What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES

Beginning Reading March 19, 2007

Reading Recovery®

Program description

Reading Recovery® is a short-term tutoring intervention program intended to serve the lowest achieving (bottom 20%) first-grade students. Students are chosen for *Reading Recovery*® by school staff, and selection is based on prior reading achievement, diagnostic testing (the Clay Observation Survey of Early Literacy Achievement), and teacher recommendations. The goals of *Reading Recovery*® are to promote literacy skills and reduce the number of first-grade students who are struggling to read. The program supplements classroom teaching with one-on-one tutoring sessions, generally conducted as pull-out sessions during the school day. Tutoring, which is conducted by trained *Reading Recovery*® teachers, takes place daily for 30 minutes over 12–20 weeks.

Research

Four studies of *Reading Recovery*® met the What Works Clearinghouse (WWC) evidence standards, and one study met WWC evidence standards with reservations. These five studies included about 700 first-grade students attending elementary schools in diverse settings across the United States. All studies focused on low-achieving students who received the *Reading Recovery*® intervention in first grade. Generally, outcomes at the end of first grade were used by the WWC to calculate a rating of effectiveness.[1] In one study, longer range effects were included.[2]

Effectiveness

Reading Recovery® was found to have positive effects on students' alphabetics skills and general reading achievement outcomes. The program was found to have potentially positive effects on comprehension and fluency.

	Alphabetics	*Fluency*	*Comprehension*	*General reading achievement*
Rating of effectiveness	Positive effects	Potentially positive effects	Potentially positive effects	Positive effects
Improvement index[3]	Average: +34 percentile points Range: –10 to +50 percentile points	Average: +46 percentile points Range: +32 to +49 percentile points	Average: +14 percentile points Range: +6 to +21 percentile points	Average: +32 percentile points Range: –5 to +50 percentile points

1. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
2. Additional findings on outcomes measured at later time points are shown in Appendix A4.4.
3. These numbers show the average and range of improvement indices for all findings across the studies.

Additional program information

Developer and contact

Developed by Dr. Marie M. Clay, University of Auckland, New Zealand. Distributed through the *Reading Recovery*® Council of North America (RRCNA). Address: 400 West Wilson Bridge Road, Suite 250, Worthington, OH 43085-5218. Email: jjohnson@readingrecovery.org. Web: http://www.readingrecovery.org. Telephone: (614) 310-7323.

Scope of use

Reading Recovery® was developed in the mid-1970s by Dr. Clay, who first tested the program in New Zealand. According to the RRCNA, more than 1.5 million first graders in 48 states and the Department of Defense Dependents Schools have been served in the United States since *Reading Recovery*® was introduced in 1984. *Reading Recovery*® is also used in New Zealand, Australia, Canada, and the United Kingdom.

Teaching

Lessons incorporate principles of early intervention and individual tutoring and include attention to phonological awareness, phonics/decoding skills, vocabulary/word work, fluency, and comprehension. Each *Reading Recovery*® lesson consists of reading familiar and novel stories, manipulating letters and words, and writing and assembling stories. Lessons are interactive between teacher and student, with the teacher carefully monitoring each child's reading behavior. *Reading Recovery*® lessons are discontinued when children demonstrate the ability to consistently read at the average level for their grade—between weeks 12 and 20 of the program. Those who make progress but do not reach average classroom performance after 20 weeks are referred for further evaluation and a plan for future action. Teacher training includes a one-year, university-based training program.

Cost

Reading Recovery® is available on a nonprofit, no royalty basis. Costs for the program involve start-up costs and ongoing costs. To establish a *Reading Recovery*® site—a district or group of districts representing multiple schools—a teacher leader must first be trained. This start-up cost includes paying salary, paying university tuition for the *Reading Recovery*® coursework, and covering the costs of books and materials. Sites must also build a one-way mirror and sound system to monitor training for the teachers. In addition to salary, travel, and program support costs for the teacher leader, costs for teachers include paying salaries and benefits for the time they dedicate to *Reading Recovery*® and paying tuition for training. Books and materials for lessons and evaluation as well as ongoing professional development for both teacher leaders and teachers should also be figured into the costs.

In addition to the teacher training described above, in 2006 the cost of program materials was approximately $100 per student served (calculated by the RRCNA as an average over the past five years, 2002–06). Sites pay an annual data evaluation fee of $250 a site plus $3.50 per student served. Sites implementing the program also pay annual technical support fees, which vary by the university that provides the *Reading Recovery*® training. Because of the cost and staff needed for the intervention, a typical school with one *Reading Recovery*® teacher will serve 4 or 5 students a semester.

Research

The WWC reviewed 78 studies that examined the effects of *Reading Recovery*®. Four studies (Baenen, Bernhole, Dulaney, & Banks, 1997; Pinnell, DeFord, & Lyons, 1988; Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994; and Schwartz, 2005) met WWC evidence standards. One study (Iverson & Tunmer, 1993) met WWC standards with reservations. The remaining 73 studies did not meet WWC evidence screens.[4]

4. Because *Reading Recovery*® is designed to improve the reading skills of low-achieving first-grade readers, the appropriate comparison groups for determining the intervention's effectiveness are similar low-achieving first-grade readers who did not receive *Reading Recovery*®. Many of the studies screened did not meet evidence standards because they used inappropriate comparison groups, such as higher achieving first-grade readers, to draw conclusions about the effectiveness of the program.

Met evidence standards

Baenen et al. (1997) was a randomized controlled trial that focused on first-grade students from Wake County, North Carolina. The WWC review focuses on the outcomes of students who qualified for and were randomly assigned to either the *Reading Recovery*® intervention or a comparison group. From an original sample size of 168, outcomes were assessed at three time points: end of first grade (n = 147), end of second grade (n = 147), and end of third grade (n = 127). Although the WWC used only the results at the end of first grade to determine the intervention rating, information on the additional findings can be found in Appendix A4.4.

Pinnell, DeFord, and Lyons (1988) was a randomized controlled trial. The study sample was first-grade students distributed across 14 schools in Columbus, Ohio. Two groups were formed by randomly assigning students to an intervention group, which received *Reading Recovery*® in addition to their regular classroom instruction (n = 38), or to a control group, which received an alternate compensatory program (n = 53). This comparison met WWC evidence standards.[5]

Pinnell et al. (1994) was a randomized controlled trial that randomly assigned 10 low-achieving first-grade students in each of 10 Ohio schools. The WWC review focuses only on the eight schools that successfully implemented randomization for the intervention (n = 31) and comparison (n = 48) conditions.[6]

Schwartz (2005) was a randomized controlled trial of first-grade students from 14 states. The WWC focused on the 37 students across several schools who were randomly assigned to receive the intervention during the first half of the year. The other 37 students, who were randomly assigned to receive the intervention during the second half of the year, served as the comparison group during the first half of the year.[7] The groups were compared at mid-year, before the comparison group had begun receiving *Reading Recovery*®.

Met evidence standards with reservations

Iverson and Tunmer (1993) was a quasi-experimental design study that included first-grade students from 30 school districts in Rhode Island. The study compared outcomes for students participating in *Reading Recovery*® (n = 32) with students in a comparison group who did not receive *Reading Recovery*® (n = 32), who were matched on the basis of pretest scores.[8] The comparison group received standard small group, out-of-class support services.

5. A third group of students qualified for and received *Reading Recovery*® outside of regular classroom instruction, but were also taught by a *Reading Recovery*®–trained teacher when they were in their regular classroom (n = 96). Although this comparison met evidence standards with reservations, it was not considered in the intervention rating because it went beyond the standard delivery of the program. However, results are reported in Appendices A4.1–A4.3.
6. Although the original study included analyses of additional interventions implemented at additional schools, only the schools that randomly assigned students to *Reading Recovery*® or the comparison group were relevant to this review. For more details about the original study, see Appendix A1.3.
7. Assessments were also made at the end of the year, but they were not appropriate for the WWC's analysis because by then both groups of low-achieving students had received the intervention. Additional comparison groups of low-average and high-average readers were not used by the WWC because these students were not eligible for *Reading Recovery*®.
8. The study also included a third group of students (n = 32) who used a modified version of *Reading Recovery*®, which provided explicit instruction in letter-phoneme patterns instead of the letter identification segment. This group was also compared with the comparison group. Although this comparison met evidence standards with reservations, it was not considered in the intervention rating because it went beyond the standard delivery of the program. However, results are reported in Appendices A4.1 and A4.3.

Effectiveness

Findings

The WWC review of interventions for beginning reading addresses student outcomes in four domains: alphabetics, reading fluency, comprehension, and general reading achievement.[9] *Reading Recovery®* studies included in this report cover all four domains and most of the constructs within each domain. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Reading Recovery®* on students. The results are presented by domain and construct for all the *Reading Recovery®* studies that the WWC reviewed.

For the four beginning reading domains, subtests of the Clay Observation Survey were used in some of the studies. The Clay Observation Survey was developed by Dr. Marie Clay, who also developed *Reading Recovery®*.

Alphabetics. Two studies examined the effects of *Reading Recovery®* on the phonemic awareness construct. Schwartz (2005) reported no statistically significant effects for the phonemic awareness measures—the deletion task and the Yopp-Singer Phoneme Segmentation Test—but the effects on both measures were positive and considered substantively important based on the WWC criteria (that is, at least 0.25). Iverson and Tunmer (1993) reported, and the WWC confirmed, statistically significant positive effects of the *Reading Recovery®* intervention on two phonemic awareness measures—a phoneme deletion task and the Yopp-Singer Phoneme Segmentation Test.

Three studies examined the effects of *Reading Recovery®* on the print awareness construct in the alphabetics domain. Pinnell, DeFord, and Lyons (1988) reported, and the WWC confirmed, a statistically significantly positive effect of *Reading Recovery®* on the Concepts about Print subtest of the Observation Survey of Early Literacy Achievement. Schwartz (2005) reported, and the WWC confirmed, a statistically significant positive effect of *Reading Recovery®* on the Concepts about Print subtest of the Observation Survey. Iverson and Tunmer (1993) found a statistically significant positive effect of *Reading Recovery®* on the Concepts about Print subtest of the Observation Survey. The significance of the effect was confirmed by the WWC.

Three studies examined the effects of *Reading Recovery®* on the letter knowledge construct in the alphabetics domain. Pinnell, DeFord, and Lyons (1988) did not find a statistically significant effect for *Reading Recovery®* on the Letter Identification subtest of the Observation Survey. Schwartz (2005) reported a statistically significant positive effect of *Reading Recovery®* on the Letter Identification subtest of the Observation Survey, but according to WWC criteria this effect was not statistically significant or large enough to be considered substantively important.[10] Iverson and Tunmer (1993) found, and the WWC confirmed, a statistically significant positive effect of *Reading Recovery®* on the Letter Identification subtest of the Observation Survey.

