

# What Works Clearinghouse



## Corrective Reading

**Program description<sup>1</sup>** *Corrective Reading* is designed to promote reading accuracy (decoding), fluency, and comprehension skills of students in third grade or higher who are reading below their grade level. The program has four levels that address students’ decoding skills and six levels that address students’ comprehension skills. All lessons in the program are sequenced and scripted.

*Corrective Reading* can be implemented in small groups of four to five students or in a whole-class format. *Corrective Reading* is intended to be taught in 45-minute lessons four to five times a week. For the single study reviewed in this report, only the word-level skills components of the *Corrective Reading* program were implemented.

**Research** One study of *Corrective Reading* met the What Works Clearinghouse (WWC) evidence standards. This study included 79 third-grade students in Pennsylvania.<sup>2</sup>

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

**Effectiveness** *Corrective Reading* was found to have potentially positive effects on alphabets and fluency and no discernible effects on comprehension.

	Alphabets	Fluency	Comprehension	General reading achievement
<b>Rating of effectiveness</b>	Potentially positive	Potentially positive	No discernible effects	na
<b>Improvement index<sup>3</sup></b>	Average: +9 percentile points Range: +1 to +13 percentile points	Average: +11 percentile points	Average: +7 percentile points Range: +2 to +11 percentile points	na

na = not applicable

1. The descriptive information for this program was obtained from publicly available sources: the program’s web site ([www.sraonline.com](http://www.sraonline.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 evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
3. These numbers show the average and range of student-level improvement indices for all findings in the study.

## Additional program information<sup>1</sup>

### Developer and contact

*Corrective Reading* is distributed by SRA/McGraw-Hill.  
Address: 220 East Danieldale Road Desoto, TX 75115-2490.  
Web: <http://www.sraonline.com/>. Telephone: (888) 772-4543.

### Scope of use

No information is available on the number of students or schools using the program.

### Teaching

The program's 45-minute lessons are designed for groups of up to 20 students, ideally five times a week. It has two components, decoding and comprehension,<sup>4</sup> each with four levels. The number of lessons varies by component and level. Decoding level A, designed for non-readers, has 65 lessons. It emphasizes basic decoding skills: rhyming, sounding out, sentence reading, and story reading. Levels B1 and B2 (65 lessons) are designed for struggling readers who do not read fluently or who confuse similar words. These levels teach students to become automatic decoders, with the increased self-confidence to read 90 words per minute by the end of B1 and 120 words per minute by the

end of B2. Level C (125 lessons) is designed for students who experience difficulty with vocabulary and complex sentence structures. This level bridges the gap between advanced word decoding skills and the ability to read informational text. All levels contain ongoing mastery tests and individual reading checkouts to assess individual student achievement.

*Corrective Reading* uses a direct instruction approach, a scripted presentation that uses a brisk pace, carefully chosen exercises and examples, and other presentation techniques. The publisher provides approximately seven hours of staff development that focuses on how to deliver direct instruction and use the program materials. Follow-up observations and coaching are recommended. A Teaching Tutor CD-Rom provides ongoing support for teachers using *Corrective Reading*.

### Cost

Prices range by level (A, B1, B2, C) and content (decoding, comprehension). The cost of student materials ranges from \$10 a student for level A programs to \$50 a student for level C materials. Teacher materials cost approximately \$200 per level.

## Research

Twenty-five studies reviewed by the WWC investigated the effects of *Corrective Reading*. One study (Torgesen et al., 2006) was a randomized controlled trial that met WWC evidence standards. The remaining 24 studies did not meet evidence screens.

Torgesen et al. (2006) examined the effects of *Corrective Reading* on 79 third-grade students in eight school units<sup>5</sup> in Pennsylvania.<sup>6</sup> Students in the comparison group participated in the regular reading program at their schools.

### 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.<sup>7</sup>

4. The comprehension component, which was not evaluated in studies reviewed by the WWC, is not fully described here.
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 of the beginning reading review, only data of the third graders were included in this review.
6. For the purposes of this study, only the word-level skill components of *Corrective Reading* 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.

## Research *(continued)*

The WWC considers the extent of evidence for *Corrective Reading* to be small for alphabetics, fluency, and comprehension.

