

What Works Clearinghouse



Wilson Reading System®

Program description¹

Wilson Reading System® is a supplemental reading and writing curriculum designed to promote reading accuracy (decoding) and spelling (encoding) skills for students with word-level deficits. The program is designed to teach phonemic awareness, alphabetic principles (sound-symbol relationship), word study, spelling, sight word instruction, fluency, vocabulary, oral expressive language development, and comprehension. Students engage in a variety of activities in the classroom, including hearing sounds, practicing with syllable and word cards, listening

to others read, and reading aloud and repeating what they have read in their own words. The program is designed to help children master new skills, with reviews reinforcing previous lessons. This program was designed for students in grade 2 and above. *Foundations*®, a related program not reviewed in this report, was recently developed with the same principle for students in Kindergarten through third grade. In the single study reviewed by the WWC for this report, only the word-level components of *Wilson Reading System*® were implemented.

Research

One study of a modified version of *Wilson Reading System*® met the What Works Clearinghouse (WWC) evidence standards.² This one study included more than 70 third grade students in Pennsylvania.³ The WWC considers the extent of

evidence for *Wilson Reading System*® to be small for alphabets, fluency, and comprehension. No studies that met WWC evidence standards with or without reservations addressed general reading achievement.

1. The descriptive information for this program was obtained from a publicly available sources: the program's web site (www.wilsonlanguage.com, downloaded April, 2007) and the research literature (Torgesen et al., 2006). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
2. The fluency, comprehension, and vocabulary components of the *Wilson Reading System*® were eliminated from instruction at the request of Torgesen et al. for the purposes of the study. For further information about the program implemented, please see the research and findings sections in this report.
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

Effectiveness *Wilson Reading System*[®] was found to have potentially positive effects on alphabets and no discernible effects on fluency and comprehension.

	Alphabets	Fluency	Comprehension	General reading achievement
Rating of effectiveness	Potentially positive	No discernible effects	No discernible effects	na
Improvement index⁴	Average: +13 percentile points Range: +6 to +22 percentile points	Average: +6 percentile points	Average: +7 percentile points Range: +3 to +11 percentile points	na

na = not applicable

Additional program information¹

Developer and contact

Developed by Barbara Wilson, *Wilson Reading System*[®] is distributed by Wilson Language Training. Address: 47 Old Webster Road, Oxford, MA 01540. Email: info@WilsonLanguage.com. Web: www.wilsonlanguage.com/w_wrs.htm. Telephone: (508) 368-2399.

Scope of use

The *Wilson Reading System*[®] was originally designed in 1988 to teach reading and writing to students who experienced difficulties with written language from upper elementary school through adulthood. *Wilson Reading System*[®] has been implemented in public and private schools, clinics, adult education classes, family literacy programs, and home school settings across the United States. In 2002 *Foundations*[®] was designed for students in Kindergarten through third grade, building on the principles of the *Wilson Reading System*[®]. The number of students and schools using the *Wilson Reading System*[®] is not available.

Teaching

The *Wilson Reading System*[®] has a daily 10-part lesson plan that builds on interaction between the teacher and student. It is divided into three blocks: parts one through five emphasize word study, parts six through eight emphasize spelling, and parts nine and 10 emphasize fluency and comprehension. The *Wilson*

Reading System[®] teaches the structure of words in the English language focusing first on basic word skills and then on more complex language structure, including morphological principles. The program provides two levels of vocabulary. Level A uses reading material appropriate for younger or ESL students, while Level B is for older students. The intervention model can be used in reading classes, small groups, or tutorials, for 45–90 minute daily lessons in general or special education classrooms. The intensive model of 60–90 minute instruction in small groups or individually is recommended for the most challenged readers. The *Wilson Reading System*[®] provides teachers and students with materials necessary to implement the program, including a *Wilson* instructors’ manual that helps the teacher prepare the daily lesson.

Training for *Wilson Reading System*[®] includes a formal professional development process. Offerings include an intensive program certification, workshops, videos, online education, and onsite visits for feedback. In addition, ongoing support, during and after training, is provided by phone, email, an annual national conference, and resources posted on the online *Wilson Academy*.

