementation

Three teachers and one teacher supervisor implemented the *SpellRead*TM program. The supervisor had previously taught the program for two years, and one of the three teachers had a teaching certificate. All instructors were screened to ensure that they had strong phonological skills. The four instructors participated in an intensive six-day training program provided by experienced *SpellRead*TM staff.

Appendix A.2: Research details for Torgesen et al. (2006)

Torgesen, J., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., Stancavage, F., Durno, D., Javorsky, R., and Haan, C. (2006). *National assessment of Title I. Interim report. Volume II: Closing the reading gap: First year findings from a randomized trial of four reading interventions for striving readers*. Washington, DC: National Center for Education Evaluation and Regional Assistance.

Table A2. Summary of findings

Meets WWC evidence standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Alphabetics	104 students	+8	No
Reading fluency	104 students	+3	No
Comprehension	104 students	0	No

Setting The study took place in 32 school units in the Allegheny Intermediate Unit (AIU), outside Pittsburgh, Pennsylvania. Each school unit consisted of several schools and included two third-grade and two fifth-grade instructional groups. Torgesen et al. (2006) does not report an exact number of participating schools.

Study sample The study design is the random assignment of 32 school units to one of four interventions (*SpellRead™*, *Corrective Reading*, *Failure Free Reading™*, and *Wilson Reading System®*). Within each school, students were randomly assigned either to the treatment group that would receive the intervention assigned to its school or to the comparison group that would receive the standard reading curriculum. This report focuses on schools assigned to *SpellRead™* and on findings for fifth graders (as specified by the Adolescent Literacy review protocol). At the time of the analysis, the sample relevant to this review included 104 fifth-grade students (59 in *SpellRead™* and 45 in the comparison group) in eight school units. Students were eligible for participation if their teacher identified them as a struggling reader and if they scored at or below the 30th percentile on a word-level reading test and at or above the 5th percentile on a vocabulary test. Students scored about one-half to one standard deviation below national norms on baseline measures used to assess their ability to decode words.

Among participating intervention group students, 26% had a learning or other disability, 46% were females, and 52% were eligible for free or reduced price lunches. For the comparison group, these proportions were 35%, 42%, and 43%, respectively.

Intervention group The intervention was implemented from the first week of November 2003 through the first weeks in May 2004. During this time students received an average of 90 hours of *SpellRead™*, which was delivered in 50-minute sessions five days a week to groups of three students. The three-student groups were heterogeneous with regard to students’ basic reading skills. The average skills of each group determined the pace of learning. Many of the sessions took place during the students’ regular classroom reading instruction, but outside their regular classrooms. Implementation fidelity was examined by trainers who observed the teachers and coached them over a period of months and by project coordinators who observed a sample of instructional sessions. In addition, ratings of a sample of videotaped sessions were used. Trainers and project coordinators rated implementation as acceptable.

Comparison group

The comparison group students received their regular reading instruction, which included typical classroom instruction and, in many cases, other services (such as another pull-out program).

Outcomes and measurements

The study reported student outcomes after six months of program implementation. The primary outcomes in the alphabetic domain were the Word Identification and Word Attack subtests of the WRMT-R, and the Phonemic Decoding Efficiency and the Sight Word Efficiency subtests of the TOWRE. The primary outcome in the reading fluency domain was the Oral Reading Fluency test. The primary outcomes in the comprehension domain were the WRMT-R Passage Comprehension subtest and the GRADE Passage Comprehension subtest. For a more detailed description of these outcome measures, see Appendix B. Additional findings reflecting students' outcomes one year after the end of the implementation of the intervention can be found in Appendices D1–D3.

Support for implementation

Professional development on how to use *SpellRead*[™] included training and coaching by *SpellRead*[™] program staff, teachers' independent study of program materials, and telephone conferences between teachers and *SpellRead*[™] staff. On average, the *SpellRead*[™] group teachers participated in 78.1 professional development hours (30.1 hours for initial training, 24.9 hours for a practice phase, and 23.1 hours for training during the six-month *SpellRead*[™] intervention period).

