

Reading Recovery®

Program Description¹

Reading Recovery® is a short-term intervention that provides one-on-one tutoring to first-grade students who are struggling in reading and writing. The supplementary program aims to promote literacy skills and foster the development of reading and writing strategies by tailoring individualized lessons to each student. Tutoring is delivered by trained *Reading Recovery*® teachers in daily 30 minute pull-out sessions over the course of 12–20 weeks.

Research²

The What Works Clearinghouse (WWC) identified three studies of *Reading Recovery*® that both fall within the scope of the Beginning Reading topic area and meet WWC evidence standards. All three studies meet standards without reservations. Together, these studies included 227 students in first grade in at least 14 states.

The WWC considers the extent of evidence for *Reading Recovery*® on the reading skills of beginning readers to be small for four outcome domains—alphabetics, reading fluency, comprehension, and general reading achievement. (See the Effectiveness Summary on p. 4 for further description of these domains.)

Effectiveness

Reading Recovery® was found to have positive effects on general reading achievement and potentially positive effects on alphabetics, reading fluency, and comprehension for beginning readers.

Report Contents	
Overview	p. 1
Program Information	p. 2
Research Summary	p. 3
Effectiveness Summary	p. 4
References	p. 7
Research Details for Each Study	p. 24
Outcome Measures for Each Domain	p. 28
Findings Included in the Rating for Each Outcome Domain	p. 30
Endnotes	p. 34
Rating Criteria	p. 35
Glossary of Terms	p. 36

Table 1. Summary of findings³

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
Alphabetics	Potentially positive effects	+21	+9 to +42	2	148	Small
Reading fluency	Potentially positive effects	+46	+32 to +49	1	74	Small
Comprehension	Potentially positive effects	+14	+6 to +26	2	145	Small
General reading achievement	Positive effects	+27	+19 to +38	3	227	Small

Program Information

Background

Reading Recovery[®] was developed by Dr. Marie M. Clay at the University of Auckland, New Zealand, and is distributed through about 20 university training centers in the United States and supported by the *Reading Recovery*[®] Council of North America (RRCNA). Address: 500 West Wilson Bridge Road, Suite 250, Worthington, OH 43085-5218. Email: jjohnson@readingrecovery.org. Web: <http://www.readingrecovery.org>. Telephone: (614) 310-7323.

Program details

In *Reading Recovery*[®], teachers tailor one-to-one tutoring lessons to accommodate each student's needs. Depending on these needs, teachers incorporate instruction in topics such as phonemic awareness, phonics, vocabulary, fluency, comprehension, writing, motivation, and oral language. Each lesson consists of reading familiar or novel stories, manipulating letters and words, and writing and assembling stories. Lessons are interactive between the teacher and student, with the teacher carefully monitoring each student's reading behavior. *Reading Recovery*[®] lessons are discontinued when students 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 1-year university-based training program and ongoing professional development.

Cost

Reading Recovery[®] is available on a nonprofit, no royalty basis. Because *Reading Recovery*[®] in the United States is a collaboration between universities and school districts, costs include tuition for initial training and continuing professional development. To establish a *Reading Recovery*[®] site—comprised of multiple schools in a district or group of districts—a teacher leader must first be trained. This start-up cost includes paying the teacher leader's salary, paying university tuition for the *Reading Recovery*[®] coursework, and covering the costs of books and materials. Each site must also equip a room with a one-way mirror and sound system to provide subsequent training for teachers.

Teacher leaders work at the site level and provide professional development to *Reading Recovery*[®] teachers. Ongoing costs include support for the teacher leader and a portion of the *Reading Recovery*[®] teachers' salaries and benefits. These specially trained *Reading Recovery*[®] teachers work part of the day in *Reading Recovery*[®] and the remaining part of the day in other capacities such as teaching small literacy groups or classrooms. According to the program developer, the average US *Reading Recovery*[®] teacher worked with eight *Reading Recovery*[®] students and approximately 40 additional students during the 2010–11 school year.

Other related ongoing costs include professional development for both teacher leaders and *Reading Recovery*[®] teachers, books and materials for lessons, student program materials, and data evaluation fees (which cover the cost of updating a site's roster of teachers and schools, data entry, plus ongoing phone and email support from the Help Desk for teacher leaders). The cost of program materials is approximately \$100 per student served (calculated by the RRCNA as an average over the 5-year period from 2007–11). Sites pay an annual data evaluation fee of \$350 a site plus \$45 per *Reading Recovery*[®] teacher. Sites implementing the program also pay annual technical support fees, which vary by the university that provides the *Reading Recovery*[®] training.

Research Summary

The WWC identified 202 studies that investigated the effects of *Reading Recovery*[®] on the reading skills of beginning readers.

The WWC reviewed 79 of those studies against group design evidence standards. Three studies (Pinnell, DeFord, & Lyons, 1988; Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994; and Schwartz, 2005) are randomized controlled trials that meet WWC evidence standards without reservations. Those three studies are summarized in this report. Seventy-six studies do not meet WWC evidence standards.

The remaining 123 studies do not meet WWC eligibility screens for review in this topic area. Citations for all 202 studies are in the References section, which begins on p. 7.

Table 2. Scope of reviewed research

Grade	1
Delivery method	Individual
Program type	Supplement

Summary of studies meeting WWC evidence standards without reservations

Pinnell, DeFord, & Lyons (1988) examined the effect of *Reading Recovery*[®] on the reading skills of first-grade students in urban public schools in Columbus, Ohio who were designated as the lowest 20% of readers in their classroom. In the portion of this study that meets WWC evidence standards without reservations⁴, students attending classrooms in which teachers had not previously been implementing the intervention were randomly assigned either to the *Reading Recovery*[®] intervention or to an alternative compensatory program focused on skills-oriented drill activities. Students in the intervention condition participated in individualized instruction for 30 minutes daily until they reached average levels for the class. Students who reached average levels received, on average, 67 daily lessons. The analysis sample included 74 students (37 in each condition). Outcomes were measured in the spring of first grade.

Pinnell et al. (1994) measured the effect of *Reading Recovery*[®] on the reading skills of first-grade students enrolled in geographically diverse school districts in Ohio. In the portion of the study that meets WWC evidence standards without reservations, low-achieving students within the same schools were randomly assigned either to the *Reading Recovery*[®] condition or to a comparison group in which they continued their regular reading program and existing federally-supported educational assistance services. Comparison group teachers were given the opportunity to select the materials to use with comparison group students; options included materials related to basic reading skills and vocabulary development. Students in the intervention condition read an average of five books per lesson and received an average of 33 minutes of daily individualized instruction. The analysis sample included eight schools with 31 students in the intervention condition and 48 students in the comparison condition. Outcomes were measured in February of first grade.

Schwartz (2005) examined the effect of *Reading Recovery*[®] on the reading skills of first-grade students attending elementary schools in 14 states. Within each participating school, teachers identified two students eligible for *Reading Recovery*[®]; these students were then randomly assigned to receive the program during the first or the second half of the school year. During the transition period between the first and second half of the school year, students assigned to receive the intervention during the first half of the year (intervention group) had finished the program (by either reaching classroom averages or attending the program for 20 weeks), and students assigned to receive the intervention in the second half of the year (comparison group) had not yet been exposed to *Reading Recovery*[®]. During this transition period, reading outcomes were measured for 74 students (37 in each condition).

Summary of studies meeting WWC evidence standards with reservations

No studies of *Reading Recovery*[®] met WWC evidence standards with reservations.

Effectiveness Summary

The WWC review of *Reading Recovery*[®] for the Beginning Reading topic area includes student outcomes in four domains: alphabetics, reading fluency, comprehension, and general reading achievement. The three studies of *Reading Recovery*[®] that meet WWC evidence standards reported findings in all four domains. Findings in the alphabetics domain for this review are differentiated by three constructs (as described in the Beginning Reading review protocol): phonemic awareness, letter knowledge, and phonics. Findings in the comprehension domain are differentiated by two constructs: reading comprehension and vocabulary development. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Reading Recovery*[®] on beginning readers. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 35.

Summary of effectiveness for the alphabetics domain

Two studies that meet WWC standards without reservations reported findings in the alphabetics domain.

One study examined the effect of *Reading Recovery*[®] on the phonemic awareness construct in the alphabetics domain. Schwartz (2005) reported no statistically significant differences for the phonemic awareness measures—the deletion task and the Yopp-Singer Test of Phoneme Segmentation—but the effects on both measures were positive and considered substantively important based on the WWC criteria (that is, at least 0.25).

Two studies examined the effect of *Reading Recovery*[®] on the letter knowledge construct in the alphabetics domain. Pinnell, DeFord, & Lyons (1988) did not find a statistically significant effect for *Reading Recovery*[®] on the Letter Identification subtest of the Observation Survey of Early Literacy Achievement, but the effect was positive and considered substantively important according to WWC criteria. Schwartz (2005) also reported a statistically insignificant effect of *Reading Recovery*[®] on the Letter Identification subtest of the Observation Survey; this difference was positive but not considered substantively important based on WWC criteria.