Three studies examined the effects of *Reading Recovery®* on the phonics construct of the alphabetics domain. Pinnell, DeFord, and Lyons (1988) found a statistically significant positive effect on the Word Recognition subtest of the Observation Survey. In WWC calculations, there was no statistically significant effect, but the positive effect was large enough to be considered substantively important. Schwartz (2005) found, and the WWC confirmed, a statistically significant positive effect of *Reading Recovery®* on the Word Recognition subtest of the Observation Survey. Iverson and Tunmer (1993) found statistically significant positive effects of *Reading Recovery®* on the Dolch Word Recognition Test, the Word Recognition subtest of the Observation Survey, and a pseudoword decoding task. The significance of the effects was confirmed by the WWC.

Overall, in the alphabetics domain, two studies with strong designs met WWC evidence standards and demonstrated statistically significant positive effects. One additional study met

9. For definitions of the domains, see the Beginning Reading Protocol.
10. In this case, the author did not control for pretest differences between groups; however, the WWC did account for pretest differences.

Effectiveness *(continued)*

WWC evidence standards with reservations and showed statistically significant positive effects.

Fluency. Schwartz (2005) found, and the WWC confirmed, positive and statistically significant effects of *Reading Recovery®* on the Slosson Oral Reading Test–Revised and the Text Reading Level subtest of the Observation Survey.

In the fluency domain, there was one study with a strong design that demonstrated statistically significant positive effects.

Comprehension. Two studies examined the effects of *Reading Recovery®* on the reading comprehension construct. Pinnell, DeFord, and Lyons (1988) found a positive and statistically significant effect of *Reading Recovery®* on the Reading Comprehension subtest of the Comprehensive Test of Basic Skills (CTBS). The significance of the effect was confirmed by the WWC. Schwartz (2005) reported no statistically significant effect of *Reading Recovery®* on the Degrees of Reading Power Test.

One study examined the effect of *Reading Recovery®* on the vocabulary construct of the comprehension domain. Pinnell, DeFord, and Lyons (1988) found, and the WWC confirmed, a positive and statistically significant effect of *Reading Recovery®* on the Reading Vocabulary subtest of the CTBS.

In the comprehension domain, there were two studies with strong designs. One study showed statistically significant positive effects, and the other study showed an indeterminate effect.

General reading achievement. Baenen et al. (1997) did not find a statistically significant effect of *Reading Recovery®* on grade retention. Pinnell, DeFord, and Lyons (1988) found, and the WWC confirmed, positive and statistically significant effects of *Reading Recovery®* on two subtests of the Observation Survey: Hearing and Recording Sounds in Words (Dictation) and Writing Vocabulary. Pinnell et al. (1994) found statistically significant positive effects of *Reading Recovery®* on the Gates-MacGinitie, the Dictation subtest of the Observation Survey, and the Woodcock Reading Mastery Test–Revised. The statistical significance of the effects was confirmed by the WWC. Schwartz (2005) found, and the WWC confirmed, positive and statistically significant effects of *Reading Recovery®* on two subtests of the Observation Survey: Dictation and Writing Vocabulary. Iverson and Tunmer (1993) found, and the WWC confirmed, statistically significant positive effects of *Reading Recovery®* on two subtests of the Observation Survey: Dictation and Writing Vocabulary.

In the general reading achievement domain, there were three studies with strong designs and statistically significant positive effects. One study had a strong design with indeterminate effects. One additional study met WWC evidence standards with reservations and demonstrated statistically significant positive effects.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings,[11] 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).

11. 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 Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In this report, such adjustments were made for Pinnell, DeFord, and Lyons (1988); Iverson and Tunmer (1993); and Pinnell et al. (1994).

The WWC found *Reading Recovery*® to have positive effects in the alphabetics and general reading domains and potentially positive effects in the fluency and comprehension domains

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 entirely based 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 alphabetics is +34 percentile points across three studies, with a range of –10 to +50 percentile points. For fluency, the average improvement index is +46 percentile points, with a range of +32 to +49 percentile points across outcomes in one study. For comprehension, the average improvement index is +14 percentile points across two studies, with a range of +6 to +21 percentile points. For the general reading domain, the average improvement index was +32 percentile points across five studies, with a range of –5 to +50 percentile points.

Summary

The WWC reviewed 78 studies that investigated the effects of *Reading Recovery*®. Four studies met WWC evidence standards and one met WWC evidence standards with reservations. Based on these five studies, the WWC found *Reading Recovery*® to have positive effects in the alphabetics and general reading achievement domains. *Reading Recovery*® was found to have potentially positive effects in the fluency and comprehension domains. The evidence presented in this report may change as new research emerges.

References

Met WWC evidence standards

Baenen, N., Bernhole, A., Dulaney, C., & Banks, K. (1997). Reading Recovery: Long-term progress after three cohorts. *Journal of Education for Students Placed at Risk, 2*(2), 161.

Additional sources:

Donley, J., Baenen, N., & Hundley, S. (1993, April). *A study of the long-term effectiveness of the Reading Recovery program.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Wake County Public School System. (1995). *Evaluation Report: WCPSS Reading Recovery 1990–94.* E&R Report No. 95.09A. Winston-Salem, NC: Wake County Public School System.

Pinnell, G. S., DeFord, D. E., & Lyons, C. A. (1988). *Reading Recovery: Early intervention for at-risk first graders* (Educational Research Service Monograph). Arlington, VA: Educational Research Service.

Additional sources:

Pinnell, G. S. (1988, January). *Success of at-risk children in a program that combines writing and reading* (Technical Report No. 417). Urbana-Champaign, IL: University of Illinois, Center for the Study of Reading.

Pinnell, G. S. (1989a). Reading Recovery: Helping at-risk children learn to read. *The Elementary School Journal, 90,* 161–183.

References *(continued)*

Pinnell, G. S. (1989b). Success of at-risk children in a program that combines writing and reading. In J. M. Mason (Ed.), *Reading and Writing Connections* (pp. 237–259). Boston: Allyn & Bacon.

Pinnell, G. S., Short, K. G., Lyons, C. A., & Young, P. (1986). *The Reading Recovery Project in Columbus, OH Year 1: 1985–1986.* Columbus, OH: Ohio State University.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly, 29*(1), 8–39.

Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology, 97*(2), 257–267.

Met WWC evidence standards with reservations

Iverson, S., & Tunmer, W. E. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology, 85*(1), 112–126.

Additional source:

Tunmer, W. E., & Hoover, W. A. (1993). Phonological recoding skills in beginning reading. *Reading and Writing: An Interdisciplinary Journal, 5,* 161–179.

Did not meet WWC evidence screens

Acalin, T. A. (1995). A comparison of Reading Recovery to Project READ. *Masters Abstracts International, 33*(06), 1660. (UMI No. 1361908)[12]

Ashdown, J., & Simic, O. (2003). Is early literacy intervention effective for English language learners? Evidence from Reading Recovery. In S. Forbes & C. Briggs (Eds.), *Research in Reading Recovery.* Portsmouth, NH: Heinemann.[13]

Askew, B. J., & Frasier, D. F. (1997). Sustained effects of Reading Recovery intervention on the cognitive behaviors of second grade children and the perceptions of their teachers. In S. L. Swartz & A. F. Klein (Eds.), *Research in Reading Recovery* (pp. 18–38). Portsmouth, NH: Heinemann.[13]

Bermel, S. (1987). *Language development component, CLEAR-Reading Recovery program 1985–86. Final evaluation report.* Columbus, OH: Columbus Public Schools, Ohio Department of Evaluation Services. (ERIC Document Reproduction Service No. ED281157)[14]

Brown, K. L. (1999). The impact of Reading Recovery intervention on the reading achievement of selected second grade students. *Dissertation Abstracts International*, *62*(08), 2636A. (UMI No. 3023650)[14]

Brown, W., Denton, E., Kelly, P., & Neal, J. (1999). Reading Recovery effectiveness: A five-year success story in San Luis Costal Unified School District. *ERS Spectrum*, *17*(1), 3–12.[13]

Bufalino, J. M. (1993). The sustained effects of Reading Recovery intervention on the reading comprehension of second graders. *Dissertation Abstracts International*, *54*(11), 145A. (UMI No. 9407866)[13]

Center, Y., Freeman, L., & Robertson, G. (2001). The relative effect of a code-oriented and a meaning-oriented early literacy program on regular and low progress Australian students in year 1 classrooms which implement Reading Recovery. *International Journal of Disability, Development and Education*, *48*(2), 207–232.[13]

Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly*, *30*(2), 240–263.[14]

Chapman, J., Tunmer, W. E., & Prochnow, J. E. (2001). Does success in the Reading Recovery program depend on developing proficiency in phonological-processing skills? A longitudinal study in a whole language instructional context. *Scientific Studies of Reading*, *5*(2), 141–176.[14]

Christman, M. S. (2003). An examination of the effects and costs of the Reading Recovery Program in an urban school district. *Dissertation Abstracts International*, *64*(08), 2824A. (UMI No. 3102259)[13]

Collins, E. W. (2000). The immediate and sustained effects of the Reading Recovery program on grade one and grade four

References *(continued)*

at-risk students: A longitudinal study. *Dissertation Abstracts International*, *61*(05), 1784A. (UMI No. 9971239)[13]

Collins, V. K. (1994). Automaticity in information processing. *Dissertation Abstracts International*, *55*(9), 2708A. (UMI No. 9502838)[13]

Curtin, J. (1993). *The effectiveness of the Reading Recovery Program on reading achievement*. Chicago: Chicago Public Schools. (ERIC Document Reproduction Service No. ED363863)[13]

DeFord, D. E. (1997). Early writing: Teachers and children in Reading Recovery. In S. L. Swartz & A. F. Klein (Eds.), *Research in Reading Recovery* (pp. 148–172). Portsmouth, NH: Heinemann.[15]

Department of Evaluation Services. (1995). *Compensatory Education (CE) product evaluation: Elementary and secondary programs 1994–95.* Saginaw, MI: Saginaw Public Schools. (ERIC Document Reproduction Service No. ED391853)[16]

Additional source:

Department of Evaluation Services. (1992). *Compensatory education product evaluation: Elementary and secondary programs 1991–1992*. Saginaw, MI: Saginaw Public Schools, Michigan Department of Evaluation Services. (ERIC Document Reproduction Service No. ED350370)