Appendix B: Outcome measures for each domain

Alphabetics	
Phonological awareness construct	
<i>Comprehensive Test of Phonological Processing (CTOPP): Blending Words subtest</i>	This norm-referenced assessment provides an overall measure of the student's phonological awareness skills. The Blending Words subtest includes 20 items that measure the extent to which the student can combine sounds to form words (as cited in Rashotte, MacPhee, & Torgesen, 2001).
<i>CTOPP: Elision subtest</i>	This norm-referenced assessment provides an overall measure of the student's phonological awareness skills. The Elision subtest includes 20 items that measure the extent to which the student can say a word and then say what is left after dropping out designated sounds (as cited in Rashotte, MacPhee, & Torgesen, 2001).
<i>CTOPP: Segmenting Words subtest</i>	This norm-referenced assessment provides an overall measure of the student's phonological awareness skills. The 20-item Segmenting Words subtest has the student repeat words and then say them one sound at a time (as cited in Rashotte, MacPhee, & Torgesen, 2001).
Phonics construct	
<i>Test of Word Reading Efficiency (TOWRE): Phonetic Decoding Efficiency subtest</i>	The TOWRE is a standardized, nationally normed measure. The Phonetic Decoding Efficiency subtest measures the number of nonwords of increasing difficulty that students can pronounce within 45 seconds (as cited in Rashotte, MacPhee, & Torgesen, 2001, and Torgesen et al., 2006).
<i>TOWRE: Sight Word Efficiency subtest</i>	The TOWRE is a standardized, nationally normed measure. The Sight Word Efficiency subtest measures the number of real words of increasing difficulty that students can pronounce within 45 seconds (as cited in Rashotte, MacPhee, & Torgesen, 2001, and Torgesen et al., 2006).
<i>Woodcock Reading Mastery Test—Revised (WRMT-R): Word Identification subtest</i>	The Word Identification subtest is a test of decoding skills. The standardized test requires students to pronounce real words from a list of increasing difficulty (as cited in Rashotte, MacPhee, & Torgesen, 2001, and Torgesen et al., 2006).
<i>WRMT-R: Word Attack subtest</i>	This standardized test measures phonemic decoding skills by asking students to pronounce printed pseudo-words. Students are aware that the words are not real (as cited in Rashotte, MacPhee, & Torgesen, 2001, and Torgesen et al., 2006).
<i>Schonell Spelling test</i>	This 100-item test requires students to correctly spell each word. Answers are scored either right or wrong (as cited in Rashotte, MacPhee, & Torgesen, 2001).
Reading fluency	
<i>Oral Reading Fluency test</i>	This test (also referred to as AIMSweb in the study) measures the number of words correct per minute that students read using three brief grade-level passages. These passages include both fiction and nonfiction text. The norms for this test are updated by Edformation each school year (as cited in Torgesen et al., 2006).
<i>Gray Oral Reading Test, Third Edition (GORT-3): Word Accuracy subtest</i>	The Word Accuracy subtest of the GORT-3 is a standardized reading test that measures the number of word reading errors that occur while reading a series of short paragraphs that increase in difficulty (as cited in Rashotte, MacPhee, & Torgesen, 2001).
<i>GORT-3: Text Reading Rate subtest</i>	The Text Reading Rate subtest of the GORT-3 is a standardized reading test that measures the amount of time taken to read short paragraphs that increase in difficulty (as cited in Rashotte, MacPhee, & Torgesen, 2001).
Comprehension	
Reading comprehension construct	
<i>Group Reading Assessment and Diagnostic Evaluation (GRADE): Passage Comprehension subtest</i>	The GRADE is a norm-referenced reading assessment that can be used with students at any level. The GRADE has four subtests: (1) Vocabulary, (2) Sentence Comprehension, (3) Passage Comprehension, and (4) Listening Comprehension. The Passage Comprehension subtest includes a passage of text and corresponding multiple-choice comprehension questions (as cited in Torgesen et al., 2006).
<i>GORT-3: Comprehension subtest</i>	In this standardized test, students read paragraphs and answer five comprehension questions for each paragraph. The questions are read to students by the tester (as cited in Rashotte, MacPhee, & Torgesen, 2001).
<i>WRMT-R: Passage Comprehension subtest</i>	In this standardized test, comprehension is measured by having students read silently and fill in missing words in a short paragraph (as cited in Rashotte, MacPhee, & Torgesen, 2001, and Torgesen et al., 2006).
<i>Woodcock Diagnostic Reading Battery (WDRB): Passage Comprehension subtest</i>	The Passage Comprehension subtest of the WDRB asks students to read a series of paragraphs silently and complete the missing words in each paragraph (as cited in Rashotte, MacPhee, & Torgesen, 2001).