### Effectiveness Findings

The WWC review of interventions for beginning reading addresses student outcomes in four domains: alphabetics, fluency, comprehension, and general reading achievement.<sup>8</sup> The study included here covers three domains: alphabetics, fluency, and comprehension. Within the alphabetics domain, the study reported on one construct: phonics.

*Alphabetics.* Torgesen et al. (2006) examined four outcomes under the phonics construct of the alphabetics 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 subtest). The authors reported statistically significant positive effects of *Corrective Reading* on two of these outcomes (WRMT–R word identification subtest and TOWRE sight word efficiency subtest). The statistical significance of these findings was confirmed by the WWC. The average effect size across the four outcomes was neither statistically significant nor large enough to be considered substantively important (that is, an effect size greater than 0.25).

*Fluency.* Torgesen et al. (2006) examined one outcome in this domain (the Oral Reading Fluency test) and reported statistically

significant positive effects on this outcome. The WWC analysis confirmed the statistical significance of the finding.

*Comprehension.* Torgesen et al. (2006) examined two outcomes in this domain (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 according to WWC criteria.

*Comprehension.* Torgesen et al. (2006) examined two outcomes in this domain (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 according to WWC criteria.

### 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,<sup>9</sup> 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 *Corrective Reading* to have potentially positive effects on alphabetics and fluency and no discernible effects on 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.

8. For definitions of the domains, see the [Beginning Reading Protocol](#).

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 *Corrective Reading*, no corrections for clustering or multiple comparisons were needed.

**The WWC found *Corrective Reading* to have potentially positive effects on alphabetic and fluency and no discernible effects on comprehension** *(continued)*

The average improvement index for alphabetic is +9 percentile points across all findings in the single study, with a range of +1 to +13 percentile points. The improvement index for the single outcome for fluency is +11 percentile points. The average improvement index for comprehension is +7 percentile points across all findings in the single study, with a range of +2 to +11 percentile points.

### Summary

The WWC reviewed 25 studies on *Corrective Reading*. One study met the WWC evidence standards.<sup>10</sup> Based on this one study, the WWC found potentially positive effects in the alphabetic and fluency domains and no discernible effects in the comprehension domain. The evidence presented in this report 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>

#### Did not meet WWC evidence screens

Arthur, C. (1988). Progress in a high school LD class. *ADI News*, 27(4), 17–18.<sup>11</sup>

Byron, D. (1988). Corrective Reading in a comprehensive school: The Hartcliffe Project. *Educational and Child Psychology*, 5(4), 35–41.<sup>12</sup>

Campbell, M. L. (1984). Corrective Reading program evaluated with secondary students in San Diego. *ADI News*, 3, 3.<sup>11</sup>

Department of Accountability and Organizational Evaluation. (2002). *Evaluation of the 2001–02 Corrective Reading program*. Retrieved from San Juan Unified School District Web site: <http://www.sanjuan.edu/accountability/program-evaluations/corrective-reading-2002.pdf><sup>11</sup>

Drakeford, W. (2002). The impact of an intensive program to increase the literacy skills of incarcerated youth. *Journal of Correctional Education*, 53(4), 139–144.<sup>11</sup>

Gunn, B., Smolkowski, K., & Biglan, A. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and Non-Hispanic students. *Journal of Special Education*, 39(2), 66–85.<sup>13</sup>

Gunn, B., Smolkowski, K., Biglan, A. & Black, C. (2005). Supplemental instruction in decoding skills for Hispanic and Non-Hispanic students in early elementary school: A follow-up. *Journal of Special Education*, 36(2), 69–80.<sup>13</sup>

#### Additional source:

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education*, 34(2), 90–103.

Harris, R. E., Marchand-Martella, N. E., Martella, R. C. (2000). Effects of a peer-delivered Corrective Reading program. *Journal of Behavioral Education*, 10, 21–36.<sup>11</sup>

Hempenstall, K. J. (1997). *The effects on the phonological processing skills of disabled readers participating in Direct Instruction reading programs*. Unpublished doctoral dissertation.

10. A single-case design study was identified but is not included in this review because the WWC does not yet have standards for reviewing single-case design studies.
11. The sample is not appropriate to this review: the parameters for this WWC review specified that students should be in grades kindergarten through third grade during the time of the intervention; this study did not focus on the targeted grades.
12. Does not use a strong causal design: this study did not use a comparison group.
13. Does not use a causal design: this study, which used a randomized controlled trial design, combined two interventions and therefore the effects of *Corrective Reading* could not be isolated.