Dunkeld, C. (1990). *Gaining experience with Reading Recovery: A pilot project between Portland Public Schools and Portland State University.* Portland, OR: Portland State University. (ERIC Document Reproduction Service No. ED321246)[17]

Escamilla, K. (1997). Descrubriendo la lectura: An early intervention literacy program in Spanish. In S. L. Swartz & A. F. Klein (Eds.), *Research in Reading Recovery* (pp. 109–121). Portsmouth, NH: Heinemann.[18]

Escamilla, K., Loera, M., Ruiz, O., & Rodriguez, Y. (2003). An examination of sustaining effects in Descubriendo la Lectura programs. In S. Forbes & C. Briggs (Eds.), *Research in Reading Recovery.* Portsmouth, NH: Heinemann.[18]

Evans, T. L. P. (1996). 'I can read deze books!': A qualitative comparison of the Reading Recovery program and a small group reading intervention. *Dissertation Abstracts International*, *57*(02), 565A. (UMI No. 9619200)[19]

Fullerton, S. K., & DeFord, D. E. (2001, December). *Conversations before writing during Reading Recovery lessons: Negotiations or tug of war?* Paper presented at the meeting of the National Reading Conference, Chicago.[15]

Gilmer, V. B. (2003). Sustained success of former Reading Recovery students. *Dissertation Abstracts International*, *64*(02), 444A. (UMI No. 3081577)[13]

Gomez-Bellenge, F. X., Rodgers, E., & Fullerton, S. K. (2003). *Reading Recovery and Descubriendo la Lectura national report 2001–2002*. Columbus, OH: Ohio State University.[18]

Gomez-Bellenge, F. X., & Rodgers, E. M. (2004). *Reading Recovery and Descubriendo la Lectura national report 2002–2003*. Columbus, OH: Ohio State University, College of Education, School of Teaching and Learning.[13]

Hovest, C. M. (2000). An examination of the achievement of phonological skills for three groups participating in an early intervention program. *Dissertation Abstracts International*, *61*(08), 3107A. (UMI No. 9982583)[13]

Johnson, J. A. (1996). *Reading Recovery: Early intervention*. Hays, KS: Fort Hays State University. (ERIC Document Reproduction Service No. ED398564)[14]

LaFave, C. E. (1995). Impact of Reading Recovery on phonemic awareness. *Dissertation Abstracts International*, *56*(07), 2621A. (UMI No. 9540370)[14]

Leitner, D. (1990). *Portland Reading Recovery program 1989–90 evaluation report. Year 1.* Portland, OR: Portland Public Schools, Oregon Research and Evaluation Department. (ERIC Document Reproduction Service No. ED326839)[15]

Litt, D. G. (2003). An exploration of the double-deficit hypothesis in the Reading Recovery population. *Dissertation Abstracts International*, *64*(06), 2028A. (UMI No. 3094512)[13]

Lyons, C. (1989). Reading Recovery: A preventative for mislabeling young 'at-risk' learners. *Urban Education*, *24*(2), 125–139.[15]

References *(continued)*

Marina, B., & Gilman, D. A. (2003). *Is Reading Recovery worth the cost?* Vigo County, IN: Vigo County School Corporation. (ERIC Document Reproduction Service No. ED473957)[13]

Marvin, C. A., & Gaffney, J. S. (2003). The effects of Reading Recovery on children's home literacy experiences. In S. Forbes & C. Briggs (Eds.), *Research in Reading Recovery.* Portsmouth, NH: Heinemann.[13]

McManus, D. A. (1999). A modified Reading Recovery program can be successful in a second grade Title I reading program. *Dissertation Abstracts International*, *60*(10), 3597A. (UMI No. 9947733)[20]

Miller, S. D. (2003). Partners-in-Reading: Using classroom assistants to provide tutorial assistance to struggling first-grade readers. *Journal of Education for Students Placed at Risk*, *8*(3), 333–349.[21]

Moore, M., & Wade, B. (1998). Reading Recovery: Its effectiveness in the long term. *Support for Learning*, *13*(3), 123–128.[13]

Murphy, J. A. (2003). An application of growth curve analysis: The evaluation of a reading intervention program. *Dissertation Abstracts International*, *64*(12), 4358A. (UMI No. 3114448)[14]

Neal, J. C., & Kelly, P. R. (2003). The success of Reading Recovery for English language learners and Descubriendo la Lectura for bilingual students in California. In S. Forbes & C. Briggs (Eds.), *Research in Reading Recovery.* Portsmouth, NH: Heinemann.[18]

O'Connor, E. A., & Simic, O. (2002). The effect of Reading Recovery on special education referrals and placements. *Psychology in the Schools*, *39*(6), 635–646.[22]

Pinnell, G. S. (1997). An inquiry-based model for educating teachers of literacy. In S. L. Swartz & A. F. Klein (Eds.), *Research in Reading Recovery* (pp. 6–17). Portsmouth, NH: Heinemann.[22]

Plewis, I. (2000). Evaluating educational interventions using multilevel growth curves: The case of Reading Recovery. *Educational Research and Evaluation*, *6*(1), 83–101.[14]

Pollock, J. S. (1993). *Final evaluation report: Reading Recovery™ program 1991–92*. Columbus, OH: Columbus Public Schools, Ohio Department of Program Evaluation. (ERIC Document Reproduction Service No. ED358440)[15]

Pollock, J. S. (with Morgan, K. L., Williams, E. J., & Amorose, R. A.). (1991). *Reading Recovery program 1990–91. Final Evaluation report. Elementary and Secondary Education Act—Chapter 1*. Columbus, OH: Columbus Public Schools, Ohio Department of Program Evaluation. (ERIC Document Reproduction Service No. ED343108)[15]

Pollock, J. S., Williams, E. J., Morgan, K. L., & Amorose, R. A. (1990). *Language development component compensatory language experiences and reading, CLEAR-Reading Recovery program, 1989–90. Final Evaluation Report*. Columbus, OH: Columbus Public Schools. (ERIC Document Reproduction Service No. ED327821)[15]

Potter, T. (2004). *Reading Recovery evaluation*. Madison, WI: Planning, Research and Evaluation, Madison Metropolitan School District.[14]

Quay, L. C., Steele, D. C., Johnson, C. I., & Hortman, W. (2001). Children's achievement and personal and social development in a first-year Reading Recovery program with teachers in training. *Literacy Teaching and Learning*, *5*, 7–25.[23]

Ramaswami, S. (1994). *The differential impact of Reading Recovery on achievement of first graders in the Newark School District, 1991–1993*. Newark, NJ: Newark Board of Education, Office of Planning, Evaluation and Testing.[13]

Rhodes, J. A. (1998). A comparison of the effects of individualized writing instruction with and without phonemic segmentation on the standard spelling performance of at-risk first graders. *Dissertation Abstracts International*, *59*(07), 2426A. (UMI No. 9839187)[24]

Rodgers, E., Gomez-Bellenge, F. X., & Fullerton, S. K. (2003). *Reading Recovery in Ohio: 2001–2002 state report* (National Data Evaluation Center Tech. Rep. No. 2003-03). Columbus, OH: Ohio State University, College of Education, School of Teaching and Learning.[13]

Rodgers, E. M., Fullerton, S. K., & DeFord, D. E. (2001). What does it take to reform instructional practices? In J.

References *(continued)*

V. Hoffman, D. L. Schallert, C. M. Fairbanks, J. Worthy, & B. Maloch (Eds.), *Fiftieth yearbook of the National Reading Conference* (pp. 519–532). Chicago: National Reading Conference.[22]

Rodgers, E. M., Gomez-Bellenge, F. X., & Schulz, M. M. (2005). *Reading Recovery in Ohio: 2003–2004 state report* (National Data Evaluation Center Tech. Rep. No. 2005-01). Columbus, OH: Ohio State University, College of Education, School of Teaching and Learning.[13]

Roehrig, A. D., Pressley, M., & Sloup, M. (2001). Reading strategy instruction in regular primary-level classrooms by teachers trained in Reading Recovery. *Reading & Writing Quarterly*, *17*(4), 323–348.[22]

Ross, S. M., Nunnery, J. A., & Smith, L. J. (1996). *Evaluation of Title I reading programs: Amphitheater public schools—Year 1: 1995–1996*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.[25]

Salinas, A., Williams, I., & Kohler, L. (1993). *Reading Recovery program evaluation: Report 1992-93.* McAllen, TX: McAllen Independent School District, Office of Research and Evaluation. (ERIC Document Reproduction Service No. ED366938)[15]

Schmitt, M. C. (2001). The development of children's strategic processing in Reading Recovery. *Reading Psychology*, *22*, 129–151.[26]

Schmitt, M. C. (2003). Metacognitive strategy knowledge: Comparison of former Reading Recovery children and their current classmates. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, *7*(1–2), 57–76.[26]

Schmitt, M. C., & Gregory, A. E. (2001, June). *The impact of early intervention: Where are the children now?* Paper presented at the meeting of the National Reading Conference, San Antonio, TX.[13]

Schotanus, H. (1991). *Reading Recovery pilot project (Laws 1989: 301): Report of results and effectiveness*. Concord, NH: New Hampshire State Department of Education. (ERIC Document Reproduction Service No. ED363859)[15]

Schotanus, H., Chase, C., & Fontaine, A. (1992). *Reading Recovery program, implementation year two, school year 1991–1992. Report of results and effectiveness*. Concord, NH: New Hampshire State Department of Education. (ERIC Document Reproduction Service No. ED363860)[15]

Schotanus, H., Chase, C., Fontaine, A., & Tilton, S. (1993). *Reading Recovery program implementation year three, school year 1992–1993. Report of results and effectiveness*. Concord, NH: New Hampshire Department of Education. (ERIC Document Reproduction Service No. ED364869)[15]

Schotanus, H., Fontaine, A., Tilton, S., Westergren, G., & Anderson, A. (1996). *Reading Recovery program implementation year six, school year 1995–96. Report of results and effectiveness.* Concord, NH: New Hampshire Department of Education. (ERIC Document Reproduction Service No. ED405573)[14]

Shoulders, M. D. (2004). The long-term effectiveness of the Reading Recovery Program. *Dissertation Abstracts International*, *65*(03), 836A. (UMI No. 3127549)[13]

Simpson, S. H. (1997). A principal's perspective of the implementation of Reading Recovery in six metropolitan Nashville elementary schools. *Dissertation Abstracts International*, *58*(08), 2948A. (UMI No. 9806596)[13]

Smith, N. (1994). Reading Recovery data and observations from one Illinois site (Part II). *Illinois Reading Council Journal*, *22*(3), 29–46.[14]

Additional source:

Smith, N. (1994). Reading Recovery data and observations from one Illinois site (Part I). *Illinois Reading Council Journal*, *22*(2), 7–27.