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

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Rashotte, MacPhee, & Torgesen, 2001^a								
<i>CTOPP: Elision subtest</i>	Grades 5–6	33 students	84.70 (13.60)	81.60 (12.20)	3.10	0.51	+19	< 0.05
<i>CTOPP: Blending Words subtest</i>	Grades 5–6	33 students	104.60 (10.60)	90.30 (11.20)	14.30	1.80	+46	< 0.05
<i>CTOPP: Segmenting Words subtest</i>	Grades 5–6	33 students	99.70 (8.50)	84.40 (7.00)	15.30	2.38	+49	< 0.05
<i>TOWRE: Phonetic Decoding Efficiency subtest</i>	Grades 5–6	33 students	86.80 (11.10)	80.80 (8.10)	6.00	0.88	+31	> 0.05
<i>TOWRE: Sight Word Efficiency subtest</i>	Grades 5–6	33 students	91.60 (11.80)	92.70 (9.20)	-1.10	-0.22	-9	< 0.05
<i>WRMT-R: Word Identification subtest</i>	Grades 5–6	33 students	93.90 (11.70)	90.90 (6.70)	3.00	0.64	+24	< 0.05
<i>WRMT-R: Word Attack subtest</i>	Grades 5–6	33 students	102.30 (8.90)	84.40 (6.90)	17.90	2.20	+49	< 0.05
<i>Schonell Spelling test</i>	Grades 5–6	33 students	50.30 (11.90)	47.70 (8.00)	2.60	0.06	+2	< 0.05
Domain average for alphabetics (Rashotte, MacPhee, & Torgesen, 2001)						1.03	+35	Statistically significant
Torgesen et al., 2006^b								
<i>TOWRE: Phonetic Decoding Efficiency subtest</i>	Grade 5	8 school units/ 104 students	92.50 (15.00)	88.40 (15.00)	4.10	0.27	+11	< 0.05
<i>TOWRE: Sight Word Efficiency subtest</i>	Grade 5	8 school units/ 104 students	92.50 (15.00)	91.40 (15.00)	2.10	0.14	+6	> 0.05
<i>WRMT-R: Word Identification subtest</i>	Grade 5	8 school units/ 104 students	90.90 (15.00)	90.80 (15.00)	0.10	0.01	0	> 0.05
<i>WRMT-R: Word Attack subtest</i>	Grade 5	8 school units/ 104 students	102.00 (15.00)	96.70 (15.00)	5.30	0.35	+14	< 0.05
Domain average for alphabetics (Torgesen et al., 2006)						0.19	+8	Not statistically significant
Domain average for alphabetics across all studies						0.61	+21	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 student outcomes, representing the average change expected for all students 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 student's percentile rank that can be expected if the student 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. na = not applicable. CTOPP = Comprehensive Test of Phonological Processing. TOWRE = Test of Word Reading Efficiency. WRMT-R = Woodcock Reading Mastery Test-Revised.

^a For Rashotte, MacPhee, & Torgesen (2001), a correction for multiple comparisons was needed and resulted in significance levels that differ from those in the original study. The *CTOPP Elision*, *WRMT-R Word Identification*, and *Schonell Spelling* contrasts were not found to be statistically significant, after adjusting for multiple comparisons. The p -values and effect sizes presented here were reported in the original study. The WWC calculated the program group mean using a difference-in-differences approach (see the WWC Procedures and Standards Handbook, Appendix B) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons.

