What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES

Students with Learning Disabilities July 2012

Reading Mastery

Report Contents

Program Description[1]

Reading Mastery is designed to provide explicit reading instruction to students in grades pre-K–5. One of several Direct Instruction curricula from McGraw-Hill, *Reading Mastery* is available in two versions:

1. *Reading Mastery Classic* (for grades pre-K–2) aims to help beginning readers identify letter sounds, segment words into sounds, blend sounds into words, develop vocabulary, and begin to learn comprehension strategies. *Reading Mastery Classic* consists of two levels. *Reading Mastery Fast Cycle* is an accelerated program that condenses Levels I and II.
2. *Reading Mastery Signature Edition* (for grades K–5) is organized by grade level and includes three strands: Reading, Language Arts, and Literature. The Reading strand addresses phonemic awareness, phonics, word analysis, fluency, vocabulary, comprehension, spelling, decoding, and word recognition skills. The Language Arts strand focuses on oral language, communication, and writing skills. The Literature strand is designed to provide students with opportunities to read different types of text and to develop their vocabulary.

Reading Mastery can be used as a supplement to a core reading program or as a stand-alone reading program for students with or without disabilities. This intervention report specifically focuses on the use of *Reading Mastery* to improve the reading and writing skills of students with learning disabilities.

Research[2]

Two studies of *Reading Mastery* that fall within the scope of the Students with Learning Disabilities review protocol meet What Works Clearinghouse (WWC) evidence standards without reservations, and no studies meet WWC evidence standards with reservations. The two studies included 113 students with learning disabilities in grades 2–5 from schools in the southeastern United States. One of the studies compared *Reading Mastery* to *Horizons Fast Track*; both interventions share a developer and many important design features, and *Horizons* was developed in response to feedback on *Reading Mastery*. Based on these two studies, the WWC considers the extent of evidence for *Reading Mastery* on students with learning disabilities to be small for four domains: alphabetics, reading fluency, reading comprehension, and writing. Five other domains are not reported in this intervention report. (See the Effectiveness Summary for further description of all domains.)

(continued)

Effectiveness

Reading Mastery was found to have no discernible effects on reading comprehension and potentially negative effects on alphabetics, reading fluency, and writing for students with learning disabilities.

Table 1. Summary of findings[3]

Outcome domain	Rating of effectiveness	Improvement index *(percentile points)*		Number of studies	Number of students	Extent of evidence
		Average	Range			
Alphabetics	Potentially negative effects	–2	–13 to +7	2	113	Small
Reading fluency	Potentially negative effects	–12	–16 to –9	1	83	Small
Reading comprehension	No discernible effects	–1	na	1	30	Small
Writing	Potentially negative effects	–25	na	1	83	Small

na = not applicable

Program Information

Background

Reading Mastery is based on the original DISTAR (Direct Instruction System for Teaching Arithmetic and Reading) program. Early versions of *Reading Mastery* were developed by Siegfried Engelmann as part of the Direct Instruction teaching model. *Reading Mastery* is distributed by McGraw-Hill Education, P.O. Box 182605, Columbus, OH 43218. Email: SEG_customerservice@mcgraw-hill.com. Website: https://www.mheonline.com. Telephone: (800) 334-7344.

Program details

Reading Mastery is designed for elementary-age students at all levels of reading performance, and can be used with English language learners and students with learning disabilities. A typical 30- to 45-minute *Reading Mastery* lesson includes seven to nine short activities. The activities encompass multiple strands of content, such as phonemic awareness, letter-sound correspondence, sounding out words, word recognition, vocabulary, oral reading fluency, and comprehension. Program materials include fully scripted lessons to guide teachers through repeated instructional steps (model new content, provide guided practice, and implement individual practice and application), and signals and group responses are used to keep students involved, help them stay on task, and to control lesson pacing. Teachers assess student performance throughout the program, and struggling students receive remedial exercises. The program typically spans one academic year.