Smith, P. E. (1994). Reading Recovery and children with English as a second language. *New Zealand Journal of Educational Studies*, *29*(2), 141–155.[13]

Stahl, K. A. D., Stahl, S. A., & McKenna, M. C. (2004). The development of phonological awareness and orthographic processing in Reading Recovery. In S. Forbes & C. Briggs (Eds.), *Research in Reading Recovery.* Portsmouth, NH: Heinemann.[14]

References *(continued)*

Thomas, P. M. (with Morgan, K. L., Amorose, R. A., & Williams, E. J.). (1989). *Language development component, compensatory language experiences and reading: CLEAR—Reading Recovery program 1987–88. Final evaluation report*. Columbus, OH: Columbus Public Schools, Ohio Department of Evaluation Services. (ERIC Document Reproduction Service No. ED313668)[15]

Townsend, M. A. R., Townsend, J. E., & Seo, K. J. (2001, December). *Children's motivation to read following Reading Recovery*. Paper presented at the meeting of the National Reading Conference, Chicago.[13]

Wang, Y. L., & Johnstone, W. (1997, March). *Evaluation of Reading Recovery program*. Paper presented at the meeting of the American Educational Research Association, Chicago.[13]

Weeks, D. (1992). A study of the implementation of Reading Recovery in Scarborough: 1990–1991. *Masters Abstracts International, 3*(03), 1005. (UMI No. MM74064)[21]

Wilkes Pendergrass, P. V. (2004). The short-term effects of Reading Recovery on children's reading development: Process and product. *Dissertation Abstracts International*, *65*(03), 823A. (UMI No. 3127551)[13]

Wright, A. (1992). Evaluation of the first British Reading Recovery programme. *British Educational Research Journal*, *18*(4), 351–368.[13]

Yukish, J. F., & Fraas, J. W. (1997). Success of old order Amish children in a strategy-oriented program for children at-risk of failure in reading. In S. L. Swartz & A. F. Klein (Eds.), *Research in Reading Recovery* (pp. 39–51). Portsmouth, NH: Heineman.[13]

Zielinski, L. A. (1997). The long term effectiveness of Reading Recovery in a small, rural school district. *Dissertation Abstracts International*, *59*(01), 0077A. (UMI No. 9822883)[13]

For more information about specific studies and WWC calculations, please see the WWC *Reading Recovery*® Technical Appendices.

12. 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.
13. Does not use a strong causal design: this study was a quasi-experimental design but did not use achievement pretests to establish that the comparison group was equivalent to the intervention group at baseline.
14. Incomparable groups: this study was a quasi-experimental design that used achievement pretests but it did not establish that the comparison group was comparable to the treatment group prior to the start of the intervention.
15. Does not use a strong causal design: this study did not use a comparison group.
16. Does not use a strong causal design: in this study, which used a quasi-experimental design, data were presented only for students who successfully completed the intervention, which does not provide a direct test of the intervention as a whole.
17. Does not use a strong causal design: this study, which used a quasi-experimental design, had a confounding factor. The *Reading Recovery* intervention was used without proper *Reading Recovery* materials, and the instructors had not been fully trained. This makes it difficult to attribute study outcomes to *Reading Recovery.*
18. The sample is not appropriate to this review: this study did not focus on students learning to read in English, one of the parameters for this WWC review.
19. Does not use a strong causal design: this is a qualitative study.
20. Complete data are not reported: the WWC could not evaluate the design because complete data were not reported. Attempts to contact the authors for more information were unsuccessful.
21. High overall attrition: this study reported an attrition rate of greater than 20 percentage points.
22. The sample is not appropriate to this review: the parameters for this WWC review specified student outcome measures, but this study did not focus on students.
23. Disruption: this study, which used a quasi-experimental design, exhibited disruption problems that made it difficult to attribute study outcomes to the intervention, as delivered.
24. Does not use a strong causal design: this study, which used a quasi-experimental design, tested only a portion of the *Reading Recovery* curriculum, making it difficult to attribute study outcomes to *Reading Recovery.*

References *(continued)*

25. Does not use a strong causal design: for the portion of the sample of interest for this WWC review, there was one of two issues. Either there was a confound, with the *Reading Recovery* intervention being modified or combined with other interventions, making it difficult to attribute study outcomes to the intervention. Or the study did not establish that the comparison group was comparable to the treatment group prior to the start of the intervention.
26. The outcome measures are not relevant to this review: the outcomes in this study did not address one of the domains of interest in this review.

Appendix

Appendix A1.1 Study characteristics: Baenen, Bernhole, Dulaney, & Banks, 1997 (randomized controlled trial)

Characteristic	Description
Study citation	Baenen, N., Bernhole, A., Dulaney, C., & Banks, K. (1997). Reading Recovery: Long-term progress after three cohorts. *Journal of Education for Students Placed at Risk, 2*(2), 161.
Participants	The study involved 772 first-grade students who were studied in four cohorts between 1990 and 1994. Of this total number, only one cohort (who received treatment in 1990–91 at 10 schools[1]) with 168 students met WWC criteria for inclusion. After 1990–91, the comparison group was made up of students who were not comparable to the intervention group in terms of their achievement level—they were not the lowest achieving students in participating schools, as the *Reading Recovery®* students were. Due to attrition, the author's final analysis sample included 147 first-grade students in the 1990–91 cohort.[2] All 147 students were followed longitudinally into second grade; 127 of these students were followed and included in the third-grade analysis. No information was reported regarding ethnicity or gender, but all students were low-achieving and a high percentage of students qualified for free or reduced-price lunch (Baenen, personal communication, January 9, 2006).
Setting	The study took place in 10 elementary schools in the Wake County public schools in North Carolina.
Intervention	The intervention group was originally composed of 84 students who qualified for *Reading Recovery®* on the basis of three subtests from the Observation Survey of Early Literacy Achievement—Text Reading Level (running record), Dictation, and Writing Vocabulary. Intervention students, among the lowest achieving students at their schools, were randomly assigned to receive the *Reading Recovery®* intervention. They were taught by one of 12 teachers. Results were presented for 72 students for first-and second-grade analyses. Of these, 27% of the students received a "partial program" instead of the full set of more than 60 lessons in *Reading Recovery®*.
Comparison	The comparison group was composed of 84 students who qualified for *Reading Recovery®* on the basis of three subtests from the Observation Survey of Early Literacy Achievement—Text Reading Level (running record), Dictation, and Writing Vocabulary. These students, among the lowest achieving students at their schools, were randomly assigned to a waitlist for *Reading Recovery®* intervention. They did not receive *Reading Recovery®* during the time of the study but received the regular services available to them. Results were presented for 75 students for first- and second-grade analyses and 68 students for third-grade analysis.
Primary outcomes and measurement	Grade retention was measured at the end of first and second grade, and the North Carolina End-of-Grade test in reading was used at the end of third grade. The authors measured referral to special education and Title I services, but the outcomes are not included in this report because these types of outcomes are not specified by the WWC Beginning Reading Protocol. The authors used three subtests of the Observation Survey of Early Literacy Achievement, but these measures are not included in the review because collection of these data did not focus on the full sample of students. The authors also used a measure to gauge teacher perception of student achievement, but it is not included in this report because the WWC focuses on direct student achievement measures only. (See Appendix A2.4 for more detailed descriptions of outcome measures.)
Teacher training	*Reading Recovery®* teachers participated in a graduate-level course that included 33 training sessions from September to June of the intervention year. The 1990–91 school year was the first year of implementing *Reading Recovery®* for the district; therefore it was the first year the teachers taught the intervention.

1. One of the manuscripts related to this study reported 11 schools.
2. Pretest comparability of the students was demonstrated for 146 students. It is reasonable to assume that the pretest mean for the 146 students would be similar to that based on 147 students.

Appendix A1.2 Study characteristics: Pinnell, DeFord, & Lyons, 1988 (randomized controlled trial)

Characteristic	Description
Study citation	Pinnell, G. S., DeFord, D. E., & Lyons, C. A. (1988). *Reading Recovery: Early intervention for at-risk first graders* (Educational Research Service Monograph). Arlington, VA: Educational Research Service.
Participants	The study involved 187 first-grade students from 14 schools. While information about the specific schools included in the study was not presented, the district has a racial composition of 45% nonwhite students and a gender composition of 51% male students. The city had a mobility rate of approximately 20%, with 66% of students receiving free or reduced-price lunch. Students in the study were low achieving, as defined by scoring in the lowest 20% of their class on reading measures. Due to attrition, results were reported for 184 students.
Setting	The study took place in 14 urban public schools in Columbus, Ohio.
Intervention	There were two intervention groups with a total of 134 students taught by 32 teachers. One group was randomly assigned to receive the standard *Reading Recovery®* pull-out program (n = 38). These students had regular classroom teachers who were not trained in *Reading Recovery®*. This group experienced attrition of one student. A second group of students determined to be eligible for *Reading Recovery®* received the standard *Reading Recovery®* pull-out program, with the addition of having regular classroom teachers trained in *Reading Recovery®* (n = 96). The second group was not randomly assigned to *Reading Recovery®* nor to their classroom teacher, so this portion of the study is considered a quasi-experimental design. It is not included in the intervention rating because the second intervention group with a *Reading Recovery®*–trained teacher as regular classroom teacher goes beyond the standard implementation of the program.[1]
Comparison	The comparison group was composed of low-achieving students who did not have regular classroom teachers trained in *Reading Recovery®* and who were randomly assigned to an alternative compensatory program (n = 53) (Pinnell, personal communication, September 9, 2006). Students in this group received from a trained paraprofessional a series of skill-oriented drill activities conducted in small groups or individual sessions of 30–45 minutes. The comparison group was similar to both intervention groups on pretest measures (Pinnell, personal communication, December 5, 2005). The final analysis sample included 51 students.
Primary outcomes and measurement	Five subtests of the Observation Survey of Early Literacy Achievement were included—Letter Identification, Word Recognition, Concepts about Print, Writing Vocabulary, and Dictation. Additional measures included a writing assessment, the Reading Vocabulary subtest of the Comprehensive Test of Basic Skills (CTBS), and the Reading Comprehension subtest of the CTBS. Results from the Observation Survey: Text Reading Level subtest were not reported because effect sizes could not be calculated that were comparable to other measures.[2] (See Appendices A2.1–A2.4 for more detailed descriptions of outcome measures.)
Teacher training	Teachers received a full year of special training during which they practiced teaching and observed other teachers through a one-way mirror. Twelve teachers received training from a university program and were in their second year of teaching the intervention during the time of the study. The other 20 teachers received training from a local teacher leader and were in their first year of teaching the intervention during the time of the study.