Cost

The *Wilson Reading System*[®] instructional sets range from \$149 to \$500. Materials for the teacher include an instructor

4. These numbers show the average and range of student-level improvement indices for all findings across the study.

Additional program information *(continued)*

manual, rules notebook, dictation books, assessment materials, instructional videos, and manipulatives (including sound and word cards). For \$59, teachers can access additional lesson plans, demonstrations, and weekly current event stories with an annual subscription to *Wilson Academy*. Student materials are purchased separately and include text readers (Steps 1–12),

workbooks, and a magnetic journal with letter tiles. *Wilson* provides different levels of professional development and support for teachers, offering in-service professional development to school districts as well as public workshops. A two-day public professional development course costs \$325.

Research

Nine studies reviewed by the WWC investigated the effects of *Wilson Reading System*[®]. One study (Torgesen et al., 2006) was a randomized controlled trial that met WWC evidence standards. The remaining studies did not meet evidence screens.

Torgesen et al. (2006) examined the effects of *Wilson Reading System*[®] on 71 third-grade students in eight school units⁵ in Pennsylvania. Students in the comparison group participated in the regular reading program at their schools.⁶

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or moderate to large (see the [What Works Clearinghouse Extent of Evidence Categorization Scheme](#)). The extent of evidence takes into account the number of studies and the total sample size across the studies that met WWC evidence standards with or without reservations.⁷

The WWC considers the extent of evidence for *Wilson Reading System*[®] to be small for alphabets, fluency, and comprehension. No studies that met WWC evidence standards with or without reservations addressed general reading achievement.

Effectiveness

Findings

The WWC review of interventions for beginning reading addresses student outcomes in four domains: alphabets, fluency, comprehension, and general reading achievement.⁸ The study included in this *Wilson Reading System*[®] report covers three domains: alphabets, fluency, and comprehension. Within the alphabets domain, the study reported on one construct: phonics.

Alphabets. Torgesen et al. (2006) analyzed the group differences on four phonics outcomes in the alphabets domain (Woodcock Reading Mastery Test–Revised (WRMT–R) word

identification and word attack subtests and the Test of Word Reading Efficiency (TOWRE) phonetic decoding efficiency and sight word efficiency subtests). The authors reported statistically significant effects of the *Wilson Reading System*[®] on two of these outcomes (WRMT–R word identification and word attack subtests). The statistical significance of these findings was confirmed by the WWC. The average effect size across the three outcomes was large enough to be considered substantively important according to WWC criteria (that is, an effect size at least 0.25).

Fluency. Torgesen et al. (2006) examined the effect of the intervention on one outcome in this domain (the Oral Reading

5. A school unit consists of several partnered schools so that the cluster included two third-grade and two fifth-grade instructional groups. Because of the age range defined by the Beginning Reading review, only data of the third graders were included in this review.
6. For the purposes of this study, only word-level skill components of *Wilson Reading System*[®] were implemented, but the study noted that the complete version contains instructional routines and materials that also focus on comprehension and vocabulary.
7. The Extent of Evidence Categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept, external validity, such as the students' demographics and the types of settings in which studies took place, are not taken into account for the categorization.
8. For definitions of the domains, see the [Beginning Reading Protocol](#).

Effectiveness *(continued)*

Fluency test). They reported no statistically significant differences between groups for the outcome.

Comprehension. Torgesen et al. (2006) examined two outcomes in this domain (the WRMT–R passage comprehension subtest and the 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.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings,⁹ the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the [WWC Intervention Rating Scheme](#)).

The WWC found the modified version of the Wilson Reading System® used in this study to have potentially positive effects on alphabets and no discernible effects on fluency and comprehension

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see [Technical Details of WWC-Conducted Computations](#)). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is based entirely on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analyses. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.