## References *(continued)*

- tation, Royal Melbourne Institute of Technology, Melbourne, Victoria, Australia.<sup>14</sup>
- Herr, C. M. (1989). Using Corrective Reading with adults. *ADI News*, 8(2), 18–21.<sup>11</sup>
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80(4), 437–447.<sup>12</sup>
- Keel, M. C., Federick, L. D., Hughes, T. A., & Owens, S. H. (1999). Using paraprofessionals to deliver Direct Instruction reading programs. *Effective School Practices*, 18(2), 16–22.<sup>12</sup>
- Malmgren, K. W., & Leone, P. E. (2000). Effects of a short-term auxiliary reading program on the reading skills of incarcerated youth. *Education & Treatment of Children*, 23, 239–247.<sup>11</sup>
- Marchand-Martella, N. E., & Martella, R. C. (2002). An overview and research summary of peer-delivered Corrective Reading instruction. *Behavior Analysis Today*, 3, 213–220.<sup>11</sup>
- Marchand-Martella, N. E., Martella, R. C., Bettis, D. F., & Riley Blakely, M. (2004). Project Pals: A description of a high school-based tutorial program using Corrective Reading and peer-delivered instruction. *Reading and Writing Quarterly*, 20, 179–201.<sup>11</sup>
- Marchand-Martella, N. E., Martella, R. C., Orlob, M., & Ebey, T. (2000). Conducting action research in a rural high school setting using peers as Corrective Reading instructors for students with disabilities. *Rural Special Education Quarterly*, 19(2), 20–29.<sup>11</sup>
- Polloway, E. A., Epstein, M. H., Polloway, C. H., Patton, J. R., & Ball, D. W. (1986). Corrective Reading program: An analysis of effectiveness with learning disabled and mentally retarded students. *Remedial and Special Education*, 7(4), 41–47.<sup>11</sup>
- Scarlato, M. C., & Asahara, E. (2004). Effects of Corrective Reading in a residential treatment facility for adjudicated youth. *Journal of Direct Instruction*, 4, 211–217.<sup>11</sup>
- Short, C., Marchand-Martella, N. E., Martella, R. C., & Ebey, T. L. (1999). The benefits of being high school Corrective Reading peer instructors. *Effective School Practices*, 18(2), 23–29.<sup>11</sup>
- Slaton, D. (2006). Effects of Corrective Reading on the reading abilities and classroom behaviors of middle school students with reading deficits and challenging behavior. *Behavioral Disorders* 31(3), 265–283.<sup>11</sup>
- Sommers, J. (1995). Seven-year overview of Direct Instruction programs used in basic skills classes at Big Piney Middle School. *Effective School Practices*, 14(4), 29–32.<sup>11</sup>
- Somerville, D. E., & Leach, D. J. (1988, February). Direct or indirect instruction: An evaluation of three types of intervention programs for assisting students with specific reading difficulties. *Educational Research*, 30(1), 46–53.<sup>11</sup>
- Steventon, C. E., & Frederick, L. D. (2003). The effects of repeated readings on student performance in the Corrective Reading program. *Journal of Direct Instruction*, 3(1), 17–27.<sup>11</sup>
- Vitale, M., Medland, M., Romance, N., & Weaver, H. P. (1993). Accelerating reading and thinking skills of low-achieving elementary students: Implications for curricular change. *Effective School Practices*, 12(1), 26–31.<sup>11</sup>

### Disposition Pending

- Flores, M. M., Shippen, M. E., Alberto, P., & Crowe, L. (2004). Teaching letter-sound correspondence to students with moderate intellectual disabilities. *Journal of Direct Instruction*, 4, 173–188.<sup>15</sup>

---

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

14. The sample is not appropriate to this review: the parameters for this WWC review specified that students should be in grades K–3; this study did not disaggregate students in the eligible range from those outside the range.