1. Results are reported in Appendices A4.1–4.3.
2. Findings based on the Observation Survey: Text Reading Level subtest are not included in the effectiveness ratings because effect sizes and the statistical significance of the findings could not be calculated given the information provided in the study. The Observation Survey: Text Reading Level subtest is reported as reading levels based on ordinal, rather than equal-interval, scales. For example, the increase in fluency measured by scoring at level 3 compared with level 2 on the scale may not be equal to the increase in fluency as measured by scoring at level 24 compared with level 23. The author no longer had information on the number of students scoring at each level. For more detail, see Denton, C. A., Ciancio, D. J., & Fletcher, J. M. (2006). Validity, reliability, and utility of the Observation Survey of Early Literacy Achievement. *Reading Research Quarterly, 41*(1), 8–34.

Appendix A1.3 Study characteristics: Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994 (randomized controlled trial)

Characteristic	Description
Study citation	Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly, 29*(1), 8–39.
Participants	The study was designed to examine outcomes of 403 first-grade students distributed across 43 schools. Percentages of children receiving Aid to Dependent Children ranged between 9% and 42% among districts. For this report, the WWC looked at results for students in 10 districts at schools that were using *Reading Recovery®*. In those schools, eligible students were randomly assigned to the *Reading Recovery®* group or the comparison group, which did not receive any special instruction. The WWC did not look at the portions of the study that examined results for students who received three additional reading interventions—Reading Success, Direct Instruction Skills Plan, and Reading and Writing Group. These were delivered at schools other than the *Reading Recovery®* schools and were compared with comparison students in those schools. In the original study design, 100 students were randomly assigned to receive *Reading Recovery®* or the comparison condition at 10 schools. However, random assignment was not successfully implemented at two schools and there was minor attrition at the remaining schools, resulting in a final analytic sample of 79 students from eight schools. Treatment and comparison students attended the same schools. All students were low achieving.
Setting	The study took place in 10 school districts (two rural, two suburban, and six urban) in Ohio.
Intervention	The intervention group was composed of 31 low-achieving students from one school in each of the districts that already had a *Reading Recovery®* program in place. Intervention students participated in the *Reading Recovery®* curriculum.
Comparison	The comparison group included 48 students who were randomly assigned to receive no special instruction, but continued to participate in their regular reading program. These students attended the same schools as the intervention students.
Primary outcomes and measurement	The Dictation subtest of the Observation Survey of Early Literacy Achievement was administered in addition to the Woodcock Reading Mastery Test–Revised and the Gates-MacGinitie Reading Test. Results from the Observation Survey: Text Reading Level subtest were not reported because effect sizes that were comparable to other measures could not be calculated.[1] (See Appendices A2.2 and A2.4 for more detailed descriptions of outcome measures.)
Teacher training	At least two years prior to the study, *Reading Recovery®* teachers received specialized training—weekly 2.5 hour sessions for one year—during which they practiced teaching and observed other teachers through a one-way mirror.

1. Findings based on the Observation Survey: Text Reading Level subtest are not included in the effectiveness ratings because effect sizes and the statistical significance of the findings could not be calculated given the information provided in the study. The Observation Survey: Text Reading Level subtest is reported as reading levels based on ordinal, rather than equal-interval, scales. For example, the increase in fluency measured by scoring at level 3 compared with level 2 on the scale may not be equal to the increase in fluency as measured by scoring at level 24 compared with level 23. The author no longer had information on the number of students scoring at each level. For more detail, see Denton, C. A., Ciancio, D. J., & Fletcher, J. M. (2006). Validity, reliability, and utility of the Observation Survey of Early Literacy Achievement. *Reading Research Quarterly, 41*(1), 8–34.

Appendix A1.4 Study characteristics: Schwartz, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology, 97*(2), 257–267.
Participants	The study was designed to examine the outcomes of first-grade students distributed across 47 *Reading Recovery®* teachers. Each teacher identified for random assignment two students eligible for *Reading Recovery®* based on their low scores on the Observation Survey of Early Literacy Achievement. These 94 students were randomly assigned to enter the *Reading Recovery®* program during either the first or second half of the school year. The study included two additional comparison groups of 47 low-average and 47 high-average readers from the same classrooms as the *Reading Recovery®* students who were not expected to participate in the *Reading Recovery®* program. Analysis involving these comparison groups was not eligible for WWC review because the WWC considers only comparisons of students with similar achievement backgrounds in assessing the effectiveness of *Reading Recovery®*. Because of missing testing data, the author's final analysis of the randomly assigned groups included 74 students distributed across 37 teachers.
Setting	The study took place in elementary schools in 14 states.
Intervention	The intervention group was composed of the lowest 20%–30% of students who qualified for *Reading Recovery®* based on the six subtests of the Observation Survey and on judgments by teachers. All teachers agreed to dedicate one of their four *Reading Recovery®* slots to the study, allowing random assignment of one of two eligible students to receive intervention delivery in the first half of the year. Thus students in the intervention group were randomly assigned to receive *Reading Recovery®* during the first half of the year. Originally, participants were taught by 47 *Reading Recovery®* teachers who had volunteered to be part of the study, but because of missing information, data from only 37 teachers and 37 students were included in the author's final analysis. The intervention group was 61% male, 38% white, 47% black, 12% Hispanic, and 3% Asian. About 60% of the group received free or reduced-price lunch.
Comparison	The comparison group of interest to the WWC was composed of the lowest 20%–30% of students who qualified for *Reading Recovery®* based on six subtests of the Observation Survey and on judgments by teachers. Participants in the comparison group were randomly assigned to receive *Reading Recovery®* during the second half of the year. Thus these participants served as comparisons only during the first part of the year when they received instruction in their regular classroom only. Data from 37 teachers and 37 students were included in the author's final analysis. The comparison group was 41% male, 47% white, 38% black, and 15% Hispanic. Approximately 57% of the group received free or reduced-price lunch.
Primary outcomes and measurement	Six subtests of the Observation Survey of Early Literacy Achievement were administered—Letter Identification, Word Recognition, Concepts about Print, Writing Vocabulary, Dictation, and Text Reading Level.[1] Other assessments included the Yopp-Singer Phoneme Segmentation Task, the deletion task, the Slosson Oral Reading Test–Revised, and the Degrees of Reading Power Test. (See Appendices A2.1–A2.4 for more detailed descriptions of outcome measures.)
Teacher training	No information on training for the specific teachers participating in this study was provided, but *Reading Recovery®* teachers typically undergo graduate-level training and must implement the program under the supervision of a teacher leader.

1. For many studies, findings based on the Observation Survey: Text Reading Level subtest were not included in the effectiveness ratings because effect sizes and the statistical significance of the findings could not be calculated given the information provided in the study. In this case, however, the author was able to provide the number of students scoring at each level (Schwartz, personal communication, December 5, 2006). The outcome was calculated as a logged-odds ratio, with the number of students reading at or above the first-grade level compared with students reading below the first-grade level.

Appendix A1.5 Study characteristics: Iverson & Tunmer, 1993 (quasi-experimental design)

Characteristic	Description
Study citation	Iverson, S., & Tunmer, W. E. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology, 85*(1), 112–126.
Participants	The study included 160 first-grade students distributed across 47 classrooms in 30 schools. Ninety-six students who performed at the lowest levels on tests administered at the beginning of the first grade were placed into one of three matched groups: a standard *Reading Recovery®* group, a modified *Reading Recovery®* group, and a comparison group that did not participate in *Reading Recovery®*. Students across the three groups were matched based on pretest scores from the Letter Identification and Dictation subtests of the Observation Survey. The study also compared the *Reading Recovery®* groups with comparison groups consisting of 64 children from the same classrooms as the *Reading Recovery®* participants who were deemed by the classroom teachers to be performing at average reading level. But analyses involving these comparison groups were not eligible for this WWC review because the WWC considers only comparisons of students with similar achievement backgrounds to provide information on the effectiveness of *Reading Recovery®*.
Setting	Thirty schools from 13 school districts in Rhode Island.
Intervention	The study authors used two intervention groups drawn from the 23 schools operating *Reading Recovery®* programs. Sixty-four low-scoring students from 34 classrooms were assigned to receive standard or modified *Reading Recovery®*. The 32 students in the standard *Reading Recovery®* group received one-on-one lessons that followed the procedures described by the program developer, which included a letter identification segment as needed. Once a child mastered letter identification, time allocated to this segment could be used for incidental word analysis in other segments. For the 32 students in the modified *Reading Recovery®* program, explicit instruction in letter-phoneme patterns replaced the letter identification segment beginning in the fourth week of the program. The WWC did not include the second comparison group in its intervention ratings because it was a modified version of the standard program.[1]
Comparison	The matched comparison group of interest to the WWC was drawn from the seven schools that did not operate *Reading Recovery®* programs. Thirty-two low-scoring students from 13 classrooms were matched to the intervention groups based on pretest scores. Students in the comparison group received the support services normally available to at-risk readers, which consisted of small-group instruction provided at least four days a week through either the federally funded Chapter One program or the state-supported literacy program.[1]
Primary outcomes and measurement	Five subtests of the Observation Survey of Early Literacy Achievement were included—Letter Identification, Word Recognition, Concepts about Print, Writing Vocabulary, and Dictation. Additional measures included the Dolch Word Recognition Test, Yopp-Singer Phoneme Segmentation Test, a phoneme deletion task, and a pseudoword decoding task. Results from the Observation Survey: Text Reading Level subtest were not reported because effect sizes that were comparable to other measures could not be calculated.[2] (See Appendices A2.1, A2.2, and A2.4 for more detailed descriptions of outcome measures.)
Teacher training	*Reading Recovery®* teachers in the study were reading specialists with master's degrees. To learn how to administer the assessments, teachers participated in four two-hour training sessions. To learn how to deliver the intervention, teachers attended weekly two-hour classes. During the session, teachers were able to practice teaching and observe other teachers through a one-way mirror. They also received feedback from the teacher leader during individual visits to the classes.

1. Results are reported in Appendices A4.1 and A4.3.
2. Findings based on the Observation Survey: Text Reading Level subtest are not included in the effectiveness ratings because effect sizes and the statistical significance of the findings could not be calculated given the information provided in the study. The Observation Survey: Text Reading Level subtest is reported as reading levels based on ordinal, rather than equal-interval, scales. For example, the increase in fluency measured by scoring at level 3 compared with level 2 on the scale may not be equal to the increase in fluency as measured by scoring at level 24 compared with level 23. The author no longer had information on the number of students scoring at each level. For more detail, see Denton, C. A., Ciancio, D. J., & Fletcher, J. M. (2006). Validity, reliability, and utility of the Observation Survey of Early Literacy Achievement. *Reading Research Quarterly, 41*(1), 8–34.

