

WWC Review of the Report “Evaluation of the i3 Scale-up of Reading Recovery Year One Report, 2011–12”¹

The findings from this review do not reflect the full body of research evidence on *Reading Recovery*[®].²

What is this study about?

The study examined the Year 1 impacts of the Investing in Innovation (i3) scale-up of *Reading Recovery*[®], a short-term intervention for struggling readers, on the reading achievement of first-grade students.³

A total of 628 schools participated in a scale-up of *Reading Recovery*[®].⁴ Of those, 209 were randomly selected to participate in the randomized controlled trial in the first year of the study.⁵ The design of this study is an individual-level randomized controlled trial, where the students with the greatest reading needs in each school are randomly assigned either to receive *Reading Recovery*[®] or to continue with normal classroom instruction. First-grade students at the selected schools were screened using the *Reading Recovery*[®] Observation Survey of Early Literacy Achievement to identify the eight students with the lowest reading levels for inclusion in the study sample.⁶

Of the 209 randomly selected schools, a total of 158 completed the random assignment of students to the intervention and comparison groups.⁷ Four matched pairs of students were formed at each school using screening scores and English learner status. One student from each matched pair was randomly assigned to the intervention group, which received 30-minute sessions of *Reading Recovery*[®] 5 days a week for 12–20 weeks. The other student was assigned to a delayed-implementation comparison group, which received normal classroom instruction during the intervention period.

The analytic sample consisted of 866 students (433 in the intervention condition and 433 in the comparison condition) from 147 schools. The primary outcome, general reading achievement, was measured mid-year using the Iowa Test of Basic Skills (ITBS) Total Reading Score. Two subtests, ITBS Reading Words and ITBS Reading Comprehension, were also assessed.⁸

Features of *Reading Recovery*[®]

Reading Recovery[®] is a short-term intervention that provides one-on-one tutoring to first-grade students who are struggling in reading. The supplementary program aims to promote literacy skills and foster the development of reading strategies by tailoring individualized lessons to each student. Tutoring is delivered by *Reading Recovery*[®] teachers in 30-minute pull-out sessions, which include reading familiar books, story composition, assembling stories using cut-up sentences, and previewing and reading new books. Sessions are held daily for 12–20 weeks. *Reading Recovery*[®] teachers receive extensive training on the design and implementation of *Reading Recovery*[®] lessons, documenting lesson activities, and collecting data to track student progress and inform lesson planning.

What did the study find?

The study authors found, and the WWC confirmed, that *Reading Recovery*[®] had a significant positive impact on the general reading achievement of struggling readers in the first grade. The authors also reported, and the WWC confirmed, statistically significant positive impacts of *Reading Recovery*[®] in the general reading achievement and reading comprehension domains.

WWC Rating

The research described in this report meets WWC group design standards without reservations

This study is a well-executed randomized controlled trial with low levels of sample attrition.

Appendix A: Study details

May, H., Gray, A., Gillespie, J. N., Sirinides, P., Sam, C., Goldsworthy, H., Armijo, M., & Tognatta, N. (2013). *Evaluation of the i3 scale-up of Reading Recovery year one report, 2011-12*. Philadelphia, PA: Consortium for Policy Research in Education.

Setting The study was conducted in first-grade classrooms in schools in the United States.

Study sample From a total of 628 schools in multiple states participating in an i3 scale-up study of *Reading Recovery*[®], 209 schools were randomly selected to participate in this randomized controlled trial. Of those, 158 schools carried out the student-level random assignment process, forming matched pairs of students and randomly assigning one student from each pair to the intervention group and one to the comparison group. In total, 628 students were assigned to the intervention group and 625 students to the comparison group.

The analytic sample included only student pairs for whom complete data were available: 866 students in 147 schools, with 433 students in the intervention group and 433 students in the comparison group. In the analytic sample, 61% of the students in the intervention group were male, 17% were English learners, 57% were White, 22% were Hispanic, 18% were African American, and 3% were categorized as other race. In the comparison group, 61% of students were male, 18% were English learners, 56% were White, 20% were Hispanic, 19% were African American, and 5% were categorized as other race.