Two studies examined the effect of *Reading Recovery*[®] on the phonics construct in the alphabetics domain. Pinnell, DeFord, & 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.

The WWC characterizes student findings for Schwartz (2005) as a statistically significant positive effect because the average effect of the four outcomes (across constructs) is positive and statistically significant. Also, the effect on the Word Recognition subtest of the Observation Survey is positive and statistically significant, and no effects are negative and statistically significant for this study. For Pinnell, DeFord, & Lyons (1988), the average effect for the two outcome measures (across constructs) is not statistically significant but is considered to be substantively important based on WWC evidence criteria; therefore, the WWC characterizes these study findings as a substantively important positive effect.

Thus, for the alphabetics domain, among the two studies with a strong design, one showed a statistically significant positive effect and one showed a substantively important positive effect. This results in a rating of potentially positive effects, with a small extent of evidence.

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

Rating of effectiveness	Criteria met
Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i>	In the two studies that reported findings, the estimated impact of the intervention on outcomes in the <i>alphabetics</i> domain was a statistically significant positive effect in one study and a substantively important positive effect in one study.
Extent of evidence	Criteria met
Small	Two studies that included 148 students reported evidence of effectiveness in the <i>alphabetics</i> domain.

Summary of effectiveness for the reading fluency domain

One study that meets WWC standards without reservations reported findings in the reading fluency domain.

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 of Early Literacy Achievement.

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

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

Rating of effectiveness	Criteria met
Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i>	In the one study that reported findings, the estimated impact of the intervention on outcomes in the <i>reading fluency</i> domain was a statistically significant positive effect.
Extent of evidence	Criteria met
Small	One study that included 74 students reported evidence of effectiveness in the <i>reading fluency</i> domain.

Summary of effectiveness for the comprehension domain

Two studies that meet WWC standards without reservations reported findings in the comprehension domain.

Two studies examined the effect of *Reading Recovery*[®] on the reading comprehension construct in the comprehension domain. Pinnell, DeFord, & Lyons (1988) reported, and the WWC confirmed, a substantively important (but statistically insignificant) positive effect on the Reading Comprehension subtest of the Comprehensive Test of Basic Skills (CTBS). Schwartz (2005) reported neither a statistically significant nor a substantively important effect of *Reading Recovery*[®] on the Degrees of Reading Power Test.

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

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

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

Rating of effectiveness	Criteria met
Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i>	In the two studies that reported findings, the estimated impact of the intervention on outcomes in the <i>comprehension</i> domain was a statistically significant positive effect in one study and an indeterminate effect in one study.
Extent of evidence	Criteria met
Small	Two studies that included 145 students reported evidence of effectiveness in the <i>comprehension</i> domain.

Summary of effectiveness for the general reading achievement domain

Three studies reported findings in the general reading achievement domain.

Pinnell, DeFord, & Lyons (1988) found, and the WWC confirmed, positive and statistically significant effects of *Reading Recovery*[®] on three subtests of the Observation Survey of Early Literacy Achievement: Concepts About Print, Hearing and Recording Sounds in Words (Dictation), and Writing Vocabulary.

Pinnell et al. (1994) found, and the WWC confirmed, 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.

Schwartz (2005) found, and the WWC confirmed, positive and statistically significant effects of *Reading Recovery*[®] on three subtests of the Observation Survey: Concepts About Print, Dictation, and Writing Vocabulary.

Thus, for the general reading achievement domain, three studies with strong designs reported statistically significant positive effects. This results in a rating of positive effects, with a small extent of evidence.

Table 6. Rating of effectiveness and extent of evidence for the general reading achievement domain

Rating of effectiveness	Criteria met
Positive effects <i>Strong evidence of a positive effect with no overriding contrary evidence.</i>	In the three studies that reported findings, the estimated impact of the intervention on outcomes in the <i>general reading achievement</i> domain was a statistically significant positive effect.
Extent of evidence	Criteria met
Small	Three studies that included 227 students reported evidence of effectiveness in the <i>general reading achievement</i> domain.

References

Studies that meet WWC evidence standards without reservations

- 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). *Success of at-risk children in a program that combines writing and reading* (Technical Report No. 417). Urbana-Champaign: 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.
- 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: 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.

Studies that meet WWC evidence standards with reservations

None.

Studies that do not meet WWC evidence standards

- 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* (Vol. 2, pp. 115–132). Portsmouth, NH: Heinemann. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Austin Independent School District, Office of Program Evaluation. (2001). *Literacy support plan evaluation, 1999–2000*. Austin, TX: Author. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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: Author.
- Bermel, S. (1987). *Language development component, CLEAR–Reading Recovery program 1985–86. Final evaluation report*. Columbus: Columbus Public Schools, Ohio Department of Evaluation Services. The study does

not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

- Briggs, C., & Young, B. K. (2003). Does Reading Recovery work in Kansas? A retrospective longitudinal study of sustained effects. *Journal of Reading Recovery*, 3(1), 59–64. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Bufalino, J. M. (1993). The sustained effects of Reading Recovery intervention on the reading comprehension of second graders. *Dissertation Abstracts International*, 54(11), 145A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Burroughs-Lange, S. (2006). *Evaluation of Reading Recovery in London schools: Every child a reader 2005–2006*. University of London: Institute of Education. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Additional sources:

- Burroughs-Lange, S. (2008). *Comparison of literacy progress of young children in London schools: A Reading Recovery follow-up study*. Retrieved from <http://www.ioe.ac.uk>
- Burroughs-Lange, S., & Douëttil, J. (2007). Literacy progress of young children from poor urban settings: A Reading Recovery comparison study. *Literacy*, 12(1), 19–46.
- Every Child a Reader. (2006). *Every child a reader: The results of the first year...* Retrieved from <http://readingrecovery.ioe.ac.uk>
- Every Child a Reader. (2007). *Every child a reader: The results of the second year...* Retrieved from <http://readingrecovery.ioe.ac.uk>
- Gross, J. (2006). Every Child a Reader. *Literacy Today*, 49, 10–11.
- Burroughs-Lange, S. (2008). Reading Recovery in London schools. *Literacy Today*, (55), 29. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Caraway, M. A. H. (2006). A cross-sectional study of performance on high-stakes state assessment by at-risk students who were served in an early intervention program. *Dissertation Abstracts International*, 68(01A), 185–125. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly*, 30(2), 240–263. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Collins, E. W. (2000). The immediate and sustained effects of the Reading Recovery program on grade one and grade four at-risk students: A longitudinal study. *Dissertation Abstracts International*, 61(05), 1784A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Collins, V. K. (1994). Automaticity in information processing. *Dissertation Abstracts International*, 55(9), 2708A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Curtin, J. (1993). *The effectiveness of the Reading Recovery Program on reading achievement*. Chicago: Chicago Public Schools. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- D'Agostino, J. V., & Williams, A. K. (2011). *Reading Recovery and Descubriendo la Lectura national report 2009–2010* (No. IDEC 2011-01). Columbus, OH: International Data Evaluation Center. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Additional sources:**
- Gómez-Bellengé, F. (2006, April). *Accounting for every child: The evaluation of Reading Recovery*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Gómez-Bellengé, F. (2007). 2005–06 national data preview: measuring the impact of Reading Recovery. *Journal of Reading Recovery*, 6(2), 53–56.
- International Data Evaluation Center. (2011). *2010–2011 Descubriendo la Lectura national statistical abstract for the U.S.* (No. IDEC Rep. No. 2011-2). Columbus: The Ohio State University.
- Dewoody, M. (2011). *A program evaluation of developmental first grade and Reading Recovery* (Unpublished doctoral dissertation). University of Oklahoma, Norman. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Dorn, L., & Allen, A. (1995). Helping low-achieving first-grade readers: A program combining Reading Recovery tutoring and small-group instruction. *Journal of Reading Recovery*, 13(3), 16–24. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Dunkeld, C. (1990). *Gaining experience with Reading Recovery: A pilot project between Portland Public Schools and Portland State University*. Portland, OR: Portland State University. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was not implemented as designed.
- 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* (Vol. 2, pp. 193–214). Portsmouth, NH: Heinemann. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