^b For Torgesen et al. (2006), a correction for multiple comparisons was needed, but the WWC could not apply this correction because exact p -values were not reported by the authors. The p -value ranges presented here were reported in the original study. For Torgesen et al. (2006), the mean outcomes were computed using information reported in the paper. For the comparison group, the mean outcome is the comparison group baseline mean standard score (Table II.3, p. 11) plus the comparison group gain. For the intervention group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain plus the impact of the intervention. The standard deviations in the Torgesen et al. (2006) study were the population standard deviations for these standardized outcomes. This study is characterized as having an indeterminate effect because no effects are statistically significant within the domain, accounting for multiple comparisons, and the mean effect is neither statistically significant nor substantively important.

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

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Rashotte, MacPhee, & Torgesen, 2001^a								
<i>GORT-3: Accuracy subtest</i>	Grades 5–6	33 students	98.80 (16.40)	94.70 (14.20)	4.10	0.38	+15	< 0.05
<i>GORT-3: Rate subtest</i>	Grades 5–6	33 students	89.80 (13.70)	81.60 (14.50)	8.20	0.92	+32	< 0.05
Domain average for reading fluency (Rashotte, MacPhee, & Torgesen, 2001)						0.65	+24	Statistically significant
Torgesen et al., 2006^b								
<i>Oral Reading Fluency test</i>	Grade 5	8 school units/ 104 students	103.50 (47.00)	99.90 (47.00)	3.60	0.08	+3	> 0.05
Domain average for reading fluency (Torgesen et al., 2006)						0.08	+3	Not statistically significant
Domain average for reading fluency across all studies						0.37	+14	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 student outcomes, representing the average change expected for all students 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 student’s percentile rank that can be expected if the student 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. na = not applicable. GORT-3 = Gray Oral Reading Test, Third Edition.

^a For Rashotte, MacPhee, & Torgesen (2001), a correction for multiple comparisons was needed and resulted in significance levels that differ from those in the original study. The *GORT-3 Accuracy* contrast was not found to be statistically significant, after adjusting for multiple comparisons. The p-values and effect sizes presented here were reported in the original study. The WWC calculated the program group mean using a difference-in-differences approach (see the WWC Procedures and Standards Handbook, Appendix B) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons.

^b For Torgesen et al. (2006), no corrections for clustering or multiple comparisons were needed. The p-value range presented here was reported in the original study. For the comparison group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain. For the intervention group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain plus the impact of the intervention. The standard deviations in the study were the population standard deviations for these standardized outcomes. This study is characterized as having an indeterminate effect because the effect is neither statistically significant nor substantively important.

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

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Rashotte, MacPhee, & Torgesen, 2001^a								
<i>GORT-3: Comprehension subtest</i>	Grades 5–6	33 students	100.70 (14.60)	91.60 (12.60)	9.10	0.64	+24	< 0.05
<i>WDRB: Comprehension subtest</i>	Grades 5–6	33 students	100.50 (12.20)	97.80 (10.20)	2.70	0.43	+17	< 0.05
Domain average for comprehension (Rashotte, MacPhee, & Torgesen, 2001)						0.54	+20	Not statistically significant
Torgesen et al., 2006^b								
<i>WRMT-R: Passage Comprehension subtest</i>	Grade 5	8 school units/ 104 students	92.60 (15.00)	92.00 (15.00)	0.60	0.04	+2	> 0.05
<i>GRADE: Passage Comprehension subtest</i>	Grade 5	8 school units/ 104 students	89.20 (15.00)	89.90 (15.00)	–0.70	–0.05	–2	> 0.05
Domain average for comprehension (Torgesen et al., 2006)						0.00	0	Not statistically significant
Domain average for comprehension across all studies						0.27	+11	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 student outcomes, representing the average change expected for all students 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 student’s percentile rank that can be expected if the student 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. na = not applicable. GORT-3 = Gray Oral Reading Test, Third Edition. WDRB = Woodcock Diagnostic Reading Battery. WRMT-R = Woodcock Reading Mastery Test–Revised. GRADE = Group Reading Assessment and Diagnostic Evaluation.