Intervention group Students in the intervention group were pulled out of the classroom for 30 minutes a day for one-on-one sessions with a *Reading Recovery*[®] teacher. The sessions included reading familiar books, story composition, assembling stories using cut-up sentences, and previewing and reading new books. Frequent progress monitoring by the *Reading Recovery*[®] teacher allowed sessions to be tailored to each student's needs.

Reading Recovery[®] lessons are discontinued when students demonstrate the ability to consistently read at the average level for their grade—this typically occurs 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.⁹

Comparison group Students in the comparison group received regular classroom instruction in the reading curriculum; they received no supplemental instruction during the intervention period. After the mid-year administration of the posttest, students in the comparison group were eligible to receive instruction in *Reading Recovery*[®] during the remainder of the school year.

Outcomes and measurement The ITBS Total Reading test was used to assess students' general reading achievement levels. The Total Reading test includes two subtests: Reading Comprehension and Reading Words. For a more detailed description of these outcome measures, see Appendix B. The test was administered mid-year, after the completion of the intervention.

Support for implementation

Reading Recovery[®] teachers participated in training sessions at designated facilities or at the schools where the teachers worked. In the sessions, teachers were trained to design and implement daily lessons tailored to the needs of the individual student. Teachers also learned to document lesson activities and collect data to track student progress and inform lesson planning. Teacher learning was supported in three main ways: (a) Teachers completed a 1-week summer course that addressed the interpretation and scoring of the Observation Survey of Early Literacy Achievement (the pretest given to students in the evaluation to assess their reading level); (b) Teachers completed a year-long academic course taught by a *Reading Recovery*[®] teacher leader, where they attended weekly 3-hour training sessions; and (c) Teachers were observed by and received feedback from their teacher leader.

Reason for review

This study was identified for review by receiving media attention.

Appendix B: Outcome measures for each domain

General reading achievement

*Iowa Test of Basic Skills (ITBS)
Reading Total*

The ITBS is a norm-referenced standardized test. The Reading Total test includes the Reading Words and Reading Comprehension subtests. The version used in this study (form A, level 6) is intended for use with students in kindergarten and first grade.

ITBS Reading Words subtest

The ITBS Reading Words subtest includes three parts of the Reading Total test: Words, Pictures, and Word Attack.

Reading comprehension

*ITBS Reading Comprehension
subtest*

The ITBS Reading Comprehension subtest includes three parts of the Reading Total test: Sentences, Picture Story, and Story.

Appendix C: Study findings for the general reading achievement domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
General reading achievement								
<i>Iowa Test of Basic Skills (ITBS) Reading Total (scale score)</i>	First-grade students, post intervention	866 students	139.24 (7.60)	135.00 (6.20)	4.24	0.61	+23	< .01
Domain average for general reading achievement						0.61	+23	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 individual outcomes, representing the average change expected for all individuals 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 individual's percentile rank that can be expected if the individual is given the intervention. The statistical significance of the study's domain average was determined by the WWC.

Study Notes: The WWC calculated the intervention group mean by adding the impact of the intervention, as estimated by a 3-level hierarchical linear model (HLM) analysis (students nested within matched pairs and matched pairs nested within schools), to the unadjusted comparison group posttest mean. The WWC effect sizes differ slightly from those reported by the authors. The WWC calculates effect sizes using Hedges' *g*, in which the mean difference between the intervention and comparison groups is divided by the pooled standard deviation of the intervention and comparison groups. The authors reported effect sizes in terms of Glass' Δ , in which the mean difference between the two groups is divided by the standard deviation of the comparison group.

No corrections for clustering or multiple comparisons and no difference-in-differences adjustment were needed. The *p*-value presented here was reported in the original study. Results for the ITBS subtests are presented in Appendix D. This study is characterized as having a statistically significant positive effect because the effect reported is positive and statistically significant. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 24.

Appendix D: Supplemental study findings by domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
General reading achievement								
<i>Iowa Test of Basic Skills (ITBS) Reading Words subtest (scale score)</i>	First-grade students, post intervention	866 students	141.26 (9.00)	136.70 (7.60)	4.56	0.55	+21	< .01
Reading comprehension								
<i>ITBS Reading Comprehension subtest (scale score)</i>	First-grade students, post intervention	866 students	140.01 (8.90)	135.50 (7.40)	4.51	0.55	+21	< .01

Table Notes: The supplemental findings presented in this table are additional findings that do not factor into the determination of the evidence rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals 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 individual's percentile rank that can be expected if the individual is given the intervention.