- Galluzzo, C. (2010). *The long-term effectiveness of Reading Recovery and the cost-efficiency of Reading Recovery relative to the learning disabled classification rate* (Unpublished doctoral dissertation). State University of New York at Buffalo. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Gilmer, V. B. (2003). Sustained success of former Reading Recovery students. *Dissertation Abstracts International*, 64(02), 444A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Gómez-Bellengé, F. X., & Rodgers, E. M. (2004). *Reading Recovery and Descubriendo la Lectura national report 2002–2003*. Columbus: Ohio State University, College of Education, School of Teaching and Learning. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Gómez-Bellengé, F. X., Rodgers, E., & Fullerton, S. K. (2003). *Reading Recovery and Descubriendo la Lectura national report 2001–2002*. Columbus: Ohio State University, College of Education, School of Teaching and Learning. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Greaney, S. (2011). *Comparing the reading behaviors in three groups of first grade students—students who discontinued from Reading Recovery, students who did not discontinue from Reading Recovery, and students who never needed Reading Recovery* (Unpublished doctoral dissertation). University of Maine, Orono. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Harvey, M. W. (2011). Union County Public Schools action research: *Comparing early literacy interventions used in Union County Public Schools. Reading Recovery vs. Leveled Literacy intervention* (Unpublished doctoral dissertation). Wingate University, NC. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Holcomb, L. L. (2010). An exploration into the longitudinal reading achievement of students in the Cherokee County (NC) school district's application of Reading Recovery. *Dissertation Abstracts International*, 70(9A), 3398. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Iverson, S., & Tunmer, W. E. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology*, 85(1), 112–126. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Additional source:**
- Tunmer, W. E., & Hoover, W. A. (1993). Phonological recoding skills in beginning reading. *Reading and Writing: An Interdisciplinary Journal*, 5, 161–179.
- Johnson, J. A. (1996). *Reading Recovery: Early intervention*. Hays, KS: Fort Hays State University. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Kahl, K. M. (2005). *Comparing outcomes of two early reading interventions: Reading Recovery and direct instruction* (Unpublished master's thesis). Widener University, Chester, PA. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

- LaFave, C. E. (1995). Impact of Reading Recovery on phonemic awareness. *Dissertation Abstracts International*, 56(07), 2621A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Litt, D. G. (2003). An exploration of the double-deficit hypothesis in the Reading Recovery population. *Dissertation Abstracts International*, 64(06), 2028A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Marina, B., & Gilman, D. A. (2003). *Is Reading Recovery worth the cost?* Vigo County, IN: Vigo County School Corporation. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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* (Vol. 2, pp. 231–256). Portsmouth, NH: Heinemann. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- McClendon, I. D. (2012). *A longitudinal case study of a literacy program titled Reading Recovery for students in a struggling midwestern school district* (Unpublished doctoral dissertation). Lindenwood University, St. Charles, MO. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- McIntyre, E., Jones, D., Powers, S., Newsome, F., Petrosko, J., Powell, R., & Bright, K. (2005). Supplemental instruction in early reading: Does it matter for struggling readers? *The Journal of Educational Research*, 99(2), 99–107. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.
- Michigan Department of Evaluation Services. (1995). *Compensatory Education (CE) product evaluation: Elementary and secondary programs 1994–95*. Saginaw, MI: Saginaw Public Schools. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Additional source:**
- Michigan Department of Evaluation Services. (1992). *Compensatory education product evaluation: Elementary and secondary programs 1991–1992*. Saginaw, MI: Saginaw Public Schools.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Moore, M., & Wade, B. (1998). Reading Recovery: Its effectiveness in the long term. *Support for Learning*, 13(3), 123–128. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Murphy, J. A. (2003). An application of growth curve analysis: The evaluation of a reading intervention program. *Dissertation Abstracts International*, 64(12), 4358A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Narramore, J. (2010). *The effectiveness of Reading Recovery on struggling first grade students* (Unpublished doctoral dissertation). Trevecca Nazarene University, Nashville, TN. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Plewis, I. (2000). Evaluating educational interventions using multilevel growth curves: The case of Reading Recovery. *Educational Research and Evaluation*, 6(1), 83–101. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

- Potter, T. (2004). *Reading Recovery evaluation*. Madison, WI: Planning, Research and Evaluation, Madison Metropolitan School District. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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: An International Journal of Early Reading and Writing*, 5(2), 7–25. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Redding, L. R. (2012). *An investigation of the sustained effects of Reading Recovery® on economically disadvantaged fifth grade students* (Unpublished doctoral dissertation). Widener University, Chester, PA. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was not implemented as designed.
- Rodgers, E., & Gómez-Bellengé, F. X. (2006). *Reading Recovery in Ohio: 2005–2006 state report* (National Data Evaluation Center Tech. Rep. No. 2006-08). Columbus, OH: The Ohio State University, National Data Evaluation Center. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Rodgers, E., Gómez-Bellengé, F., Wang, C., & Schulz, M. (2005, April). *Predicting the literacy achievement of struggling readers: Does intervening early make a difference*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Rodgers, E., Gómez-Bellengé, F. X., & Fullerton, S. K. (2003). *Reading Recovery in Ohio: 2001–2002 state report* (National Data Evaluation Center Tech. Rep. No. 2003-03). Columbus: The Ohio State University, College of Education, School of Teaching and Learning. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Rodgers, E. M., Gómez-Bellengé, F. X., & Schulz, M. M. (2005). *Reading Recovery in Ohio: 2003–2004 state report* (National Data Evaluation Center Tech. Rep. No. 2005-01). Columbus: The Ohio State University, College of Education, School of Teaching and Learning. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Ross, S. M., Smight, L. J., Casey, J., & Slavin, R. E. (1995). Increasing the academic success of disadvantaged children: An examination of alternative early intervention programs. *American Educational Research Journal*, 32(4), 773–800. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

Additional sources:

- Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). Success for All: A summary of research. *Journal of Education for Students Placed at Risk*, 1(1), 41–76.
- Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994, April). *Success for All: Longitudinal effects of systemic school-by-school reform in seven districts*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Ruhe, V. (2006). The impact of Reading Recovery on later literacy achievement in Maine: Year 2 report. *ERS Spectrum*, 24(3), 19–28. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Ruhe, V., & Paula, M. (2005). The impact of Reading Recovery on later achievement in reading and writing. *ERS Spectrum*, 23(1), 20–30. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Schmitt, M. C., & Gregory, A. E. (2001, December). *The impact of early intervention: Where are the children now?* Paper presented at the annual meeting of the National Reading Conference, San Antonio, TX. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Shamey, T. (2009). Effects of early elementary Reading Recovery programs on middle-school students: A longitudinal evaluation. *Dissertation Abstracts International*, 69(12A), 4619. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Shoulders, M. D. (2004). The long-term effectiveness of the Reading Recovery program. *Dissertation Abstracts International*, 65(03), 836A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Simpkins, J. (1995). *Longitudinal study of Reading Recovery: School years 1990–91 through 1993–94*. Unpublished manuscript. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Smith, N. (1994). Reading Recovery data and observations from one Illinois site (Part II). *Illinois Reading Council Journal*, 22(3), 29–46. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- 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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Stahl, K. A. D., Stahl, S. A., & McKenna, M. C. (2003). The development of phonological awareness and orthographic processing in Reading Recovery. In S. Forbes & C. Briggs (Eds.), *Research in Reading Recovery* (Vol. 2, pp. 99–114). Portsmouth, NH: Heinemann. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Sylva, K., & Hurry, J. (1996). Early intervention in children with reading difficulties: An evaluation of Reading Recovery and a Phonological Training. *Literacy Teaching and Learning: An International Journal of Early Literacy*, 2(2), 49–73. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Additional source:

Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading*, 30(3), 227–248.

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, IL. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Wang, Y. L., & Johnstone, W. (1997, April). *Evaluation of Reading Recovery program*. Paper presented at the meeting of the American Educational Research Association, Chicago, IL. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Weeks, D. (1992). A study of the implementation of Reading Recovery in Scarborough: 1990–1991. *Masters Abstracts International*, 3(03), 1005. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

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. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Wright, A. (1992). Evaluation of the first British Reading Recovery programme. *British Educational Research Journal*, 18(4), 351–368. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

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: Heinemann. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Zielinski, L. A. (1997). The long term effectiveness of Reading Recovery in a small, rural school district. *Dissertation Abstracts International*, 59(01), 0077A. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Studies that are ineligible for review using the Beginning Reading Evidence Review Protocol

Acalin, T. A. (1995). A comparison of Reading Recovery to Project READ. *Masters Abstracts International*, 33(06), 1660. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Allington, R. L. (2005). How much evidence is enough evidence? *Journal of Reading Recovery*, 4(2), 8–11. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Alvermann, D. E., Simpson, M. L., & Fitzgerald, J. (2006). Teaching and learning in reading. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 427–455). Mahwah, NJ: Lawrence Erlbaum Associates. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Askew, B. J., Fountas, I. C., & Lyons, C. A. (1989). *Reading Recovery review: Understandings, outcomes and implications*. Columbus, OH: Reading Recovery Council of North America. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