Appendix A2.1 Outcome measures in the alphabetics domain by construct

Outcome measure	Description
Phonemic awareness	
Deletion task	A 10-item version of the Rosner task that requires students to repeat a word and then say it again after omitting a given syllable or sound. The assessment is not standardized (as cited in Schwartz, 2005).
Phoneme deletion task	Developed by Calfee, this measure requires students to delete the first consonant segment in a word and read the word aloud. It includes four training lists and six transfer lists that increase in difficulty as students move through each list. The most difficult lists consist of both real words and pseudowords. Student scores are based on the number of correct words read from the transfer lists (as cited in Iverson and Tunmer, 1993).
Yopp-Singer Phoneme Segmentation Test	Developed by Yopp, the test is an orally administered assessment. A teacher works with each student individually and introduces the test as a word game. The teacher has a list of 22 words that the student is not allowed to see. After the teacher reads each word, the student must repeat all of the sounds in the word separately and slowly (as cited in Iverson and Tunmer, 1993; Schwartz, 2005).
Print awareness	
Observation Survey of Early Literacy Achievement: Concepts about Print subtest[1]	Students perform tasks related to printed language concepts (for example, directionality and word concepts) while reading a book. This assessment, developed by Clay, is not standardized (as cited in Iverson and Tunmer, 1993; Pinnell, DeFord, and Lyons, 1988; Schwartz, 2005).
Letter knowledge	
Observation Survey of Early Literacy Achievement: Letter Identification subtest[1]	Students identify upper- and lowercase letters. This assessment, developed by Clay, is not standardized (as cited in Iverson and Tunmer, 1993; Pinnell, DeFord, and Lyons, 1988; Schwartz, 2005).
Phonics	
Dolch Word Recognition Test	A standardized, individually administered word recognition test that contains a list of 220 short, frequently occurring words. Students are asked to read aloud from the list of words to measure their automatic word recognition and word retrieval skills, not their ability to sound out words (as cited in Iverson and Tunmer, 1993).
Observation Survey of Early Literacy Achievement: Word Recognition subtest (also known as Ready to Read or Ohio Word Test)[1]	Students read 20 common sight words from basic reading texts, and their accuracy is scored. This assessment, developed by Clay, is not standardized (as cited in Iverson and Tunmer, 1993; Pinnell, DeFord, and Lyons, 1988; Schwartz, 2005).
Pseudoword decoding task	A 40-item test that requires students to name single-syllable synthetic words that conform to the rule of English orthography. The assessment is not standardized (as cited in Iverson and Tunmer, 1993).

1. The Clay Observation Survey was developed by Dr. Marie Clay, who also developed *Reading Recovery®*. Although there is no evidence of obvious overalignment between the measure and the intervention (intervention student receiving exposure to the measure during the course of treatment), it should be noted that the developer of the intervention and the measure were the same.

Appendix A2.2 Outcome measures in the fluency domain

Outcome measure	Description
Observation Survey of Early Literacy Achievement: Text Reading Level subtest[1,2]	The percentage of students scoring at the first-grade reading level or higher compared with those scoring lower than first grade. To determine this, students read from passages of increasing difficulty, and student's error rate and self-correcting behavior were recorded using the running record technique. Students read from leveled texts drawn from a basal reading series until their accuracy rate fell below 90%. Results were translated to a numerical reading level from level 1 to level 30, which in turn match up to grade-level equivalency. This assessment method is not standardized (as cited in Schwartz, 2005).
Slosson Oral Reading Test–Revised (SORT–R3)	Developed by Nicholson, the measure consists of 200 words arranged in order of difficulty, with 20 words per list. Each list represents an approximate reading grade level (for example, list one is equivalent to first grade). Administration ends after all the words on one list are missed. The measure is standardized and norm-referenced (as cited in Schwartz, 2005).

1. The Clay Observation Survey was developed by Dr. Marie Clay, who also developed *Reading Recovery®*. Although there is no evidence of obvious overalignment between the measure and the intervention (intervention student receiving exposure to the measure during the course of treatment), it should be noted that the developer of the intervention and the measure were the same.
2. Findings based on the Observation Survey: Text Reading Level subtest are not included in the effectiveness ratings for most studies because effect sizes and the statistical significance of the findings could not be calculated given the information provided in the studies. The author of one study (Schwartz, 2005) was able to provide the percentage of students scoring at each level. The outcome was changed to a logged-odds ratio for students scoring at the first-grade level or higher compared with those scoring lower than first grade.

Appendix A2.3 Outcome measures in the comprehension domain by construct

Outcome measure	Description
Reading comprehension	
Comprehensive Test of Basic Skills (CTBS): Reading Comprehension subtest	A group-administered, standardized assessment of reading comprehension (as cited in Pinnell, DeFord, and Lyons, 1988).
Degrees of Reading Power Test	An untimed standardized test requiring students to read a nonfiction passage with a word or set of words missing. Students select an appropriate answer to complete the sentence from a set of four or five alternatives (as cited in Schwartz, 2005).
Vocabulary development	
CTBS: Reading Vocabulary subtest	A group-administered, standardized assessment of vocabulary (as cited in Pinnell, DeFord, and Lyons, 1988).

Appendix A2.4 Outcome measures in the general reading achievement domain

Outcome measure	Description
Gates-MacGinitie (1978)	A standardized test that covers vocabulary and comprehension aspects of reading. It evaluates students' abilities to decode initial consonants, consonant clusters, final consonants, and vowels in real English words and also measures their ability to recognize commonly used words without decoding. For reading comprehension, answer choices are given as pictures and words (as cited in Pinnell et al., 1994).
North Carolina End-of-Grade reading test	A standardized state assessment designed to match the North Carolina curriculum. It uses multiple-choice questions with reading passages and is designed to measure reading and thinking skills (as cited in Baenen et al., 1997).
Observation Survey of Early Literacy Achievement: Hearing and Recording Sounds in Words (Dictation) subtest[1]	Students write the words that are dictated to them in sentence form. This assessment, developed by Clay, is not standardized (as cited in Iverson and Tunmer, 1993; Pinnell, DeFord, and Lyons, 1988; Pinnell et al., 1994; Schwartz, 2005).
Observation Survey of Early Literacy Achievement: Writing Vocabulary subtest[1]	Students are given 10 minutes to write as many words as they can on a blank sheet of paper. If needed, a standard set of prompts is used to encourage additional attempts to write. The measure is scored by counting the number of correctly spelled words (as cited in Iverson and Tunmer, 1993; Pinnell, DeFord, and Lyons, 1988; Schwartz, 2005).
Retention	The percentage of students who were retained in their current grade (as cited in Baenen et al., 1997).
Woodcock Reading Mastery Test–Revised	A standardized test composed of six subtests. The subtests measure ability to form associations between visual stimuli and oral responses; ability to recognize upper- and lowercase letters in a variety of fonts; ability to read words aloud; ability to read aloud nonsense words or uncommon words to test phonic and structural analysis skills for pronouncing unfamiliar words; vocabulary ability through the use of antonyms, synonyms, and analogies; and passage comprehension by filling in missing words in a short paragraph (as cited in Pinnell et al., 1994).

1. The Clay Observation Survey was developed by Dr. Marie Clay, who also developed *Reading Recovery®*. Although there is no evidence of obvious overalignment between the measure and the intervention (intervention student receiving exposure to the measure during the course of treatment), it should be noted that the developer of the intervention and the measure were the same.

Appendix A3.1 Summary of findings for all domains[1]

	Domain							
	Alphabetics					Comprehension		
	Phonemic awareness	*Print awareness*	*Letter knowledge*	*Phonics*	**Fluency**	*Reading comprehension*	*Vocabulary development*	**General reading achievement**
Met evidence standards								
Baenen et al., 1997	nr	nr	nr	nr	nr	nr	nr	ind
Pinnell, DeFord, & Lyons, 1988	nr	+	ind	(+)	nr	+	+	+
Pinnell et al., 1994	nr	nr	nr	nr	nr	nr	nr	+
Schwartz, 2005	(+)	+	(+)	+	+	ind	nr	+
Met evidence standards with reservations								
Iverson & Tunmer, 1993	+	+	+	+	nr	nr	nr	+
Rating of effectiveness	Positive				Potentially positive	Potentially positive		Positive

nr = no reported outcomes under this domain or construct

+ = study average finding was positive and statistically significant

(+) = study average finding was positive and substantively important, but not statistically significant

ind = study average finding was indeterminate, that is, neither substantively important nor statistically significant

1. This appendix reports summary findings of study averages that were considered for the effectiveness rating and the improvement index in each domain. More detailed information on findings for the all measures within the domains and the constructs that factor into the domains are reported in Appendices A3.2–A3.5.