^a For Rashotte, MacPhee, & Torgesen (2001), no corrections for clustering or multiple comparisons were needed. The p-values computed by the WWC were larger than 0.05 and did not require the correction for multiple comparisons. The p-values and effect sizes presented here were reported in the original study. The WWC calculated the program group mean using a difference-in-differences approach (see the WWC Procedures and Standards Handbook, Appendix B) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. This study is characterized as having a substantively important positive effect, because no effects are statistically significant within the domain and the positive mean effect is at least 0.25.

^b For Torgesen et al. (2006), no corrections for clustering or multiple comparisons were needed. The p-values presented here were reported in the original study. For the comparison group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain. For the intervention group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain plus the impact of the intervention. The standard deviations in the study were the population standard deviations for these standardized outcomes. This study is characterized as having an indeterminate effect because no effects are statistically significant within the domain, and the mean effect is neither statistically significant nor substantively important.

Appendix D.1: Supplemental findings for the alphabetic domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Torgesen et al., 2007^a								
<i>TOWRE: Phonetic Decoding Efficiency subtest</i>	Grade 5	8 school units/ 100 students	88.10 (15.00)	84.90 (15.00)	3.20	0.21	+8	> 0.05
<i>TOWRE: Sight Word Efficiency subtest</i>	Grade 5	8 school units/ 100 students	99.60 (15.00)	87.20 (15.00)	3.40	0.22	+9	< 0.05
<i>WRMT-R: Word Identification subtest</i>	Grade 5	8 school units/ 100 students	89.30 (15.00)	89.20 (15.00)	0.10	0.01	0	> 0.05
<i>WRMT-R: Word Attack subtest</i>	Grade 5	8 school units/ 100 students	95.80 (15.00)	92.30 (15.00)	3.50	0.23	+9	> 0.05

Table Notes: The supplemental findings presented in this table are additional findings reflecting students' outcomes one year after the end of the implementation of the intervention from Torgesen et al. (2007) that do not factor in 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 student outcomes, representing the average change expected for all students 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 student's percentile rank that can be expected if the student is given the intervention. TOWRE = Test of Word Reading Efficiency. WRMT-R = Woodcock Reading Mastery Test–Revised.

^a For Torgesen et al. (2007), a correction for multiple comparisons was needed, but the WWC could not apply this correction because exact p-values were not reported by the authors. The p-value ranges presented here were reported in the original study. For the comparison group, the mean outcome is the comparison group baseline mean standard score (p. 11) plus the comparison group gain (p. xvii). For the intervention group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain plus the impact of the intervention (p. xvii).

Appendix D.2: Supplemental findings for the reading fluency domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Torgesen et al., 2007^a								
<i>Oral Reading Fluency test</i>	Grade 5	8 school units/ 100 students	102.50 (47.00)	105.80 (47.00)	-3.30	-0.07	-3	> 0.05

Table Notes: The supplemental findings presented in this table are additional findings reflecting students' outcomes one year after the end of the implementation of the intervention from Torgesen et al. (2007) that do not factor in 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 student outcomes, representing the average change expected for all students 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 student's percentile rank that can be expected if the student is given the intervention.

^a For Torgesen et al. (2007), no corrections for clustering or multiple comparisons were needed. The p-value presented here was reported in the original study. For the comparison group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain. For the intervention group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain plus the impact of the intervention. The standard deviations in the study were the population standard deviations for these standardized outcomes.