- Askew, B. J., Kaye, E., Frasier, D. F., Mobasher, M., Anderson, N., & Rodriguez, Y. (2002). Making a case for prevention in education. *Literacy Teaching and Learning*, 6(2), 43–73. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Assad, S., & Condon, M. A. (1996). Demonstrating the cost effectiveness of Reading Recovery: Because it makes a difference. An example from one school district. *Network News*, 10–14. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Batten, P. (2004). Investing equity funding in early literacy. *ERS Spectrum*, 22(1), 40–45. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Bremer, T. (2007). *Longitudinal sustainability of Reading Recovery intervention at first grade on student ranking and SRI scores on fifth grade students in Title 1 buildings*. Maryville: Northwestern Missouri State University. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Brown, T. (2007). The lasting effects of the Reading Recovery program on the reading achievement on at risk youth. *Dissertation Abstracts International*, 68(02A), 138-511. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Bufalino, J., Wang, C., Gómez-Bellengé, F. X., & Zalud, G. (2010). What's possible for first-grade at-risk literacy learners receiving early intervention services. *Literacy Teaching and Learning*, 15(1), 1–15. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Catholic Education Office Melbourne. (2011). *Reading Recovery in the archdiocese of Melbourne*. Retrieved from <http://www.ceomelb.catholic.edu.au> The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Celebrations as one-to-one tuition helps pupil's literacy. (2006). *Education*, (246), 1. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Charlesworth, A., Charlesworth, R., Raban, B., & Rickards, F. (2006). Reading Recovery for children with hearing loss. *Volta Review*, 106(1), 29–51. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Cheung, A. C. K., & Slavin, R. E. (2012) Effective reading programs for Spanish dominant English language learners (ELLs) in the elementary grades: A synthesis of research. Retrieved from <http://www.bestevidence.org> The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Coats-Kitsopoulos, G. (2011). *The relationship between early literacy assessment and first-grade reading achievement for Native American students* (Unpublished doctoral dissertation). University of South Dakota, Vermillion. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Cohen, S. G., McDonnell, G., & Osborn, B. (1989). Self-perceptions of at-risk and high achieving readers: Beyond Reading Recovery achievement data. In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction: Thirty-eighth yearbook of the National Reading Conference* (pp. 117). Chicago, IL: National Reading Conference. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Concha, J. S. (2005). *Reading Recovery children and early literacy development: Investigation into phonological awareness, orthographic knowledge, oral reading processing, and reading comprehension processing*. University of Maryland, College Park, Department of Curriculum and Instruction. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Cox, B. E., & Hopkins, C. J. (2006). Building on theoretical principles gleaned from Reading Recovery to inform classroom practice. *Reading Research Quarterly*, 41(2), 254–267. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

D'Agostino, J. V., & Murphy, J. A. (2004). A meta-analysis of Reading Recovery in United States schools. *Educational Evaluation and Policy Analysis*, 26(1), 23–38. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Additional source:

D'Agostino, J., & Murphy, J. (2011). A meta-analysis of Reading Recovery in United States schools. In D. Wyse (Ed.), *Literacy Teaching and Education* (vol. 2). London: SAGE.

Daniels, A. (2010). *The implementation of a Reading Recovery program for at-risk first-grade students* (Unpublished doctoral dissertation). Capella University, Minneapolis, MN. The study is ineligible for review because it does not use a comparison group design or a single-case design.

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. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Douëttil, J. (2004). *The long term effects of Reading Recovery on national curriculum tests at end of key stages 1 and 2*. London: Institute of Education. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Duerr, S. A. (2008). Communication, efficacy, and student reading achievement: An unbroken circle. *Dissertation Abstracts International*, 68(8A), 3326. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Dunn, M. (2010). Response to Intervention and reading difficulties: A conceptual model that includes Reading Recovery. *Learning Disabilities*, 8(1), 21–40. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Escamilla, K. (1997). Descubriendo 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. The study is ineligible for review because it does not examine an intervention conducted in English.

European Centre for Reading Recovery. (2012). *Reading Recovery™ annual report for the UK and the Republic of Ireland: 2011-12*. University of London: Institute of Education. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Additional source:

Evans, D. (2012). Early intervention boosts later success. *Times Educational Supplement*, (5023), 14–15.

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. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Florida Center for Reading Research. (2008). *Reading Recovery*. Tallahassee: Florida Center for Reading Research. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Flowers, L. J. (2006). The short- and long-term reading performance of former Reading Recovery students. *Dissertation Abstracts International*, 67(12A), 155-4496. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Foorman, B., & Al Otaiba, S. (2009). Reading remediation: State of the art. In K. Pugh & P. McCardle (Eds.), *How children learn to read: Current issues and new directions in the integration of cognition, neurobiology and genetics of reading and dyslexia research and practice* (pp. 257–274). New York: Psychology Press. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Foorman, B. R., Fletcher, J. M., & Francis, D. J. (2004). Early reading assessment. In W. M. Evers & H. J. Walberg (Eds.), *Testing student learning, evaluating teaching effectiveness* (pp. 81–125). Stanford, CA: Hoover Institution Press. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

- Fraser, H., MacDougall, A., Pirrie, A., & Croxford, L. (2001). *National evaluation of the early intervention program*. KPMG and The University of Edinburgh. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- 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, IL. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Gapp, S. C., Zalud, G., & Pietrzak, D. (2009). End of intervention Reading Recovery decisions and subsequent achievement. *Reading Improvement*, 46(1), 9–18. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Additional source:**
- Gapp, S. C. (2006). An examination of end of treatment Reading Recovery decisions and later achievement. *Dissertation Abstracts International*, 67(11A), 80-4099.
- Gardner, C. (2010). *Impacts of a Reading Recovery intervention on the reading achievement of low-income students on a standardized examination* (Unpublished doctoral dissertation). Texas Southern University, Houston. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Greaney, K., & Arrow, A. (2012). Phonological-based assessment and teaching within a first year reading program in New Zealand. *Australian Journal of Language and Literacy*, 35(1), 9–32. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Gross, J. (2009). Defending Reading Recovery. *Literacy Today*, (58), 8. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Harrison, L. (2002). *A study on the complementary effects of Reading Recovery and small group instruction for reversing reading failure*. Research Summary No 102–103. Little Rock: University of Arkansas at Little Rock. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge, New York. This study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Heenan, R. A. (2007). Literacy and deafness: A qualitative analysis into the efficacy of an adapted Reading Recovery program. *Dissertation Abstracts International*, 67(10), 3711. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% general education students.
- Heibert, E. H. (1994). Reading Recovery in the United States: What difference does it make to an age cohort? *Educational Researcher*, (23), 15–25. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Herman, R., & Stringfield, S. (1997). *Ten promising programs for educating all children: Evidence of impact*. Arlington, VA: Educational Research Service. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Hipp, D. M. (2008). A case study: The effectiveness of a modified Reading Recovery format for an intermediate struggling reader. *Dissertation Abstracts International*, 70(01A), 127-230. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Horner, S. L., & O'Connor, E. A. (2007). Helping beginning and struggling readers to develop self-regulated strategies: A Reading Recovery example. *Reading & Writing Quarterly*, 23(1), 97–109. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- How effective is Reading Recovery? (2006). *Literacy Today*, (46), 30–31. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Hummel-Rossi, B., & Ashdown, J. (2010). *Cost-effectiveness analysis as a decision tool in selecting and implementing instructional interventions in literacy*. Worthington, OH: Reading Recovery Council of North America. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

- Intensive, tailored tuition raise literacy. (2012). *Children and Young People Now*, 2(15), 32–33. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Iverson, S., Tunmer, W. E., & Chapman, J. W. (2005). The effects of varying group size on the Reading Recovery approach to preventive early intervention. *Journal of Learning Disabilities*, 38(5), 456–472. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Jones, N. (2005). *One-to-one vs. two-to-two instruction: A response to Iverson, Tunmer, and Chapman*. Worthington, OH: Reading Recovery Council of North America. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Kamberg, S. (2010). Sustained impact of the Reading Recovery early intervention as students move into more content-laden grades. *Dissertation Abstracts International*, 70(8A), 2936. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% general education students.
- Kelly, P. R., Gómez-Bellengé, F. X., Chen, J., and Schulz, M. M. (2008). Learner outcomes for English language learners. *TESOL Quarterly*, 42(2), 235–260. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Lee, M. (2008). *Annual monitoring of Reading Recovery: The data for 2007*. New Zealand: Ministry of Education. Retrieved from <http://www.educationcounts.govt.nz> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Additional sources:**
- Lee, M. (2009). *Annual monitoring of Reading Recovery: The data for 2008*. New Zealand: Ministry of Education. Retrieved from <http://www.educationcounts.govt.nz>
- Lee, M. (2010). *Annual monitoring of Reading Recovery: The data for 2009*. New Zealand: Ministry of Education. Retrieved from <http://www.educationcounts.govt.nz>
- Leitner, D. (1990). *Portland Reading Recovery program 1989–90 evaluation report. Year 1*. Portland: Portland Public Schools, Oregon Research and Evaluation Department. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Lewis, D. (2011). *The effects of Reading Recovery on the English reading development of native Spanish-speaking ESOL students* (Unpublished doctoral dissertation). University of Maryland, Baltimore. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Litt, D. (2010). Do children selected for Reading Recovery® exhibit weaknesses in phonological awareness and rapid automatic naming? *Literacy Teaching and Learning*, 14(1), 89–102. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Lose, M. K., Schmitt, M. C., Gómez-Bellengé, F. X., Jones, N., Honchell, B., & Askew, B. J. (2007). Reading Recovery and IDEA legislation: Early Intervening Services (EIS) and Response-to-Intervention (RTI). *The Journal of Reading Recovery*, 6(2), 42–47. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Lu, L. Y. (2010). Phonics and semiotics for early intervention. *International Education Studies*, 3(2), 17–26. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.
- Lukin, C., & Estraviz, L. (2010). The relationship between severe oral language impairment and progress with reading intervention. *Australian Journal of Language & Literacy*, 33(2), 126–133. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Lyons, C. (1989). Reading Recovery: A preventative for mislabeling young “at-risk” learners. *Urban Education*, 24(2), 125–139. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Mahyuddin, R., Nor, S. M., Roslan, S., & Elias, H. (2009). Intervention program for early readers: The teething problems. *International Journal of Learning*, 16(6), 169–176. The study is ineligible for review because it does not include a student outcome.