The average improvement index for alphabets is +13 percentile points across all findings in the single study, with a range

of +6 to +22 percentile points. The improvement index for the fluency outcome is +6 percentile points in the single study. The average improvement index for comprehension is +7 percentile points across all findings in the study, with a range of +3 to +11 percentile points.

Summary

The WWC reviewed nine studies on *Wilson Reading System*®. One study met WWC evidence standards, and the remaining studies did not meet WWC evidence screens. Based on this one study, the WWC found potentially positive effects in alphabets, and no discernible effects in fluency and comprehension. It should be noted, however, that the fluency, comprehension, and vocabulary components of the *Wilson Reading System*® were not used at the request of the researchers conducting the study. The evidence presented in this report is limited and may change as new research emerges.

References

Met WWC evidence standards

Torgesen, J., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., et al. (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. Retrieved from Institute of Education Sciences, U.S. Department of Education Web site: <http://www.ed.gov/rschstat/eval/disadv/title1interimreport/index.html>

9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate the statistical significance. In the case of *Wilson Reading System*®, corrections for multiple comparisons were needed.

References *(continued)*

Did not meet WWC evidence screens

- Banks, S. R., Guyer, B. P., & Guyer, K. E. (1993). Spelling improvement by college students who are dyslexic. *Annals of Dyslexia*, 43, 186–193.¹⁰
- Gustavson, K., & Watson, N. (1995). *Wilson Reading and Reading to Read*. Augusta, ME: Division of Adult & Community Education.¹⁰
- Guyer, B. P., Banks, S. R., & Guyer, K. E. (1993). Spelling improvement for college students who are dyslexic. *Annals of Dyslexia*, 43, 254–259.¹⁰
- Moats, L. C. (1998). Reading, spelling, and writing disabilities in the middle grades. In B. Wong (Ed.), *Learning about learning disabilities* (2nd edition) (pp. 1–19). Orlando, FL: Academic Press.¹⁰
- Wilson, B. A. (1998). Matching student needs to instruction: Teaching reading and spelling using the Wilson Reading System. In S. A. Vogel & S. Reder (Eds.), *Learning disabilities, literacy, and adult education* (pp. 213–234). Baltimore, MD: Brookes Publishing.¹¹
- Wilson, B. A., & O'Connor, J. R. (1995). Effectiveness of the Wilson Reading System used in public school training. In C. W. McIntyre & J. S. Pickering (Eds.), *Clinical studies of multisensory structured language education* (pp. 247–254). Salem, OR: International Multisensory Structured Language Education Council.¹¹
- Wilson Language Training. (2002). [Evidence of Effectiveness: Wake Forest University, 2002: Data analysis]. Unpublished raw data. Retrieved from http://www.wilsonlanguage.com/PDF/Evidence_Data_Analysis.pdf¹²
- Wilson Language Training Corporation. (2002). *Wilson Literacy Solutions: Evidence of effectiveness Wilson Spelling results 2000*. Retrieved from http://www.wilsonlanguage.com/PDF/Lynn_Results.pdf¹²

For more information about specific studies and WWC calculations, please see the [WWC Wilson Reading System® Technical Appendices](#).

10. The sample is not appropriate to this review: the parameters for this WWC review specified that students should be in grades kindergarten through 3 during the time of the intervention; this study did not focus on the targeted grades.
11. The sample is not appropriate to this review: the parameters for this WWC review specified that students should be in grades kindergarten through 3; this study did not disaggregate students in the eligible range from those outside the range.
12. Does not use a strong causal design: this study did not use a comparison group.

Appendix

Appendix A1.1 Study characteristics: Torgesen et al., 2006 (randomized controlled trial)