Cost

Student materials include storybooks (grades pre-K–1) or textbooks (grades 2–6), workbooks, and test books. The cost per student ranges from $200 to $300 for the first year of implementation. A full set of teaching materials—a one-time purchase—costs between $650 and $1,000 for each grade level. Additional components include literature collections, independent readers, seatwork blackline masters, and practice and review CD-ROMs for students. SRA Teaching Tutor CD-ROMs supplement consultant-led professional development. Additional information on costs of training materials and workshops is available at https://www.mheonline.com/.

Research Summary

Seventeen studies reviewed by the WWC investigated the effects of *Reading Mastery* on students with learning disabilities. Two studies (Cooke, Gibbs, Campbell, & Shalvis, 2004; Herrera, Logan, Cooker, Morris, & Lyman, 1997) are randomized controlled trials that meet WWC evidence standards without reservations.[4] These two studies are summarized in this report. The remaining 15 studies do not meet either WWC eligibility screens or evidence standards. (See references beginning on p. 8 for citations for all 17 studies.)

Table 2. Scope of reviewed research

Grade	2, 3, 4, 5
Delivery method	Whole class
Program type	Curriculum
Studies reviewed	17
Meet WWC standards without reservations	2 studies
Meets WWC standards with reservations	0 studies

Summary of studies meeting WWC evidence standards without reservations

Cooke et al. (2004) examined the effect of *Reading Mastery* on the alphabetics achievement and reading comprehension of 30 students from grades 2–4 in three elementary schools in a suburban school district in the southeastern United States. The study involved three teachers and occurred over two years. One school participated both years, the second school participated in the first year, and the third school participated in the second year. Each school had two preexisting special education groups of three to five students, and within each school, these groups were randomly assigned to implement either *Reading Mastery Fast Cycle* or *Horizons Fast Track*. *Reading Mastery Fast Cycle* and *Horizons Fast Track* were each implemented in 30- to 40-minute sessions, five days a week, over one year. *Horizons* and *Reading Mastery* share a developer and many important design features, and *Horizons* was developed in response to feedback on *Reading Mastery*. The study was conducted to examine whether the limited substantive differences between the interventions led to different effects. The final sample included 15 students in each condition. Fifty percent of the students in the combined *Reading Mastery Fast Cycle* and *Horizons Fast Track* groups were identified as learning disabled. The remaining students had other disabilities, such as behavioral/emotional disabilities or other health impairments.

Herrera et al. (1997) examined the effect of *Reading Mastery* on the alphabetics achievement, reading fluency, and writing of 83 learning disabled students from 11 classrooms serving grades 3–5 in Florida's Orange County Public Schools. All students had previously been diagnosed as learning disabled and at a reading level at least three years below grade level. The 11 classrooms were randomly assigned to two intervention conditions—the *Reading Mastery Program* (six classrooms) or the *Stabilized Learning System* (five classrooms). Both the *Reading Mastery Program* and the *Stabilized Learning System* were implemented in 45-minute sessions, five days a week, during the 68-day study period.

Summary of studies meeting WWC evidence standards with reservations

No studies of *Reading Mastery* meet WWC evidence standards with reservations.

Effectiveness Summary

The WWC review of interventions for Students with Learning Disabilities addresses student outcomes in nine domains: alphabetics, reading fluency, reading comprehension, general reading achievement, mathematics, writing, science, social studies, and progressing in school. The two studies that contribute to the effectiveness rating in this report cover four domains: alphabetics, reading fluency, reading comprehension, and writing. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Reading Mastery* on students with learning disabilities. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 18.

Summary of effectiveness for the alphabetics domain

Two studies reported findings in the alphabetics domain.

Cooke et al. (2004) found no statistically significant effects of *Reading Mastery Fast Cycle* (when compared to *Horizons Fast Track*) on three measures of alphabetics: the Letter-Word Identification and Word Attack subtests of the Woodcock-Johnson Psycho-Educational Battery–Revised and the North Carolina Literacy Assessment. The WWC-calculated effect size for this measure was not large enough to be considered substantively important. The WWC characterizes these study findings as an indeterminate effect.