- Manzo, K. K. (2006). Reading Recovery. *Education Week*, 25(28), 16. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- McCabe, M. (2006). Enhancing face value: A description of teacher and student negotiation of power and politeness in a one-to-one first-grade reading intervention. *Dissertation Abstracts International*, 67(10A), 239-3718. The study is ineligible for review because it does not include a student outcome.
- McDowall, S., Boyd, S., Hodgen, E., & van Vliet, T. (2005). *Reading Recovery in New Zealand: Uptake, implementation, and outcomes, especially in relation to Māori and Pasifika students*. Wellington: New Zealand Council for Educational Research. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McGee, L. (2010). *Reading Recovery in Ohio: Executive summary, 2009–2010*. Columbus: The Ohio State University, International Data Evaluation Center. The study is ineligible for review because it does not provide enough information about its design to assess whether it meets standards.
- Additional source:**
International Data Evaluation Center. (2011). 2010–2011 *Reading Recovery statistical abstract for the U.S.* Columbus: The Ohio State University.
- 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. The study is ineligible for review because it does not provide enough information about its design to assess whether it meets standards.
- McNaughton, S. (2008). Looking at school improvement through a Reading Recovery lens. *Literacy Teaching and Learning*, 12(2), 1–17. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Mesmer, H. A. E. (2008). *Tools for matching readers to texts: Research-based practices*. New York: Guilford Press. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Moen, A. D. (2006). A study of the transition of discontinued Reading Recovery students from grade one to grade four. *Dissertation Abstracts International*, 67(04A), 112-1273. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- 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* (Vol. 2, pp. 257–280). Portsmouth, NH: Heinemann. The study is ineligible for review because it does not examine an intervention conducted in English.
- Ng, L. (2006). *Annual monitoring of Reading Recovery: The data for 2005*. Wellington, NZ: Ministry of Education, Research Division. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Ng, L. (2007). *Annual monitoring of Reading Recovery: The data for 2006*. Wellington, NZ: Ministry of Education, Research Division. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Ng, L. (2008). *Annual monitoring of Reading Recovery: The data for 2007*. Wellington, NZ: Ministry of Education, Research Division. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- North American Trainers Group. (2002). *What evidence says about Reading Recovery*. Columbus, OH: Reading Recovery Council of North America. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- North American Trainers Group Research Committee. (n.d.) *Six Reading Recovery studies: Meeting the criteria for scientifically based research*. Worthington, OH: Reading Recovery Council of North America. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

- 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. The study is ineligible for review because it does not include a student outcome.
- Petersen, J. K. (2008). Actions and training in renewal plans designed by Reading Recovery® teacher leaders. *Dissertation Abstracts International*, 68(9A), 3785. The study is ineligible for review because it does not include a student outcome.
- Pinnell, G. S. (1994). Student, text, teacher: Interactive learning in the Reading Recovery program. In F. Lehr & J. Osborn (Eds.), *Reading language and literacy: Instruction for the twenty-first century* (pp. 149–164). New York: Routledge. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- 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. The study is ineligible for review because it does not include a student outcome.
- Pinnell, G. S. (1997). Reading Recovery: A review of research. In J. Squire, J. Flood, & D. Lapp (Eds.), *Handbook of research on teaching literacy through the communicative and visual arts* (pp. 638–654). New York: Routledge. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Pinnell, G. S. (2000). *Reading Recovery: An analysis of a research-based reading intervention*. Columbus, OH: Reading Recovery Council of North America. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Pollock, J. S. (1991). *Reading Recovery program 1990–91. Final evaluation report. Elementary and Secondary Education Act—Chapter 1*. Columbus: Columbus Public Schools, Ohio Department of Program Evaluation. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Pollock, J. S. (1993). *Final evaluation report: Reading Recovery program 1991–92*. Columbus: Columbus Public Schools, Ohio Department of Program Evaluation. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- 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. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Potter, W. (2007). An analysis of the achievement gap of discontinued Reading Recovery students: A longitudinal study of Reading Recovery students. *Dissertation Abstracts International*, 68(3A), 826. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Pullen, P. C., Lane, H. B., Lloyd, J. W., Nowak, R., & Ryals, J. (2005). Effects of explicit instruction on decoding of struggling first grade students: A data-based case study. *Education & Treatment of Children*, 28(1), 63–76. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Rathvon, N. (2008). *Effective school interventions: Evidence-based strategies for improving student outcomes* (2nd ed.). New York: Guilford Press. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Reading Recovery National Network. (2005). *Reading Recovery Annual Report for UK and Ireland 2004–2005*. University of London: Institute of Education. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Reynolds, M., & Wheldall, K. (2007). Reading Recovery 20 years down the track: Looking forward, looking back. *International Journal of Disability, Development, and Education*, 54(2), 199–223. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

- Reynolds, M., Wheldall, K., & Madelaine, A. (2009). The devil is in the detail regarding the efficacy of Reading Recovery: A rejoinder to Schwartz, Hobsbaum, Briggs, and Scull. *International Journal of Disability, Development, and Education*, 56(1), 17–35. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Rinehart, J. S., & Short, P. M. (2010). Viewing Reading Recovery as a restructuring phenomenon. *Journal of School Leadership*, 20(1), 89–109. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Rodgers, E. M., Wang, C., & Gómez-Bellengé, F. X. (2004, April). *Closing the literacy achievement gap with early intervention*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- 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. The study is ineligible for review because it does not include a student outcome.
- Rowe, K. J. (1995). Factors affecting children's progress in reading: Key findings from a longitudinal study. *Literacy Teaching and Learning*, 1(2), 57–110. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Rumbaugh, W., & Brown, C. (2000). The impact of Reading Recovery on students' self-concepts. *Reading Psychology*, 21, 13–31. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- 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. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Scammacca, N., Vaughn, S., Roberts, G., Wanzek, J., & Torgesen, J. K. (2007). *Extensive reading interventions in grades K-3 : From research to practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Schmitt, M. C. (2001). The development of children's strategic processing in Reading Recovery. *Reading Psychology*, 22(2), 129–151. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Schmitt, M. C. (2003). Metacognitive strategy knowledge: Comparison of former Reading Recovery children and their current classmates. *Literacy Teaching and Learning*, 7(1-2), 57–76. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Schotanus, H. (1991). *Reading Recovery pilot project (Laws 1989: 301): Report of results and effectiveness*. Concord: New Hampshire State Department of Education. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Additional sources:**
- Schotanus, H., Chase, C., & Fontaine, A. (1992). *Reading Recovery program, implementation year two, school year 1991–1992. Report of results and effectiveness*. Concord: New Hampshire State Department of Education.
- 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: New Hampshire Department of Education.
- 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: New Hampshire Department of Education.