15. The disposition is pending development of WWC evidence standards for single subject designs.

# Appendix

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

Characteristic	Description
<b>Study citation</b>	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: <a href="http://www.ed.gov/rschstat/eval/disadv/title1interimreport/index.html">http://www.ed.gov/rschstat/eval/disadv/title1interimreport/index.html</a>
<b>Participants</b>	The study design was based on random assignment of 37 school units <sup>1</sup> to one of four interventions, <i>Corrective Reading</i> , <i>Kaplan SpellRead</i> , <i>Failure Free Reading</i> , or <i>Wilson Reading</i> . Within each school, students were randomly assigned to the intervention or to the comparison condition. This report focuses on eight school units assigned to <i>Corrective Reading</i> . <sup>2</sup> At the time of analysis, the sample included 79 third-grade students (44 in intervention and 35 in comparison groups). The number of students at baseline was not reported. <sup>3</sup> About 44% of the intervention group students were female, compared with 52% in the comparison group. About 42% of the intervention group students were eligible for free/reduced lunch program and 48% in the comparison group.
<b>Setting</b>	Eight school units in Pennsylvania.
<b>Intervention</b>	<i>Corrective Reading</i> was implemented by nine teachers beginning in the first week in November 2003 through the first weeks in May 2004. The intervention was administered to groups of three students with heterogeneous basic reading skills. The average skills of each three-student instructional group determined the pace of instruction. Implementation fidelity was determined by reading program trainers who observed the teachers and coached them over a period of months, by project coordinators who observed a sample of instructional sessions, and by ratings based on a sample of videotaped sessions. Implementation was rated as acceptable. For purposes of this study, only the word-level skills components were implemented. However, the complete program of <i>Corrective Reading</i> also contains instructional routines and materials that focus on comprehension and vocabulary.
<b>Comparison</b>	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.
<b>Primary outcomes and measurement</b>	The primary outcome measures in the alphabetic domain were the word identification and word attack subtests of the Woodcock Reading Mastery Test–Revised (WRMT–R) and the phonetic decoding efficiency subtest of the Test of Word Reading Efficiency (TOWRE). The primary measures in the fluency domain were the TOWRE sight word efficiency subtest and the Oral Reading Fluency test. The primary measures in the comprehension domain were the passage comprehension subtests of WRMT–R and the Group Reading Assessment and Diagnostic Evaluation (GRADE). (See Appendix A2.1–2.3 for more detailed descriptions of outcome measures.)
<b>Teacher training</b>	Professional development included training and coaching by reading program staff, independent study of program materials, and telephone conferences. On average, intervention group teachers participated in 70.8 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. Findings on *Kaplan SpellRead*, *Failure Free Reading*, and *Wilson Reading* are included in other WWC beginning reading reports.
3. The study reported that 10 students in the intervention 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 that was also examined in this study. The fifth-grade sample that was 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>	
<b>Test of Word Reading Efficiency (TOWRE): Phonetic Decoding Efficiency subtest</b>	The TOWRE is a standardized 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).
<b>TOWRE: Sight Word Efficiency subtest</b>	The TOWRE is a standardized 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).
<b>Woodcock Reading Mastery Test–Revised (WRMT–R): Word Identification subtest</b>	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).
<b>WRMT–R: Word Attack subtest</b>	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
<b>Edformation Oral Fluency Assessment</b>	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>	
<b>Group Reading Assessment and Diagnostic Evaluation (GRADE): Passage Comprehension subtest</b>	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).
<b>WRMT–R: Passage Comprehension subtest</b>	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 alphabetics domain<sup>1</sup>

Outcome measure	Study sample	Sample size (school units/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> (Corrective Reading – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			Corrective Reading group	Comparison group				
<b>Torgesen et al., 2006 (randomized controlled trial)<sup>7</sup></b>								
TOWRE: Phonetic Decoding Efficiency subtest	Grade 3	8/79	89.86 (15.00)	89.48 (15.00)	0.38	0.02	ns	+1
TOWRE: Sight Word Efficiency subtest	Grade 3	8/79	90.98 (15.00)	86.41 (15.00)	4.57	0.30	Statistically significant	+12
WRMT–R: Word Identification subtest	Grade 3	8/79	91.06 (15.00)	87.77 (15.00)	3.30	0.22	Statistically significant	+9
WRMT–R: Word Attack subtest	Grade 3	8/79	100.34 (15.00)	95.15 (15.00)	5.18	0.34	ns	+13
<b>Domain average<sup>8</sup> for alphabetics</b>						0.22	na	+9