Herrera et al. (1997) found a statistically significant negative effect of the *Reading Mastery Program* (when compared to the *Stabilized Learning System*) on one measure of alphabetics: the Sight Vocabulary Accuracy subtest of the Orange County Curriculum Based Assessment. WWC calculations that were corrected for clustering do not confirm the statistical significance of this finding. However, the WWC-calculated effect size for this measure was large enough (–0.33) to be considered substantively important. The WWC characterizes these study findings as a substantively important negative effect.

Thus, for the alphabetics domain, one study found an indeterminate effect and one study found a substantively important negative effect. This results in a rating of potentially negative 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 negative effects *Evidence of a negative effect with no overriding contrary evidence.*	The review of *Reading Mastery* in the alphabetics domain had one study showing a substantively important negative effect and no studies showing a statistically significant or substantively important positive effect.
Extent of evidence	**Criteria met**
Small	The review of *Reading Mastery* in the alphabetics domain was based on two studies that included at least 11 classrooms and 113 students.

Table Note: Although there were possibly more than 14 classrooms across studies, the extent of evidence is small because the classrooms, on average, had fewer than 25 students. While there were 11 classrooms in Herrera et al. (1997), the number of classrooms is not reported in Cooke et al. (2004), and thus the total number of classrooms cannot be calculated.

Summary of effectiveness for the reading fluency domain

One study reported findings in the reading fluency domain.

Herrera et al. (1997) found a statistically significant negative effect of the *Reading Mastery Program* (when compared to the *Stabilized Learning System*) on three measures of reading fluency: the Sight Vocabulary Fluency subtest, the Familiar Passage Reading subtest, and the Unfamiliar Passage Reading subtest of the Orange County Curriculum Based Assessment. WWC calculations that were corrected for clustering and multiple comparisons do not confirm the statistical significance of these three findings. However, the WWC-calculated effect size across the three measures was large enough (–0.31) to be considered substantively important. The WWC characterizes these study findings as a substantively important negative effect.

Thus, for the reading fluency domain, one study found a substantively important negative effect. This results in a rating of potentially negative 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 negative effects *Evidence of a negative effect with no overriding contrary evidence.*	The review of *Reading Mastery* in the reading fluency domain had one study showing a substantively important negative effect and no studies showing a statistically significant or substantively important positive effect.
Extent of evidence	**Criteria met**
Small	The review of *Reading Mastery* in the reading fluency domain was based on one study that included 11 classrooms and 83 students.

Summary of effectiveness for the reading comprehension domain

One study reported findings in the reading comprehension domain.

Cooke et al. (2004) found no statistically significant effects of *Reading Mastery Fast Cycle* (when compared to *Horizons Fast Track*) on one measure of reading comprehension: the Passage Comprehension subtest of the Woodcock-Johnson Psycho-Educational Battery–Revised. The WWC-calculated effect size for this measure was not large enough to be considered substantively important. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the reading comprehension domain, no studies found statistically significant or substantively important effects. This results in a rating of no discernible effects, with a small extent of evidence.

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

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	The review of *Reading Mastery* in the reading comprehension domain had no studies showing a statistically significant or substantively important effect, either positive or negative.
Extent of evidence	**Criteria met**
Small	The review of *Reading Mastery* in the reading comprehension domain was based on one study that included an unknown number of classrooms and 30 students.

Table Note: Although there were possibly more than 14 classrooms across studies, the extent of evidence is small because the classrooms, on average, had fewer than 25 students. While there were 11 classrooms in Herrera et al. (1997), the number of classrooms is not reported in Cooke et al. (2004), and thus the total number of classrooms cannot be calculated.

Summary of effectiveness for the writing domain

One study reported findings in the writing domain.