- Schwartz, R., Hobsbaum, A., Briggs, C., & Scull, J. (2009). Reading Recovery and evidence-based practice: A response to Reynolds and Wheldall (2007). *International Journal of Disability, Development, and Education*, 56(1), 5–15. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Schwartz, R. M. (in press). Reading Recovery research in North America. In B. Watson & B. J. Askew (Eds.), *What is possible? Reading Recovery in an historical and international context: The legacy of Marie Clay*. Portsmouth, NH: Heinemann. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Scull, J. (2010). Embedding comprehension within reading acquisition processes. *Australian Journal of Language & Literacy*, 33(2), 87–107. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Scull, J. A., & Lo Bianco, J. (2008). Successful engagement in an early literacy intervention. *Journal of Early Childhood Literacy*, 8(2), 123–150. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Shanahan, T., & Barr, R. (1995). Synthesis of research on Reading Recovery. *Reading Research Quarterly*, 30, 958–996. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Slavin, R. E. (2005). *Evidence-based reform: Advancing the education of students at risk*. Retrieved from <http://www.americanprogress.org> The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, 79(4), 1391–1466. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2009). *Effective programs for struggling readers: A best evidence synthesis*. Retrieved from <http://www.bestevidence.org> The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Smith, A. B. (2011). *A case study of teacher responsivity in one-on-one and small-group lessons conducted by teachers trained in Reading Recovery* (Unpublished doctoral dissertation). University of Kentucky, Lexington. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- St. John, E. P., Loesch, S. A., & Bardzell, J. S. (2003). *Improving reading and literacy in grades 1–5: A resource guide to research-based programs*. Thousand Oaks, CA: Corwin Press. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology*, 46(6), 1556–1579. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Sullivan, L. J. R. (2008). Perceptions of Reading Recovery teachers regarding teacher change and professional development. *Dissertation Abstracts International*, 69(3A), 866. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Swartz, S. L. (1999). *California Early Literacy Learning and Reading Recovery: Two innovative programs for teaching children to read and write*. Paper presented at the Claremont Reading Conference, Claremont, CA. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Taylor, M. G. (2010). Responses to intervention: A pragmatic investigation of schools as systems, early intervention and special education. *Dissertation Abstracts International*, 71(3A), 897. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

- 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: Columbus Public Schools, Ohio Department of Evaluation Services. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Thomas, R., & Lopes, J. (2007). Systematic reviews of learning by children with learning disabilities. *Psicologia: Teoria, Investigacao e Pratica*, 12(1), 3–12. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Torgesen, J. K. (2002). Lessons learned from intervention research in reading: A way to go before we rest. *Learning and Teaching Reading*, 89–103. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., ... Francis, D. J. (2006). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities*, 39(1), 56–73. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% general education students.
- Vaughn, S., Wanzek, J., Linan-Thompson, S., & Murray, C. S. (2007). Monitoring response to supplemental services for students at risk for reading difficulties: High and low responders. In S. R. Jimerson, M. K. Burns, and A. M. VanDerHeyden (Eds.), *Handbook of Response-to-Intervention: The science and practice of assessment and intervention* (pp. 234–243). New York: Springer Science. The study is ineligible for review because it does not provide enough information about its design to assess whether it meets standards.
- Voyles, J. (2011). *The relationship between oral language proficiency and the success of English language learners in Reading Recovery* (Unpublished doctoral dissertation). Walden University, Baltimore, MD. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Wanzek, J., Vaughn, S., Kim, A., & Cavanaugh, C. L. (2006). The effects of reading interventions on social outcomes for elementary students with reading difficulties: A synthesis. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 22(2), 121–138. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, (28), 179–200. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Wilson, D. L. (2007). Teachers' self-reports of their preference for and use of Reading Recovery. *Dissertation Abstracts International*, 68(3A), 934. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Woods, A., & Henderson, R. (2008). The early intervention solution: enabling or constraining literacy learning. *Journal of Early Childhood Literacy*, 8(3), 251–268. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Appendix A.1: Research details for Pinnell, DeFord, & Lyons (1988)

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.

Table A1. Summary of findings

Meets WWC evidence standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Alphabetics	74 students	+18	No
Comprehension	71 students	+22	Yes
General reading achievement	74 students	+24	Yes

Setting The study took place in 12 urban public schools in Columbus, Ohio.

Study sample The study authors used several comparison groups to examine the effectiveness of the *Reading Recovery*[®] program. The study comparison that meets WWC evidence standards includes students attending classrooms of teachers who had not previously been trained in *Reading Recovery*[®]. Eligible first-grade students were designated as the lowest 20% of readers in their classroom, based on the scores on the Observation Survey of Early Literacy Achievement, teacher judgment, and a standardized test. Thirty-eight students were randomly assigned to participate in the *Reading Recovery*[®] program, and 37 students were randomly assigned to the comparison group. The analysis sample after sample attrition included 37 students in both the intervention and comparison groups.

Intervention group Students in the *Reading Recovery*[®] group attended regular education classes. Each student also participated in individualized instruction with a *Reading Recovery*[®] teacher for 30 minutes daily until the student reached average levels for the class (on average, students who reached average levels received 67 daily lessons).

Comparison group Students in the comparison group attended regular education classes. They also attended an alternative compensatory program focused on a series of skills-oriented drill activities. This program included primarily small group instruction (with minimal individual-level instruction) and was delivered by trained paraprofessionals for approximately 30–45 minutes per day.

Outcomes and measurement Researchers reported outcomes from nine literacy measures, seven of which were included in the WWC review and ratings of effectiveness. Five of the six reported subtests of the Observation Survey⁵ were included in the WWC review of this study: two in the alphabetics domain, including Letter Identification and Word Recognition; and three in the general reading achievement domain, including Concepts About Print, Dictation, and Writing Vocabulary. Results from the Observation Survey: Text Reading Level subtest were not reported in this review because the WWC determined that it was not possible to calculate effect sizes that were comparable to other measures. The study authors also reported two outcome measures that fall into the comprehension domain: the Reading Vocabulary subtest and the Reading Comprehension subtest of the Comprehensive Test of Basic Skills (CTBS). Finally, the study included a writing assessment that does not fall within one of the domains specified in the WWC Beginning Reading protocol. For a more detailed description of the included outcome measures, see Appendix B.

Support for implementation

Reading Recovery® teachers received a full year of special training, during which they practiced teaching using *Reading Recovery*® methods and observed other teachers through a one-way mirror. The 20 teachers who provided the *Reading Recovery*® intervention to the analysis sample included in this WWC review received training from a local teacher leader and were in their first year of teaching the intervention during the time of the study.⁶

Appendix A.2: Research details for Pinnell et al. (1994)

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.

Table A2. Summary of findings

Meets WWC evidence standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General reading achievement	79 students	+21	Yes

Setting

The study took place in ten school districts (two rural, two suburban, and six urban) in Ohio.

Study sample

The authors studied 403 first-grade students distributed across 43 schools from ten districts. The percentage of students in each district who received public assistance in the form of Aid to Dependent Children ranged from 9% and 42%. Four schools per district implemented one of four reading interventions—*Reading Recovery*®, *Reading Success*, *Direct Instruction Skills Plan*, and *Reading and Writing Group*. Within each school, the ten lowest-scoring students were randomly assigned either to participate in the intervention or to participate in the school’s regular reading program. For this report, the WWC looked at results for students in the ten schools (across ten school districts) who were using *Reading Recovery*® as their intervention. These schools all had prior experience implementing *Reading Recovery*®. In the original study design, 100 students were randomly assigned to receive either *Reading Recovery*® or the comparison condition at ten 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 (in eight districts). All students were low achieving, which was defined as students who scored below the 37th percentile on a standardized assessment and who were recommended for compensatory help by their teachers.

Intervention group

The intervention group was composed of 31 low-achieving students across eight schools. Intervention students received one-on-one tutoring with a trained *Reading Recovery*® teacher daily for 30 minutes. The activities led by the teacher were aimed at fostering independent reading skills and included: reading both easier and more challenging books, conducting word analysis in context, and participating in activities aimed at improving writing fluency, such as composing sentences and reconstructing cut-up versions of sentences.

Comparison group

The comparison group included 48 students attending the same eight schools as the intervention group. Students assigned to the comparison group received no special instruction, but continued to participate in their regular reading program and existing federally-funded supplemental education services with an instructional focus on developing basic reading and vocabulary skills. Some lessons from the supplemental education program included teachers reading aloud as well as group reading. Comparison group teachers, none of whom had received *Reading Recovery*[®] training, selected instructional materials based on their own discretion.

Outcomes and measurement

This WWC review focuses on outcomes measured in February of the academic year in which the study took place because, at that point, no comparison group students had been exposed to the intervention. The WWC review does not include assessments that were measured in May of the same academic year because, at that time, a portion of students who had originally been assigned to the comparison condition had participated in the intervention. Three measures were administered to assess student outcomes in the general reading achievement domain: the Dictation subtest of the Observation Survey of Early Literacy Achievement, 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. For a more detailed description of the included outcome measures, see Appendix B.

Support for implementation

At least two years prior to the study, *Reading Recovery*[®] teachers received specialized training. During this training period that took place over the course of an academic year, the teachers participated in weekly 2.5 hour long sessions, in which they practiced teaching using *Reading Recovery*[®] methods and observed other teachers through a one-way mirror. They also received a 1-day orientation at the beginning of the study.

Appendix A.3: Research details for Schwartz (2005)

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.