Appendix D.3: Supplemental findings for the comprehension domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Torgesen et al., 2007^a								
<i>WRMT-R: Passage Comprehension subtest</i>	Grade 5	8 school units/ 100 students	89.10 (15.00)	90.00 (15.00)	-0.90	-0.06	-3	> 0.05
<i>GRADE: Passage Comprehension subtest</i>	Grade 5	8 school units/ 100 students	83.30 (15.00)	84.40 (15.00)	-1.10	-0.07	-3	> 0.05

Table Notes: The supplemental findings presented in this table are additional findings reflecting students' outcomes one year after the end of the implementation of the intervention from Torgesen et al. (2007) that do not factor in 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 student outcomes, representing the average change expected for all students 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 student's percentile rank that can be expected if the student is given the intervention. WRMT-R = Woodcock Reading Mastery Test–Revised. GRADE = Group Reading Assessment and Diagnostic Evaluation.

^a For Torgesen et al. (2007), no correction for clustering was needed in the comprehension domain. No correction for multiple comparisons was needed because the study's reported corrections for multiple comparisons were based on the same grouping of outcomes as the domain for this review. The p-values presented here were reported in the original study. For the comparison group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain. For the intervention group, the mean outcome is the comparison group baseline mean standard score plus the comparison group gain plus the impact of the intervention. The standard deviations in the study were the population standard deviations for these standardized outcomes.

Endnotes

¹ The descriptive information for this program was obtained from publicly available sources: the WWC Beginning Reading *SpellRead*[™] intervention report and the distributor's website (<http://www.pcieducation.com/spellread/default.aspx>, downloaded January 2012). The WWC requests distributors review the program description sections for accuracy from their perspective. The program description was provided to the distributor in January 2012; however, the WWC received no response. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by December 2011.

² The studies in this report were reviewed using WWC Evidence Standards, version 2.1, as described in the Adolescent Literacy review protocol, version 2.0. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

³ One study in this intervention report, Torgesen et al. (2006), was prepared in part by staff of Mathematica Policy Research. For this reason, the study was rated by researchers unaffiliated with Mathematica. The report was reviewed by the principal investigator, a WWC quality assurance reviewer, and an external peer reviewer.

⁴ For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 22 of this report. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

⁵ The study authors conducted statistical analyses of three groups of students: grades 1 and 2, grades 3 and 4, and grades 5 and 6. This report focuses only on the impact of *SpellRead*[™] on students in grades 5 and 6, as defined in the Adolescent Literacy review protocol, version 2.0.

⁶ A school unit consists of several schools partnering so that each cluster included two third-grade and two fifth-grade instructional groups. Torgesen et al. (2006) does not report an exact number of participating schools. Only the findings on fifth graders are included in this review as specified by the Adolescent Literacy review protocol, version 2.0.

⁷ The study's authors refer to the intervention as *SpellRead P.A.T. (Phonological Auditory Training)*.

⁸ Additional findings reflecting students' outcomes one year after the intervention year can be found in Appendices D.1–D.3. Torgesen et al. (2006, 2007) also reported subgroup analyses by initial skill level (Woodcock Reading Mastery Test–Revised Word Attack subtest and Peabody Picture Vocabulary Test) and socioeconomic status. The study did not establish baseline equivalence of the intervention and comparison students in these subgroups. Therefore, these analyses are not included in this report.

⁹ The WWC computes an average effect size as a simple average of the effect sizes across all individual findings within the study domain.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, January). *Adolescent Literacy intervention report: SpellRead*[™]. Retrieved from <http://whatworks.ed.gov>.

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC evidence standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC evidence 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 evidence 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 evidence 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 participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.
Clustering adjustment	If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.
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 design of a study is the method by which intervention and comparison groups were assigned.
Domain	A domain is a group of closely related outcomes.
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 analysis sample groups are similar on observed characteristics defined in the review area protocol.
Extent of evidence	An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 22.
Improvement index	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from -50 to +50.
Multiple comparison adjustment	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
Quasi-experimental design (QED)	A quasi-experimental design (QED) is a research design in which subjects 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 investigators randomly assign eligible participants into intervention and comparison groups.
Rating of effectiveness	The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 22.
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% ($p < 0.05$).
Substantively important	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the [WWC Procedures and Standards Handbook \(version 2.1\)](#) for additional details.