Herrera et al. (1997) found a statistically significant negative effect of the *Reading Mastery Program* (when compared to the *Stabilized Learning System*) on one measure of writing: the Writing subtest of the Orange County Curriculum Based Assessment. WWC calculations that were corrected for clustering confirmed this finding. The WWC characterizes these study findings as a statistically significant negative effect.

Thus, for the writing domain, one study found a statistically significant negative effect. This results in a rating of potentially negative effects, with a small extent of evidence.

Table 6. Rating of effectiveness and extent of evidence for the writing domain

Rating of effectiveness	Criteria met
Potentially negative effects *Evidence of a negative effect with no overriding contrary evidence.*	The review of *Reading Mastery* in the writing domain had one study showing a statistically significant negative effect and no studies showing a statistically significant or substantively important positive effect.
Extent of evidence	**Criteria met**
Small	The review of *Reading Mastery* in the writing domain was based on one study that included 11 classrooms and 83 students.

References

Studies that meet WWC evidence standards without reservations

Cooke, N. L., Gibbs, S. L., Campbell, M. L., & Shalvis, S. L. (2004). A comparison of Reading Mastery Fast Cycle and Horizons Fast Track A-B on the reading achievement of students with mild disabilities. *Journal of Direct Instruction, 4*(2), 139.

Herrera, J. A., Logan, C. H., Cooker, P. G., Morris, D. P., & Lyman, D. E. (1997). Phonological awareness and phonetic-graphic conversion: A study of the effects of two intervention paradigms with learning disabled children. Learning disability or learning difference? *Reading Improvement, 34*(2), 71.

Study that does not meet WWC evidence standards

Earheart, L. S. (2002). The efficacy of the SRA reading program for disabled learners as measured by the Terra Nova achievement test. *Dissertation Abstracts International, 63*(08A), 57-2823. (AAI3061780). 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.

Studies that are ineligible for review using the Students with Learning Disabilities Evidence Review Protocol

Butler, M. T. (2002). Comparison of the effects of direct instruction and basal instruction on the reading achievement of first-grade students identified as students with reading difficulties. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 62*(9-A), 3002. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% students with learning disabilities.

Fitzpatrick, E., McLaughlin, T. F., & Weber, K. P. (2004). The effects of a first day and second day reads on reading accuracy with Reading Mastery III Textbook B for a fifth grade student with learning disabilities. *International Journal of Special Education, 19*(1), 56–63. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Frankhauser, M. A., Tso, M. E., & Martella, R. C. (2001). A comparison of curriculum-specified reading checkout timings and daily 1-minute timings on student performance in reading mastery. *Journal of Direct Instruction, 1*(2), 85. The study is ineligible for review because it does not test the effectiveness of *Reading Mastery*.

Kanfush, P. M., III. (2010). *Use of direct instruction to teach reading to students with significant cognitive impairments: Student outcomes and teacher perceptions*. Ann Arbor, MI: ProQuest, LLC. The study is ineligible for review because the WWC could not confirm that at least 50% of the sample was classified as students with learning disabilities.

Kinder, D., Kubina, R., & Marchand-Martella, N. (2005). Special education and direct instruction: An effective combination. *Journal of Direct Instruction, 5*(1), 1–36. 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.

Lovett, M. W., Steinbach, K. A., & Frijters, J. C. (2000). Remediating the core deficits of developmental reading disability: A double-deficit perspective. *Journal of Learning Disabilities, 33*(4), 334. The study is ineligible for review because it does not use a comparison group design or a single-case design.

O'Connor, R. E., Jenkins, J. R., Cole, K. N., & Mills, P. E. (1993). Two approaches to reading instruction with children with disabilities: Does program design make a difference? *Exceptional Children, 59*(4), 312–323. The study is ineligible for review because the WWC could not confirm that at least 50% of the sample was classified as students with learning disabilities.