Table A3. Summary of findings

Meets WWC evidence standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Alphabetics	74 students	+23	Yes
Reading fluency	74 students	+46	Yes
Comprehension	74 students	+6	No
General reading achievement	74 students	+35	Yes

Setting

The study took place in an unspecified number of elementary schools in 14 states.

Study sample

The study was designed to examine the effect of *Reading Recovery*[®] on the outcomes of first-grade students. Forty-seven *Reading Recovery*[®] teachers each identified two students⁷ eligible for *Reading Recovery*[®] based on their low scores on the Observation Survey of Early Literacy Achievement and their own judgment. These 94 students were randomly assigned to enter the *Reading Recovery*[®] program during either the first or second half of the school year. [Note: The study also 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 an intervention.] Because of missing test data, the author's final analytic sample included 74 students distributed across 37 teachers.

Intervention group

Students participated in the one-on-one daily 30-minute tutoring program for up to 20 weeks or until they were judged by their teacher to have met the criteria for termination of the program by reaching average levels of literacy performance. The length of program participation ranged from 12 to 20 weeks. Originally, participants were taught by 47 *Reading Recovery*[®] teachers who had volunteered to be part of the study, but because of missing test data, data from only 37 teachers and 37 students were included in the author's final analysis. The intervention group was 61% male, 47% Black, 38% White, 12% Hispanic, and 3% Asian. About 60% of the group received free or reduced-price lunch.

Comparison group

The comparison group included students who were randomly assigned to receive *Reading Recovery*[®] during the second half of the year. Thus, these participants served as a comparison group only during the first part of the year when they received instruction in their regular classroom but no additional supplemental services. The final analysis included data from 37 teachers and 37 students. The comparison group was 41% male, 47% White, 38% Black, and 15% Hispanic. Approximately 57% of the group received free or reduced-price lunch.

Outcomes and measurement

The study author reported outcomes on ten literacy measures, all of which were included in the WWC review and ratings of effectiveness. Six reported subtests of the Observation Survey were included in the WWC review of this study: two in the alphabetic domain, including Letter Identification and Word Recognition; one in the fluency domain (Text Reading Level); and three in the general reading achievement domain, including Concepts About Print, Dictation, and Writing Vocabulary. The study author also reported two additional outcome measures that fall into the alphabetic domain, Phoneme Segmentation and Deletion task, one additional outcome in the fluency domain, Slosson Oral Reading Test–Revised, and one outcome in the comprehension domain, Degrees of Reading Power. For a more detailed description of the included outcome measures, see Appendix B.

Support for implementation

Although the study provided no information about training provided to participating teachers, *Reading Recovery*[®] teachers typically must complete a year-long certification program.

Appendix B: Outcome measures for each domain

Alphabetics	
Phonemic awareness	
<i>Deletion task</i>	A ten-item version of the Rosner task, this assessment 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).
<i>Yopp-Singer Test of Phoneme Segmentation</i>	Developed by Hallie K. 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 Schwartz, 2005).
Letter knowledge	
<i>Observation Survey of Early Literacy Achievement: Letter Identification subtest</i>	Students identify upper- and lowercase letters. This assessment, developed by Dr. Marie M. Clay, is not standardized (as cited in Pinnell, DeFord, & Lyons, 1988; Schwartz, 2005).
Phonics	
<i>Observation Survey of Early Literacy Achievement: Word Recognition subtest (also known as the Ready to Read or Ohio Word Test)</i>	Students read 20 common sight words from basic reading texts, and their accuracy is scored. This assessment, developed by Dr. Marie M. Clay, is not standardized (as cited in Pinnell, DeFord, & Lyons, 1988; Schwartz, 2005).
Reading fluency	
<i>Observation Survey of Early Literacy Achievement: Text Reading Level subtest^a</i>	This subtest measures 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 each student's error rate and self-correcting behavior are recorded using the running record technique. Students read from leveled texts drawn from a basal reading series until their accuracy rate falls below 90%. Results are translated to a numerical reading level from level one to level 30, which in turn match up to grade-level equivalency. This assessment is not standardized (as cited in Schwartz, 2005).
<i>Slosson Oral Reading Test–Revised (SORT-R3)</i>	Developed by Richard L. Slosson and Charles L. Nicholson, this 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 mispronounced. The measure is standardized and norm-referenced (as cited in Schwartz, 2005).
Comprehension	
Reading comprehension	
<i>Comprehension Test of Basic Skills (CTBS): Reading Comprehension subtest</i>	This subtest is a group-administered, standardized assessment of reading comprehension (as cited in Pinnell, DeFord, & Lyons, 1988).
<i>Degrees of Reading Power Test</i>	This test is an untimed standardized assessment 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	
<i>CTBS: Reading Vocabulary subtest</i>	A group-administered, standardized assessment of vocabulary (as cited in Pinnell, DeFord, & Lyons, 1988).
General reading achievement	
<i>Gates-MacGinitie Reading Test (1978)</i>	A standardized test, this assessment 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).
<i>Observation Survey of Early Literacy Achievement: Concepts About Print subtest</i>	Students perform tasks related to printed language concepts (for example, directionality, book handling, and word concepts) while reading a book. This assessment, developed by Dr. Marie M. Clay, is not standardized (as cited in Pinnell, DeFord, & Lyons, 1988; Schwartz, 2005).

<i>Observation Survey of Early Literacy Achievement: Hearing and Recording Sounds in Words (Dictation) subtest</i>	For this subtest, students write the words that are dictated to them in sentence form. This assessment, developed by Dr. Marie M. Clay, is not standardized (as cited in Pinnell, DeFord, & Lyons, 1988; Pinnell et al., 1994; Schwartz, 2005).
<i>Observation Survey of Early Literacy Achievement: Writing Vocabulary subtest</i>	For this subtest, students are given ten 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 Pinnell, DeFord, & Lyons, 1988; Schwartz, 2005).
<i>Woodcock Reading Mastery Test–Revised</i>	A standardized test composed of six subtests, this assessment measures the 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).

^a For Pinnell et al. (1988) and Pinnell et al. (1994), findings based on the Observation Survey of Early Literacy Achievement: 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 studies. The 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 authors 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 C.1: Findings included in the rating for the alphabetics domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Construct: Phonemic awareness								
Schwartz, 2005^a								
<i>Deletion task</i>	Grade 1	74 students	6.64 (2.56)	5.58 (2.50)	1.06	0.41	+16	> 0.05
<i>Yopp-Singer Test of Phoneme Segmentation</i>	Grade 1	74 students	17.70 (4.93)	15.27 (5.43)	2.43	0.46	+18	> 0.05
Construct: Letter knowledge								
Pinnell, DeFord, & Lyons, 1988^b								
<i>Observation Survey of Early Literacy Achievement: Letter Identification subtest</i>	Grade 1	74 students	52.27 (1.41)	51.19 (3.17)	1.08	0.44	+17	0.06
Schwartz, 2005^a								
<i>Observation Survey of Early Literacy Achievement: Letter Identification subtest</i>	Grade 1	74 students	52.18 (1.27)	51.68 (2.78)	0.50	0.23	+9	> 0.05
Construct: Phonics								
Pinnell, DeFord, & Lyons, 1988^b								
<i>Observation Survey of Early Literacy Achievement: Word Recognition subtest</i>	Grade 1	74 students	13.68 (1.63)	12.51 (2.87)	1.17	0.50	+19	0.04
Schwartz, 2005^a								
<i>Observation Survey of Early Literacy Achievement: Word Recognition subtest</i>	Grade 1	74 students	14.96 (3.99)	8.87 (4.75)	6.09	1.37	+42	< 0.01
Domain average for alphabetics (Pinnell, DeFord, & Lyons, 1988)						0.47	+18	Not statistically significant
Domain average for alphabetics (Schwartz, 2005)						0.62	+23	Statistically significant
Domain average for alphabetics across all studies						0.55	+21	na

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

^a For Schwartz (2005), no corrections for clustering or multiple comparisons were needed as the authors adjusted for multiple comparisons. The p-values presented here were reported in the original study. For the *Letter Identification* and *Word Recognition* outcomes, the WWC calculated the program group means using a difference-in-differences approach (see WWC Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group post-test means. Mean gains were not available for the two phonemic awareness outcomes, and thus, the WWC reports unadjusted posttest means for the intervention group. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant.

^b For Pinnell, DeFord, & Lyons (1988), a correction for multiple comparisons was needed and resulted in a WWC-computed critical p-value of 0.025 for the *Word Recognition* test; therefore, the WWC does not find the individual results to be statistically significant. However, this study is characterized as having a substantively important positive effect because the mean effect size for the measures of outcomes in the domain is positive and greater than 0.25. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, p. 96.