Appendix A3.2 Summary of study findings included in the alphabetics domain by construct[1]

Outcome measure	Study sample	Sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Phonemic awareness construct								
Schwartz, 2005 (randomized controlled trial)[7]								
Deletion task[8]	Grade 1	74 students	6.64 (2.56)	5.58 (2.50)	1.06	0.41	ns	+16
Yopp-Singer Phoneme Segmentation Test[8]	Grade 1	74 students	17.70 (4.93)	15.27 (5.43)	2.43	0.46	ns	+18
Iverson & Tunmer, 1993 (quasi-experimental design)—standard *RR* group vs. comparison group[7]								
Phoneme deletion task[8]	Grade 1	64 students 14 schools	14.00 (6.16)	3.91 (5.31)	10.09	1.73	Statistically significant	+46
Yopp-Singer Phoneme Segmentation Test[8]	Grade 1	64 students 14 schools	17.63 (4.46)	6.21 (5.18)	11.42	2.33	Statistically significant	+49
Print awareness construct								
Pinnell, DeFord, & Lyons, 1988 (randomized controlled trial)—*RR* group without *RR* teachers vs. comparison group[7]								
Observation Survey: Concepts about Print subtest[9]	Grade 1	88 students	16.80 (2.91)	13.98 (3.31)	2.82	0.89	Statistically significant	+31
Schwartz, 2005 (randomized controlled trial)[7]								
Observation Survey: Concepts about Print subtest[9, 10]	Grade 1	74 students	19.24 (2.55)	16.68 (2.30)	2.56	1.04	Statistically significant	+35
Iverson & Tunmer, 1993 (quasi-experimental design)—standard *RR* group vs. comparison group[7]								
Observation Survey: Concepts about Print subtest[9]	Grade 1	64 students 14 schools	18.12 (1.74)	13.72 (3.49)	4.40	1.58	Statistically significant	+44
Letter knowledge construct								
Pinnell, DeFord, & Lyons, 1988 (randomized controlled trial)—*RR* group without *RR* teachers vs. comparison group[7]								
Observation Survey: Letter Identification subtest[9]	Grade 1	88 students	48.05 (1.41)	49.61 (8.33)	–1.56	–0.24	ns	–10

(continued)

Appendix A3.2 Summary of study findings included in the alphabetics domain by construct *(continued)*

Outcome measure	Study sample	Sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Schwartz, 2005 (randomized controlled trial)[7]								
Observation Survey: Letter Identification subtest[9,10]	Grade 1	74 students	52.28 (1.27)	51.68 (2.78)	0.60	0.23	ns	+9
Iverson & Tunmer, 1993 (quasi-experimental design)—standard *RR* group vs. comparison group[7]								
Observation Survey: Letter Identification subtest[9]	Grade 1	64 students 14 schools	53.53 (1.01)	49.00 (6.63)	4.53	0.94	Statistically significant	+33
Phonics construct								
Pinnell, DeFord, & Lyons, 1988 (randomized controlled trial portion)—*RR* group without *RR* teachers vs. comparison group								
Observation Survey: Word Recognition subtest[9]	Grade 1	88 students	13.29 (1.63)	11.98 (3.92)	1.31	0.41	ns	+16
Schwartz, 2005 (randomized controlled trial)[7]								
Observation Survey: Word Recognition subtest[9,10]	Grade 1	74 students	14.96 (3.99)	8.87 (4.75)	6.09	1.37	Statistically significant	+41
Iverson & Tunmer, 1993 (quasi-experimental design)—standard *RR* group vs. comparison group[7]								
Dolch Word Recognition Test[9]	Grade 1	64 students 14 schools	94.87 (29.07)	20.22 (13.93)	74.65	3.24	Statistically significant	+50
Observation Survey: Word Recognition subtest[9]	Grade 1	64 students 14 schools	12.66 (1.85)	5.56 (3.83)	7.10	2.33	Statistically significant	+49
Pseudoword decoding task[8]	Grade 1	64 students 14 schools	9.00 (7.07)	1.41 (2.31)	7.59	1.43	Statistically significant	+42
Averages								
Average[11] for alphabetics (Pinnell, DeFord, & Lyons, 1988)						0.35	ns	+14
Average[11] for alphabetics (Schwartz, 2005)						0.70	Statistically significant	+26
Average[11] for alphabetics (Iverson & Tunmer, 1993)						1.94	Statistically significant	+47

(continued)

Appendix A3.2 Summary of study findings included in the alphabetics domain by construct *(continued)*

Outcome measure	Study sample	Sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Domain average[11] for alphabetics across all studies						1.00	na	+34

ns = not statistically significant
na = not applicable

1. This appendix presents findings for four constructs that fall within the alphabetics domain. All alphabetics measures were averaged together for rating purposes (see Appendix A3.1). Additional findings from the same studies are not included in these ratings, but are reported in Appendix A4.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see 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 and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools 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 Schwartz (2005), a correction for multiple comparisons was applied by the author, so the WWC did not apply a correction. In the case of Pinnell, DeFord, and Lyons (1988), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. In the case of Iverson and Tunmer (1993), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
8. Means presented for these measures are the posttest means for each group (pretest data were not available in the study report).
9. The *RR* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
10. The pretest data were based on 36 of the 37 pairs of students in this group. It is reasonable to assume that the pretest mean for the 36 pairs of students would be similar to that based on 37 pairs.
11. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

Appendix A3.3 Summary of study findings included in the fluency domain[1]

Outcome measure	Study sample	Analysis sample size	Author's findings from the study: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Schwartz, 2005 (randomized controlled trial)[7]								
SORT–R3[8]	Grade 1	74 students	30.58 (14.41)	18.12 (11.87)	12.46	0.93	Statistically significant	+32
Observation Survey: Text Reading Level subtest[9]	Grade 1	74 students	0.78	0.05	0.73	2.49	Statistically significant	+49
Average[10] for fluency (Schwartz, 2005)						1.71	Statistically significant	+46

1. This appendix reports findings considered for the effectiveness rating and the improvement index. All fluency measures were averaged together for rating purposes (see Appendix A3.1).
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see 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 and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools 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 Schwartz (2005), a correction for multiple comparisons was applied by the author, so the WWC did not apply a correction.
8. Means presented for this measure are the posttest means for each group (pretest data were not available in the study report).
9. Means presented for this measure are the posttest proportions for each group scoring at or above a first-grade reading level. Effect size is computed as a Cox Index: logged-odds ratio transformation divided by 1.65. See Technical Details of WWC-Conducted Computations, specifically the computation of effect sizes for binary outcomes.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

Appendix A3.4 Summary of study findings included in the comprehension domain by construct[1]

Outcome measure	Study sample	Analysis sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Reading comprehension construct								
Pinnell, DeFord, & Lyons, 1988 (randomized controlled trial portion)—*RR* group without *RR* teachers vs. comparison group[7]								
CTBS: Reading Comprehension subtest[8]	Grade 1	82 students	36.67 (19.27)	27.33 (13.94)	9.34	0.56	Statistically significant	+21
Schwartz, 2005 (randomized controlled trial)[7]								
Degrees of Reading Power Test[8]	Grade 1	74 students	4.82 (3.88)	4.27 (3.88)	0.55	0.14	ns	+6
Vocabulary development construct								
Pinnell, DeFord, & Lyons, 1988 (randomized controlled trial)—*RR* group without *RR* teachers vs. comparison group[7]								
CTBS: Reading Vocabulary subtest[8]	Grade 1	81 students	36.64 (11.93)	28.07	(17.00) 8.57	0.57	Statistically significant	+21
Averages								
Average[9] for comprehension (Pinnell, DeFord, & Lyons, 1988)						0.56	Statistically significant	+21
Average[9] for comprehension (Schwartz, 2005)						0.14	ns	+6
Domain average[9] for comprehension across all studies						0.35	na	+14

ns = not statistically significant
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index. All comprehension measures were averaged together for rating purposes (see Appendix A3.1). Additional findings from the same studies are not included in these ratings, but are reported in Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see 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 and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools 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 Pinnell, DeFord, and Lyons (1988),

(continued)

a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. In the case of Schwartz (2005), a correction for multiple comparisons was applied by the author, so the WWC did not apply a correction.

8. Means presented for this measure are the posttest means for each group (pretest data were not available in the study report).
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

Appendix A3.5 Summary of study findings included in the general reading achievement domain[1]

Outcome measure	Study sample	Analysis sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Baenen et al., 1997 (randomized controlled trial)[7]								
Outcome time 1—end of first grade								
Retention[8]	Grade 1	147 students	0.06	0.05	0.01	–0.12	ns	–5
Pinnell, DeFord, & Lyons, 1988 (randomized controlled trial)—*RR* group without *RR* teachers vs. comparison group[7]								
Observation Survey: Dictation subtest[9]	Grade 1	88 students	30.52 (6.13)	23.80 (7.99)	6.72	0.92	Statistically significant	+32
Observation Survey: Writing Vocabulary subtest[9]	Grade 1	88 students	33.21 (13.49)	25.37 (14.33)	7.84	0.56	Statistically significant	+21
Pinnell et al., 1994 (randomized controlled trial)[7]								
Gates-MacGinitie[10]	Grade 1	79 students	36.19 (13.12)	31.00 (na)	5.19 (10.16)	0.51	Statistically significant	+19
Observation Survey: Dictation subtest[10]	Grade 1	79 students	31.74 (6.18)	26.75 (na)	4.99 (7.67)	0.65	Statistically significant	+24
Woodcock Reading Mastery Test–Revised[10]	Grade 1	79 students	39.81 (na)	39.49 (na)	0.32 (0.70)	0.49	Statistically significant	+19
Schwartz, 2005 (randomized controlled trial)[7]								
Observation Survey: Dictation subtest[9]	Grade 1	74 students	35.58 (2.70)	29.08 (7.37)	6.50	1.16	Statistically significant	+38
Observation Survey: Writing Vocabulary subtest[9]	Grade 1	74 students	42.67 (11.42)	31.00 (12.94)	11.67	0.95	Statistically significant	+33
Iverson & Tunmer, 1993 (quasi-experimental design)—standard *RR* group vs. comparison group[7]								
Observation Survey: Dictation subtest[9]	Grade 1	64 students 14 schools	34.63 (2.53)	23.34 (8.68)	11.29	1.74	Statistically significant	+46
Observation Survey: Writing Vocabulary subtest[9]	Grade 1	64 students 14 schools	39.66 (6.42)	15.69 (9.15)	23.97	3.00	Statistically significant	+50

(continued)

Appendix A3.5 Summary of study findings included in the general reading achievement domain *(continued)*

Outcome measure	Study sample	Analysis sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Averages								
Average[11] for general reading achievement (Baenen et al., 1997)						–0.12	ns	–5
Average[11] for general reading achievement (Pinnell, DeFord, & Lyons, 1988)						0.74	Statistically significant	+27
Average[11] for general reading achievement (Pinnell et al., 1994)						0.55	Statistically significant	+21
Average[11] for general reading achievement (Schwartz, 2005)						1.05	Statistically significant	+35
Average[11] for general reading achievement (Iverson & Tunmer, 1993)						2.37	Statistically significant	+49
Domain average[14] for general reading achievement across all studies						0.92	na	+32

ns = not statistically significant
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index. All general reading achievement measures were averaged together for rating purposes (see Appendix A3.1). Follow-up findings from the same studies are not included in these ratings, but are reported in Appendix A4.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see 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 and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools 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 Pinnell, DeFord, and Lyons (1988), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. In the case of Pinnell et al. (1994), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. In the case of Schwartz (2005), a correction for multiple comparisons was applied by the author, so the WWC did not apply a correction. In the case of Baenen et al. (1997), no corrections were needed.
8. Means presented for this measure are proportions for each group (retention rates). Effect size is computed as a Cox Index: logged-odds ratio transformation divided by 1.65. See Technical Details of WWC-Conducted Computations, specifically the computation of effect sizes for binary outcomes. The sign of the effect size was reversed so that a positive effect size favors the intervention group.
9. The *RR* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
10. The authors stated that the exact sample varied from outcome to outcome, although they did not provide specific details. The mean difference is regression-adjusted and standard deviation is a pooled within-treatment group estimate. Study authors used HLM analysis to compute results. Effect size was calculated by dividing treatment effect estimate (regression adjusted difference in the posttest measures) by the pooled within-treatment group standard deviation.
11. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