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 socioeconomic status. The study found no differences for alphabetics outcomes as a function of students' initial level of word attack scores. The study reported that Corrective Reading had a statistically significant positive effect on alphabetics for students with high pretest PPVT scores and no effects on students with low pretest PPVT scores. An additional subgroup analysis found statistically significant positive effects on the TOWRE sight word efficiency subtest for students not eligible for the free/reduced lunch program, but not for students eligible for the free/reduced lunch program.
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. The standard deviations in Torgesen et al. (2006) were the population standard deviations for these standardized 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 alphabetics domain, corrections for multiple comparisons were needed because the study's reported corrections for multiple comparisons were based on a grouping of outcomes that differed from the groupings 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 domain<sup>1</sup>

Outcome measure	Study sample	Authors' findings from the study						
		Sample size (school units/ students <sup>3</sup> )	Mean outcome (standard deviation <sup>2</sup> )		WWC calculations			
			Corrective Reading group	Comparison group	Mean difference <sup>4</sup> (Corrective Reading – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
<b>Torgesen et al., 2006 (randomized controlled trial)<sup>7</sup></b>								
Oral Reading Fluency	Grade 3	8/79	66.04 (39.20)	55.33 (39.20)	10.71	0.27	Statistically significant	+11
<b>Domain average<sup>9</sup> for fluency</b>						0.27	Statistically significant	+11

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. The study found statistically significant positive effects on the Oral Fluency Test for students not eligible for the free/reduced lunch program, but not for students eligible for the free/reduced lunch program. No differences were found between initial skill level subgroups of students for the fluency outcome.
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. The standard deviations in Torgesen et al. (2006) were the population standard deviations for these standardized outcomes.
3. The sample size for the analysis was not reported in the study. The sample size reported is the total number of third-grade students in the intervention and control conditions at baseline, which may differ from the actual number of students used in the analysis in the report.
4. 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.
5. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. 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.
8. 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 fluency domain, 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.
9. This row provides the domain average, which in this instance is also the single study finding.

### Appendix A3.3 Summary of study findings included in the rating for the comprehension domain<sup>1</sup>

Outcome measure	Study sample	Authors' findings from the study						
		Sample size (school units/ students <sup>3</sup> )	Mean outcome (standard deviation <sup>2</sup> )		WWC calculations			
			Corrective Reading group	Comparison group	Mean difference <sup>4</sup> (Corrective Reading – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
<b>Torgesen et al., 2006 (randomized controlled trial)<sup>7</sup></b>								
WRMT: Passage Comprehension subtest	Grade 3	8/79	93.16 (15.00)	92.30 (15.00)	0.86	0.06	ns	+2
GRADE: Passage Comprehension subtest	Grade 3	8/79	87.39 (15.00)	83.22 (15.00)	4.18	0.28	ns	+11
<b>Domain average<sup>9</sup> for comprehension</b>						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. The standard deviations in the Torgesen et al. (2006) study were the population standard deviations for these standardized outcomes.
3. The sample size for the analysis was not reported in the study. The sample size reported is the total number of third-grade students in the intervention and control conditions at baseline, which may differ from the actual number of students used in the various analysis in the report.
4. 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.
5. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. 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.
8. 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 corrections for multiple comparisons were needed because the study's reported corrections for multiple comparisons were based on the same group of outcomes as the domain for this review.
9. 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 *Corrective Reading* rating for the alphabetics 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.<sup>1</sup>

For the outcome domain of alphabetics, the WWC rated *Corrective Reading* as potentially positive. 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 *Corrective Reading* 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 Corrective Reading 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.<sup>1</sup>

For the outcome domain of fluency, the WWC rated *Corrective Reading* as potentially positive. It did not meet the criteria for positive effects because only one study showed statistically significant positive effects. The remaining ratings (mixed, no discernible effects, potentially negative, or negative) were not considered because *Corrective Reading* 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 A4.3 Corrective Reading 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.<sup>1</sup>

For the outcome domain of comprehension, the WWC rated *Corrective Reading* as having no discernible effects. It did not meet the criteria for other ratings (positive, potentially positive, mixed, potentially negative, and negative) 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 *Corrective Reading* 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 <sup>1</sup>
		School units	Students	
Alphabets	1	8	More than 70	Small
Fluency	1	8	More than 70	Small
Comprehension	1	8	More than 70	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.”