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

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Schwartz, 2005^a								
<i>Slosson Oral Reading Test–Revised</i>	Grade 1	74 students	30.58 (14.41)	18.12 (11.87)	12.46	0.93	+32	< 0.01
<i>Observation Survey of Early Literacy Achievement: Text Reading subtest</i>	Grade 1	74 students	0.78	0.05	0.73	2.49	+49	< 0.01
Domain average for reading fluency (Schwartz, 2005)						1.71	+46	Statistically significant

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study’s domain average was determined by the WWC.

^a For Schwartz (2005), no corrections for clustering or multiple comparisons were needed as the authors adjusted for multiple comparisons. The p-values presented here were reported in the original study. Means presented for the *Text Reading* subtest are the posttest proportions for each group scoring at or above a first-grade reading level (provided by the study author). Effect size is computed as a Cox’s index: logged-odds ratio transformation divided by 1.65. See the WWC Handbook, Version 2.1 for the computation of effect sizes for binary outcomes. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant.

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

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Construct: Reading comprehension								
Pinnell, DeFord, & Lyons, 1988^a								
<i>Comprehension Test of Basic Skills (CTBS) Reading Comprehension subtest</i>	Grade 1	70 students	36.67 (19.27)	28.88 (14.53)	7.79	0.45	+17	0.06
Schwartz, 2005^b								
<i>Degrees of Reading Power Test</i>	Grade 1	74 students	4.82 (3.88)	4.27 (3.88)	0.55	0.14	+6	> 0.05
Construct: Vocabulary development								
Pinnell, DeFord, & Lyons, 1988^a								
<i>CTBS Reading Vocabulary subtest</i>	Grade 1	71 students	36.64 (11.93)	26.11 (16.86)	10.53	0.71	+26	< 0.01
Domain average for comprehension (Pinnell, DeFord, & Lyons, 1988)						0.58	+22	Statistically significant
Domain average for comprehension (Schwartz, 2005)						0.14	+6	Not statistically significant
Domain average for comprehension across all studies						0.36	+14	na

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

^a For Pinnell, DeFord, & Lyons (1988), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-values presented here were calculated from *t*-statistics reported in the original study. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant.

^b For Schwartz (2005), no correction for clustering or multiple comparisons were needed as the authors adjusted for multiple comparisons. The *p*-values presented here were reported in the original study. This study is characterized as having an indeterminate effect because the single effect is neither statistically significant nor substantively important.

Appendix C.4: Findings included in the rating for the general reading achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			<i>p</i> -value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Pinnell, DeFord, & Lyons, 1988^a								
<i>Observation Survey of Early Literacy Achievement: Concepts About Print subtest</i>	Grade 1	74 students	15.81 (2.91)	14.30 (3.08)	1.51	0.50	+19	0.04
<i>Observation Survey of Early Literacy Achievement: Dictation subtest</i>	Grade 1	74 students	30.62 (6.13)	24.38 (6.92)	6.24	0.94	+33	< 0.01
<i>Observation Survey of Early Literacy Achievement: Writing Vocabulary subtest</i>	Grade 1	74 students	32.86 (13.49)	26.05 (14.32)	6.81	0.48	+19	0.04
Pinnell et al., 1994^b								
<i>Gates-MacGinitie Reading Test (1978)</i>	Grade 1	79 students	36.19 (13.12)	31.00 (nr)	5.19	0.51	+19	< 0.05
<i>Observation Survey of Early Literacy Achievement: Dictation subtest</i>	Grade 1	79 students	31.74 (6.18)	26.75 (nr)	4.99	0.65	+24	0.01
<i>Woodcock Reading Mastery Test—Revised</i>	Grade 1	79 students	39.81 (21.35)	39.49 (nr)	0.32	0.49	+19	< 0.05
Schwartz, 2005^c								
<i>Observation Survey of Early Literacy Achievement: Concepts About Print subtest</i>	Grade 1	74 students	19.24 (2.55)	16.68 (2.30)	2.56	1.04	+35	< 0.01
<i>Observation Survey of Early Literacy Achievement: Dictation subtest</i>	Grade 1	74 students	35.58 (2.70)	29.08 (7.37)	6.50	1.16	+38	< 0.01
<i>Observation Survey of Early Literacy Achievement: Writing Vocabulary subtest</i>	Grade 1	74 students	42.67 (11.42)	31.00 (12.94)	11.67	0.95	+33	< 0.01
Domain average for general reading achievement (Pinnell, DeFord, & Lyons, 1988)						0.64	+24	Statistically significant
Domain average for general reading achievement (Pinnell et al., 1994)						0.55	+21	Statistically significant

Domain average for general reading achievement (Schwartz, 2005)	1.05	+35	Statistically significant
Domain average for general reading achievement across all studies	0.75	+27	na

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

^a For Pinnell, DeFord, & Lyons (1988), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. The p -values presented here were calculated from t -statistics reported in the original study. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant.

^b For Pinnell et al. (1994), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. The p -values presented here were reported in the original study. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant.

^c For Schwartz (1994), no corrections for clustering or multiple comparisons were needed as the authors adjusted for multiple comparisons. The p -values presented here were reported in the original study. The WWC calculated the program group means using a difference-in-differences approach (see WWC Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, p. 96.

Endnotes

¹ The descriptive information for this program was obtained from a publicly available source: the program's website (<http://www.readingrecovery.org>; downloaded December 2011). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in March 2012, and we incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by December 2012.

² This report has been updated to include reviews of 96 studies that have been reviewed since the previous intervention report was released in December 2008. The additional 96 studies were not within the scope of the review protocol for the Beginning Reading topic area or were within the scope of the review protocol but did not meet evidence standards. In addition, two studies, (Iverson and Tunmer, 1993 and Baenen et al., 1997), which met WWC evidence standards with reservations in the previous report, do not meet WWC evidence standards with or without reservations in this report. These revised dispositions are due to a change in the review standards. In particular, in the version 1.0 standards, a statistical adjustment for baseline differences was sufficient to demonstrate equivalence in quasi-experimental studies; in the protocol version 2.1 standards, if differences are too great at baseline (greater than 25% of the pooled standard deviation), then the study cannot meet standards (even after a statistical adjustment). A complete list of all studies reviewed and their dispositions are provided in the references. The studies in this report were reviewed using WWC Evidence Standards, version 2.1, as described in the Beginning Reading protocol (version 2.1). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

³ For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 35. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

⁴ In the WWC *Reading Recovery*[®] intervention report that was published in 2008, the WWC review focused on a slightly different comparison sample of 51 students. During the revised WWC review conducted for this report, it was determined, based on the published documents combined with information obtained through an author query, that the most appropriate random assignment comparison is based on the comparison group of 37 students as reported in Pinnell et al. (1986). Since it is not entirely clear whether the remaining 14 students were randomly assigned in the same manner, the WWC assessed whether the group of 37 *Reading Recovery*[®] students was equivalent to the larger comparison group of 51 students on pretest scores. These groups were not deemed to be equivalent, and thus, this comparison does not meet WWC evidence standards. Similarly, 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 neither randomly assigned to *Reading Recovery*[®] nor randomly assigned 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 trained *Reading Recovery*[®] teacher as a regular classroom teacher goes beyond the standard implementation of the program. Also, this comparison does not meet WWC evidence standards due to lack of statistical adjustment for differences in pretest reading scores as required by the WWC.

⁵ The Observation Survey of Early Literacy Achievement was developed by Dr. Marie M. Clay, who also developed *Reading Recovery*[®]. Although there is no evidence of obvious overalignment between the measure and the intervention (intervention students receiving exposure to the measure during the course of the intervention), it should be noted that the developer of the intervention and the measure were the same.

⁶ Twelve teachers received training from a university program and were in their second year of teaching the intervention during the time of the study. These teachers provided the program to students in the non-random assignment portion of the study that did not meet WWC evidence standards.

⁷ The teachers initially identified five students. The lowest three students in the class automatically received *Reading Recovery*[®], and the remaining two were randomly assigned.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, July). *Beginning Reading intervention report: Reading Recovery*[®]. Retrieved from <http://whatworks.ed.gov>

WWC Rating Criteria

Criteria used to determine the rating of a study

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

Criteria used to determine the rating of effectiveness for an intervention

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

Criteria used to determine the extent of evidence for an intervention

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

Glossary of Terms

Attrition	Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.
Clustering adjustment	If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.
Confounding factor	A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.
Design	The design of a study is the method by which intervention and comparison groups were assigned.
Domain	A domain is a group of closely related outcomes.
Effect size	The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.
Eligibility	A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.
Equivalence	A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.
Extent of evidence	An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 35.
Improvement index	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from -50 to +50.
Multiple comparison adjustment	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
Quasi-experimental design (QED)	A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.
Randomized controlled trial (RCT)	A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.
Rating of effectiveness	The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 35.
Single-case design	A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.
Standard deviation	The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.
Statistical significance	Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).
Substantively important	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

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