Study Notes: The WWC calculated the intervention group mean by adding the impact of the intervention, as estimated by a 3-level HLM analysis (students nested within matched pairs and matched pairs nested within schools), to the unadjusted comparison group posttest mean. The WWC effect sizes differ slightly from those reported by the authors. The WWC calculates effect sizes using Hedges' *g*, in which the mean difference between the intervention and comparison groups is divided by the pooled standard deviation of the intervention and comparison groups. The authors reported effect sizes in terms of Glass' Δ , in which the mean difference between the two groups is divided by the standard deviation of the comparison group. No corrections for clustering or multiple comparisons and no difference-in-differences adjustment were needed. The *p*-values presented here were reported in the original study. The study also presents the raw scores for the ITBS Reading Words and Reading Comprehension subtests; however, those scores are not presented here because they are redundant with the scale scores.

Endnotes

¹ Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the authors) to assess whether the study design meets WWC group design standards. The review reports the WWC's assessment of whether the study meets WWC group design standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the single study review protocol, version 2.0. The WWC rating applies only to the study outcomes that were eligible for review under this topic area. The reported analyses in this single study review are only for those eligible outcomes that either met WWC group design standards without reservations or met WWC group design standards with reservations, and do not necessarily apply to all results presented in the study.

² The WWC released an intervention report on *Reading Recovery*[®] in the Beginning Reading topic area in July 2013, which does not include this study. See U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, July). *Beginning Reading intervention report: Reading Recovery*[®], available at <http://whatworks.ed.gov>.

³ As part of the American Recovery and Reinvestment Act of 2009 (ARRA), Title XIV, Public Law 111-5, the i3 Fund awards grants to educational entities for the development and expansion of innovative educational practices. These grants are known as i3 grants. i3 “scale-up” grants are awarded for the purpose of implementing, in large numbers of schools, interventions that have demonstrated effectiveness in smaller numbers of schools. *Reading Recovery*[®] was awarded such a grant in 2010 to expand the program to more than 2,000 schools, 628 of which participated in this study.

⁴ The report indicates that participating schools were recruited by *Reading Recovery*[®] 19 University Training Centers in 18 states. However, the report does not enumerate the criteria used to select the schools or the states in which the selected schools are located.

⁵ Due to concerns about the burden of the randomized controlled trial design on participating schools, the 628 i3 schools were randomly assigned to one of three evaluation components in the 2011–12 study year; the randomized controlled trial, a regression discontinuity design, or an internal *Reading Recovery*[®] evaluation. Participating schools will rotate to different evaluation components in subsequent study years. Impacts from the randomized controlled trial design only are presented in the Year 1 report. Regression discontinuity design results will be presented in a follow-up report.

⁶ In general, students with the lowest scores were selected for participation, although some schools excluded students with particular types of disabilities.

⁷ The authors did not collect data on why these 51 schools did not comply with random assignment in the 2011–12 school year. More comprehensive monitoring data on random assignment from the 2012–13 school year suggest that much of the noncompliance was beyond the control of the selected schools (for example, because *Reading Recovery*[®] was discontinued at the school).

⁸ Subtest findings are reported in Appendix D. In addition to the findings for the full sample, the study presented results for rural students and English learners. Findings for those subgroups are not included in this WWC report, however, because the study did not provide sufficient detail on attrition and baseline equivalence to determine the WWC group design rating for the subgroup results.

⁹ This information was taken from a July 2013 WWC intervention report on *Reading Recovery*[®]; see 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>.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014, October). *WWC review of the report: Evaluation of the i3 scale-up of Reading Recovery year one report, 2011–12*. Retrieved from <http://whatworks.ed.gov>

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 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 analytic sample groups are similar on observed characteristics defined in the review area protocol.
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 study participants 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 eligible study participants are randomly assigned to intervention and comparison groups.
Single-case design (SCD)	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 are 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% ($p < .05$).
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 \(version 3.0\)](#) for additional details.