Characteristic	Description
Study citation	Torgesen, J., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., et al. (2006). <i>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</i> . Retrieved from Institute of Education Sciences, U.S. Department of Education Web site: http://www.ed.gov/rschstat/eval/disadv/title1interimreport/index.html
Participants	The study design was based on random assignment of 37 school units ¹ to one of the four interventions, <i>Corrective Reading</i> , <i>Kaplan SpellRead</i> , <i>Failure Free Reading</i> , and <i>Wilson Reading System</i> [®] . Within each school, students were randomly assigned to the intervention condition or to the comparison condition. ² This report focuses on eight school units assigned to <i>Wilson Reading System</i> [®] . ³ At the time of analysis, the study included a total of 71 third-grade students (53 in the intervention and 18 in the comparison groups). Sample size at posttest by outcome measure was not reported. ⁴ In the intervention group, 61% of the students were female, 45% were African-American, and 36% were eligible for the free/reduced lunch program. In the comparison group, 79% of the students were female, 32% were African-American, and 64% were eligible for the free/reduced lunch program.
Setting	Eight school units in Pennsylvania.
Intervention	<i>Wilson Reading System</i> [®] was implemented by nine teachers from November 2003 to May 2004. For purposes of this study only word-level skills were developed, although the complete version of <i>Wilson</i> contains instructional routines and materials that also focus on comprehension and vocabulary. A 50-minute lesson was delivered five days a week to groups of three students with various basic reading levels. The average capabilities of each three-student group determined the pace of learning. Many of the sessions took place during the students' regular classroom reading instruction but were held outside their regular classrooms. Thus intervention group students received less reading instruction in the classroom than did students in the comparison group. Implementation fidelity was examined by reading program trainers who observed the teachers and coached them over a period of months, project coordinators who observed a sample of instructional sessions, and ratings based on a sample of videotaped sessions. Implementation was rated as acceptable.
Comparison	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). The comparison group students had fewer small group instructional hours than the intervention group students, but more one-on-one instructional hours.
Primary outcomes and measurement	The outcome measures in the alphabetic domain were the phonemic decoding efficiency and sight word efficiency subtests of the Test of Word Reading Efficiency (TOWRE) and the word identification and word attack subtests of the Woodcock Reading Mastery Tests–Revised (WRMT–R). The only measure in the fluency domain was the Oral Reading Fluency test. Measures in the comprehension domain were the passage comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE) and the passage comprehension subtest of WRMT–R. (See Appendix A2.1–2.3 for more detailed descriptions of outcome measures.)
Teacher training	Trainers from <i>Wilson Reading System</i> [®] provided teacher training, which included group instruction, coaching, telephone consultation, and independent study using the <i>Wilson Academy</i> online course. On average, intervention group teachers participated in 62.5 professional development hours across all phases of the study (initial training phase, practice phase, and implementation phase).

1. A school unit consists of several partnered schools so that the cluster included two third-grade and two fifth-grade instructional groups.
2. One of seven indicators of students' reading skills at baseline (TOWRE-SWE) showed statistically significant differences between the intervention and comparison groups. Baseline differences were taken into account in the WWC analysis of the program effects.
3. Findings on *Corrective Reading*, *Kaplan SpellRead*, and *Failure Free Reading* are included in other WWC beginning reading reports.
4. The study reported that four students in the intervention group and three students in the comparison group were lost to analysis. However, it is not clear whether those students were in third grade or were part of an additional sample of fifth-grade students also examined in this study. The fifth-grade sample included in this study is not reviewed in this report because it is outside the scope of the review. For sample relevancy criteria, please see the [Beginning Reading Protocol](#).

Appendix A2.1 Outcome measures in the alphabetic domain

Outcome measure	Description
<i>Phonics</i>	
Test of Word Reading Efficiency (TOWRE): Phonetic Decoding Efficiency subtest	The TOWRE is a standardized, nationally normed measure. The phonetic decoding efficiency subtest measures the number of pronounceable printed nonwords that can be accurately decoded within 45 seconds (as cited in Torgesen et al., 2006).
TOWRE: Sight Word Efficiency subtest	The TOWRE is a standardized, nationally normed measure. The sight word efficiency subtest assesses the number of real printed words that can be accurately identified within 45 seconds (as cited in Torgesen et al., 2006).
Woodcock Reading Mastery Test–Revised (WRMT–R): Word Identification subtest	The word identification subtest is a test of decoding skills. The standardized test requires the child to read aloud isolated real words that range in frequency and difficulty (as cited in Torgesen et al., 2006).
WRMT–R: Word Attack subtest	This standardized test measures phonemic decoding skills by asking students to read pseudowords. Students are aware that the words are not real (as cited in Torgesen et al., 2006).