Appendix A4.1 Summary of findings for alternative intervention groups for the alphabetics domain by construct[1]

Outcome measure	Study sample	Analysis sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group[3]	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*RR* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Phonemic awareness construct								
Iverson & Tunmer, 1993 (quasi-experimental design)—modified *RR* group vs. comparison group[8]								
Phoneme deletion task[9]	Grade 1	64 students 23 schools	11.00 (5.85)	3.91 (5.31)	7.09	1.25	Statistically significant	+40
Yopp-Singer Phoneme Segmentation Test[9]	Grade 1	64 students 23 schools	16.88 (4.53)	6.21 (5.18)	10.67	2.17	Statistically significant	+48
Print awareness construct								
Iverson & Tunmer, 1993 (quasi-experimental design)—modified *RR* group vs. comparison group[8]								
Observation Survey: Concepts about Print subtest[10]	Grade 1	64 students 23 schools	17.40 (2.40)	13.72 (3.49)	3.68	1.21	Statistically significant	+39
Pinnell, DeFord, & Lyons, 1988 (quasi-experimental design)—*RR* group with *RR* teachers vs. comparison group[8]								
Observation Survey: Concepts about Print subtest[10]	Grade 1	147 students[11] 14 schools	15.85 (2.77)	13.98 (3.31)	1.87	0.63	Statistically significant	+23
Letter knowledge construct								
Iverson & Tunmer, 1993 (quasi-experimental design)—modified *RR* group vs. comparison group[8]								
Observation Survey: Letter Identification subtest[10]	Grade 1	64 students 23 schools	52.68 (1.27)	49.00 (6.63)	3.68	0.75	Statistically significant	+27
Pinnell, DeFord, & Lyons, 1988 (quasi-experimental design)—*RR* group with *RR* teachers vs. comparison group[8]								
Observation Survey: Letter Identification subtest[10]	Grade 1	147 students[11] 14 schools	49.80 (3.83)	49.61 (8.33)	0.19	0.03	ns	+1
Phonics construct								
Iverson & Tunmer, 1993 (quasi-experimental design)—modified *RR* group vs. comparison group[8]								
Dolch Word Recognition Test[10]	Grade 1	64 students 23 schools	94.75 (23.74)	20.22 (13.93)	74.53	3.78	Statistically significant	+50

(continued)

Appendix A4.1 Summary of findings for alternative intervention groups for the alphabetics domain by construct *(continued)*

Outcome measure	Study sample	Analysis sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group[3]	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*RR* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Observation Survey: Word Recognition subtest[10]	Grade 1	64 students 23 schools	11.91 (2.15)	5.56 (3.83)	6.35	2.02	Statistically significant	+48
Pseudoword decoding task[9]	Grade 1	64 students 23 schools	8.19 (6.69)	1.41 (2.31)	6.78	1.34	Statistically significant	+41
Pinnell, DeFord, & Lyons, 1988 (quasi-experimental design)—*RR* group with *RR* teachers vs. comparison group[8]								
Observation Survey: Word Recognition subtest[10]	Grade 1	147 students[11] 14 schools	13.64 (1.85)	11.98 (3.92)	1.66	0.60	Statistically significant	+23

ns = not statistically significant

1. This appendix presents findings for alternative intervention groups on measures that fall in the alphabetics domain. Only primary group scores were used for rating purposes and are presented in Appendix A3.2
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Standard deviations were reported by study authors but cannot be compared to standard deviations on other measures because of the way the measure is created (see Appendix A2.1).
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
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 and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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 (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). 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 Pinnell, DeFord, and Lyons (1988), a correction for clustering was needed, so the significance levels may differ from those reported in the original study. The clustering correction under-adjusts clustering to the extent that the *RR* group mean and the comparison group mean were correlated within schools. In the case of Iverson and Tunmer (1993), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. Means presented for this measure are the posttest means for each group (pretest data were not available in the study report).
10. The *RR* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
11. While the authors presented data for 96 students at posttest, the pretest data were based on 94 of the 96 students in this group. It is reasonable to assume that the pretest mean for the 96 students is similar to that based on the 94 students.

Appendix A4.2 Summary of findings for alternative intervention groups for the comprehension domain by construct[1]

Outcome measure	Study sample	Analysis sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Comprehension construct								
Pinnell, DeFord, & Lyons, 1988 (quasi-experimental design)—*RR* group with *RR* teachers vs. comparison group[7]								
CTBS: Reading Comprehension subtest[8]	Grade 1	147 students[9] 14 schools	38.84 (15.31)	27.33 (13.94)	11.51	0.77	Statistically significant	+28
Vocabulary development construct								
Pinnell, DeFord, & Lyons, 1988 (quasi-experimental design)—*RR* group with *RR* teachers vs. comparison group[7]								
CTBS: Reading Vocabulary subtest[8]	Grade 1	147 students[9] 14 schools	38.16 (15.31)	28.07 (17.00)	10.09	0.63	Statistically significant	+24

1. This appendix presents findings for alternative intervention groups on measures that fall in the comprehension domain. Only primary group scores were used for rating purposes and are presented in Appendix A3.4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see 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 and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). 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 Pinnell, DeFord, and Lyons (1988), a correction for clustering was needed, so the significance levels may differ from those reported in the original study. The clustering correction under-adjusts clustering to the extent that the *RR* group mean and the comparison group mean were correlated within schools.
8. Means presented for this measure are the posttest means for each group (pretest data were not available in the study report).
9. While the authors presented data for 96 intervention students and 51 comparison students at posttest, the pretest data were based on 94 of the 96 intervention students in this group. It is reasonable to assume that the pretest mean for the 96 students is similar to that based on the 94 students.

Appendix A4.3 Summary of findings for alternative intervention groups for the general reading achievement domain[1]

Outcome measure	Study sample	Analysis sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Iverson & Tunmer, 1993 (quasi-experimental design)—modified *RR* group vs. comparison group[7]								
Observation Survey: Dictation subtest[8]	Grade 1	64 students 23 schools	34.57 (3.32)	23.34 (8.68)	11.23	1.67	Statistically significant	+45
Observation Survey: Writing Vocabulary subtest[8]	Grade 1	64 students 23 schools	40.53 (10.24)	15.69 (9.15)	24.84	2.51	Statistically significant	+49
Pinnell, DeFord, & Lyons, 1988 (quasi-experimental design)—*RR* group with *RR* teachers vs. comparison group[7]								
Observation Survey: Dictation subtest[8]	Grade 1	147 students[9] 14 schools	32.67 (5.80)	23.80 (7.99)	8.87	1.33	Statistically significant	+41
Observation Survey: Writing Vocabulary subtest[8]	Grade 1	147 students[9] 14 schools	35.96 (12.56)	25.37 (14.33)	10.59	0.80	Statistically significant	+29

1. This appendix presents findings for alternative intervention groups on measures that fall in the general reading achievement domain. Only primary group scores were used for rating purposes and are presented in Appendix A3.5.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see 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 and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). 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 Pinnell, DeFord, and Lyons (1988), a correction for clustering was needed, so the significance levels may differ from those reported in the original study. In the case of Iverson and Tunmer (1993), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
8. The *RR* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
9. While the authors presented data for 96 students at posttest, the pretest data were based on 94 of the 96 students in this group. It is reasonable to assume that the pretest mean for the 96 students is similar to that based on the 94 students.

Appendix A4.4 Summary of follow-up findings for the general reading achievement domain[1]

Outcome measure	Study sample	Analysis sample size	Author's findings from the studies: Mean outcome (standard deviation[2]) — *Reading Recovery® (RR)* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*RR* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Baenen et al., 1997 (randomized controlled trial)[7]								
Outcome time 2—end of second grade								
Retention[8]	Grade 1	147 students	0.04	0.04	0	0	ns	0
Outcome time 3—end of third grade								
North Carolina End-of-Grade reading test[9]	Grade 1	127 students	135.50 (8.40)	136.20 (9.10)	–0.70	–0.08	ns	–3

1. This appendix presents follow-up findings for measures that fall in the general reading achievement domain. Only first-grade scores were used for rating purposes and are presented in Appendix A3.5.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see 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 and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). 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 the Baenen et al. (1997), no correction for clustering was needed.
8. Means presented for this measure are the posttest data and proportions (retention rates). Pretest data is not applicable for the measure. Effect size is computed as a logged-odds ratio transformation divided by 1.65. See Technical Details of WWC-Conducted Computations, specifically the computation of effect sizes for binary outcomes.
9. Means presented for this measure are the posttest means for each group (pretest data were not available in the study report).

Appendix A5.1 *Reading Recovery*® 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.[1]

For the outcome domain of alphabetics, the WWC rated *Reading Recovery*® as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because the intervention was assigned the highest applicable rating.

Rating received
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.
Met. *Reading Recovery*® had three studies showing statistically significant positive effects, two of which met WWC evidence standards for a strong design.
• 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 A5.2 *Reading Recovery*® 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 *Reading Recovery*® as having potentially positive effects. It did not meet the criteria for positive effects because it had only one study that showed statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because the intervention 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. *Reading Recovery*® had one study showing statistically significant positive effects.

- 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 *Reading Recovery*® studies showed statistically significant or substantively important 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. *Reading Recovery*® had only one study showing statistically significant positive effects that met WWC evidence standards for a strong design.

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

  Met. No *Reading Recovery*® 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 A5.3 *Reading Recovery®* 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 *Reading Recovery®* as having potentially positive effects. It did not meet the criteria for positive effects because it had only one study that showed statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because the intervention 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. *Reading Recovery®* had one study showing statistically significant positive effects and one study showing indeterminate effects.

- 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 *Reading Recovery®* studies showed statistically significant or substantively important negative effects, and the number of studies showing indeterminate effects was equal to the number of studies showing statistically significant positive 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. *Reading Recovery®* had only one study showing statistically significant positive effects that met WWC evidence standards for a strong design.

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

  Met. No *Reading Recovery®* 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 A5.4 *Reading Recovery*® rating for the general reading achievement domain

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

For the outcome domain of general reading achievement, the WWC rated *Reading Recovery*® as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because the intervention was assigned the highest applicable rating.

Rating received
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.
Met. *Reading Recovery*® had four studies showing statistically significant positive effects, three of which met WWC evidence standards for a strong design.
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. No *Reading Recovery*® 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.