Appendix A2.2 Outcome measure in the fluency domain

Outcome measure	Description
Edformation Oral Fluency Assessment	This test measures the number of words correct per minute (WCPM) that students read using three brief grade-level passages (AIMSweb, as cited in Torgesen et al., 2006). These passages include both fiction and nonfiction text. The norms for this test are updated by Edformation each school year.

Appendix A2.3 Outcome measures in the comprehension domain

Outcome measure	Description
<i>Reading comprehension</i>	
Group Reading Assessment and Diagnostic Evaluation (GRADE): Passage Comprehension subtest	The GRADE is an untimed, norm-referenced standardized test. The passage comprehension subtest includes a passage of text and corresponding multiple-choice comprehension questions (as cited in Torgesen et al., 2006).
WRMT–R: Passage Comprehension subtest	In this standardized test, comprehension is measured by having students fill in missing words in a short paragraph (as cited in Torgesen et al., 2006).

Appendix A3.1 Summary of study findings included in the rating for the alphabetic domain¹

Outcome measure	Study sample	Sample size (school units/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (Wilson Reading System [®] – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Wilson Reading System [®] group	Comparison group				
Torgesen et al., 2006 (randomized controlled trial)⁷								
TOWRE: Phonetic Decoding Efficiency subtest	Grade 3	8/71	91.97 (15.00)	86.19 (15.00)	5.78	0.38	Statistically significant	+15
TOWRE: Sight Word Efficiency subtest	Grade 3	8/71	87.19 (15.00)	84.14 (15.00)	3.05	0.20	ns	+8
WRMT–R: Word Identification subtest	Grade 3	8/71	92.21 (15.00)	89.75 (15.00)	2.46	0.16	ns	+6
WRMT–R: Word Attack subtest	Grade 3	8/71	103.10 (15.00)	94.30 (15.00)	8.80	0.58	Statistically significant	+22
Domain average⁸ for alphabetics						0.33	na	+13

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. The study also included subgroup analyses by initial skill level (WRMT–R word attack subtest and Peabody Picture Vocabulary Test (PPVT)) and socio-economic status. The study found statistically significant positive effects on WRMT–R word attack scores at posttest only for students with initial high word attack scores and students with initial high PPVT scores. Finally, the study found statistically significant positive effects on WRMT–R word attack and TOWRE-PDE posttest scores only for students who were not eligible for free/reduced lunch program, but not for those students who were eligible for free/reduced lunch.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean is the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Torgesen et al. (2006) and the alphabetic domain, no corrections for clustering were needed because students were assigned to conditions. Corrections for multiple comparisons were needed because the study's reported corrections for multiple comparisons were based on grouping of outcomes that differs from the grouping of domains for this review.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (school units/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (Wilson Reading System [®] – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Wilson Reading System [®] group	Comparison group				
Torgesen et al., 2006 (randomized controlled trial)⁷								
Oral Reading Fluency	Grade 3	8/71	46.95 (39.20)	41.00 (39.20)	5.95	0.15	ns	+6
Domain average⁸ for fluency						0.15	ns	+6

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. The study also included subgroup analyses by initial skill level (WRMT–R word attack subtest and Peabody Picture Vocabulary Test (PPVT)) and socio-economic status. No differences were found between subgroups of students for the outcome in the fluency domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean is the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Torgesen et al. (2006) and fluency, no corrections for clustering were needed because students were assigned to conditions. No corrections for multiple comparisons were needed because there is only one outcome in this domain.
8. This row provides the domain average, which in this instance is also the single outcome finding from the one study.

Appendix A3.3 Summary of study findings included in the rating for the comprehension domain¹

Outcome measure	Study sample	Sample size (school units/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (Wilson Reading System [®] – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Wilson Reading System [®] group	Comparison group				
Torgesen et al., 2006 (randomized controlled trial)⁷								
GRADE: Passage Comprehension subtest	Grade 3	8/71	89.97 (15.00)	85.78 (15.00)	4.19	0.28	ns	+11
WRMT-R: Passage Comprehension subtest	Grade 3	8/71	93.87 (15.00)	92.87 (15.00)	1.00	0.07	ns	+3
Domain average⁸ for comprehension						0.17	ns	+7

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. The study also included subgroup analyses by initial skill level (WRMT–R word attack subtest and Peabody Picture Vocabulary Test (PPVT)) and socioeconomic status. No differences were found between subgroups of students for outcomes in the comprehension domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean is the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Torgesen et al. (2006) and the comprehension domain, no corrections for clustering were needed. No correction for multiple comparisons were needed because the study's reported corrections for multiple comparisons were based on the same grouping of outcomes as the domain for this review.
8. This row provides the domain average, which in this instance is also the study average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Wilson Reading System® rating for the alphabets domain

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

For the outcome domain of alphabets, the WWC rated *Wilson Reading System*® as potentially positive effects. It did not meet the criteria for positive effects because only one study showed a statistically significant positive effect. The remaining ratings (mixed, no discernible effects, potentially negative, or negative) were not considered because *Wilson Reading System*® was assigned the highest applicable rating.

Rating received

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

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

Met. One study showed a statistically significant positive effect.

AND

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

Met. No studies showed a statistically significant or substantively important negative effect. The single study that met the WWC standards showed a statistically significant positive effect.

Other ratings considered

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

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

Not met. Only one study showed a statistically significant positive effect.

AND

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

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

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A4.2 Wilson Reading System® rating for the fluency domain

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

For the outcome domain of fluency, the WWC rated *Wilson Reading System*® as no discernible effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, and negative effects) because the single study that met WWC standards did not show statistically significant or substantively important effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

Met. No studies showed statistically significant or substantively important positive or negative effects.

Other ratings considered

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

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

Not met. No studies showed statistically significant positive effects.

AND

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

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

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

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

Not met. No studies showed statistically significant or substantively important positive effects.

AND

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

Not met. The single study that met WWC standards showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

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

Not met. No studies showed statistically significant or substantively important effects, either positive or negative.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. No studies showed statistically significant or substantively important effects, either positive or negative.

(continued)

Appendix A4.2 Wilson Reading System® rating for the fluency domain (continued)

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

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

Not met. No studies showed statistically significant or substantively important negative effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies showed statistically significant or substantively important positive effects. In addition, no studies showed a statistically significant or substantively important negative effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

Not met. No studies showed statistically significant negative effects.

AND

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

Met. No studies showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A4.3 Wilson Reading System® rating for the comprehension domain

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

For the outcome domain of comprehension, the WWC rated *Wilson Reading System*® as no discernible effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, and negative effects) because the single study that met WWC standards did not show statistically significant or substantively important effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

Met. No studies showed statistically significant or substantively important positive or negative effects.

Other ratings considered

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

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

Not met. No studies showed statistically significant positive effects.

AND

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

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

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

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

Not met. No studies showed statistically significant or substantively important positive effects.

AND

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

Not met. The single study that met WWC standards showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

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

Not met. No studies showed statistically significant or substantively important effects, either positive or negative.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. No studies showed statistically significant or substantively important effects, either positive or negative.

(continued)

Appendix A4.3 Wilson Reading System® rating for the comprehension domain (continued)

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

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

Not met. No studies showed statistically significant or substantively important negative effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies showed statistically significant or substantively important positive effects. In addition, no studies showed a statistically significant or substantively important negative effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

Not met. No studies showed statistically significant negative effects.

AND

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

Met. No studies showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		School units	Students	
Alphabets	1	8	71	Small
Fluency	1	8	71	Small
Comprehension	1	8	71	Small
General reading achievement	0	0	0	na

na = not applicable/not studied

1. A rating of “moderate to large” requires at least two studies and two schools across studies in one domain, and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.”