Additional source:

O'Connor, R. E., Jenkins, J. R., Cole, K. N., & Mills, P. E. (1992). *Two approaches to reading instruction for children with disabilities: Does program design make a difference?* Unpublished manuscript.

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: 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.

Schieffer, C., Marchand-Martella, N. E., Martella, R. C., Simonsen, F. L., & Waldron-Soler, K. M. (2002). An analysis of the Reading Mastery program: Effective components and research review. *Journal of Direct Instruction, 2*(2), 87–119. 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.

SRA/McGraw-Hill. (2006) *Exceptional education and regular education students excel with direct instruction*. Retrieved from: http://www.sraonline.com/download/DI/EfficacyReports/Iredell_DI_FNL.pdf. The study is ineligible for review because it does not use a comparison group design or a single-case design.

SRA/McGraw-Hill. (2006). *Reading Mastery, corrective reading help students with disabilities achieve significant academic growth*. DeSoto, TX: Author. The study is ineligible for review because it does not use a comparison group design or a single-case design.

SRA/McGraw-Hill. (2009). *A report on the effects of SRA/McGraw-Hill's Reading Mastery, Signature Edition: A response to intervention solution*. DeSoto, TX: Author. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% students with learning disabilities.

Stewart, R. M., Benner, G. J., Martella, R. C., & Marchand-Martella, N. E. (2007). Three-tier models of reading and behavior: A research review. *Journal of Positive Behavior Interventions, 9*(4), 239–253. 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.

Swanson, H. L. (2011). Learning disabilities: Assessment, identification, and treatment. In M. A. Bray & T. J. Kehle (Eds.), *The Oxford handbook of school psychology* (pp. 334–350). New York: Oxford University 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.

Appendix A.1: Research details for Cooke et al., 2004

Cooke, N. L., Gibbs, S. L., Campbell, M. L., & Shalvis, S. L. (2004). A comparison of Reading Mastery Fast Cycle and Horizons Fast Track A-B on the reading achievement of students with mild disabilities. *Journal of Direct Instruction, 4*(2), 139.

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	3 schools/30 students	+2	No
Reading comprehension	3 schools/30 students	–1	No

Setting

The study was conducted in three schools in a suburban district in the southeastern United States.

Study sample

The sample for this study included a total of 30 students from grades 2–4 taught by three teachers in three elementary schools. All students in the study had been identified by school district staff as needing special education services. The study occurred over two years—one school participated in both years, the second school participated in the first year, and the third school participated in the second year. Prior to the start of the study, two groups of three to five students had been formed in each school. Within schools, the student groups were randomly assigned to receive either *Reading Mastery Fast Cycle* or *Horizons Fast Track*, resulting in 15 students receiving each intervention. In total, there were 15 students identified as learning disabled—ten in the intervention group and five in the comparison group. The remaining 15 students had other disabilities, such as behavioral/emotional disabilities or other health impairments. At each school, one teacher delivered both the *Reading Mastery Fast Cycle* and *Horizons Fast Track* interventions. The authors reported no group or student attrition.

Intervention group

Reading Mastery Fast Cycle is a version of *Reading Mastery* that teaches at a faster rate with less repetition than conventional *Reading Mastery*. In the present study, *Reading Mastery Fast Cycle* was implemented in 30- to 40-minute sessions, five days a week, over one school year.

Comparison group

Horizons Fast Track shares the same developer and many program characteristics with *Reading Mastery Fast Cycle* and was developed in response to feedback on *Reading Mastery*. The two programs differ in sequence, procedures, prompts, orthographic conventions, and teacher presentation materials. For example, *Reading Mastery Fast Cycle* teaches letter sounds before letter names, whereas *Horizons Fast Track* requires students to use letter names as assistance in learning letter sounds. *Reading Mastery Fast Cycle* does not use capital letters early in the program; *Horizons Fast Track* includes the use of capital letters in the first lessons that present sentences. Finally, *Reading Mastery Fast Cycle* uses special forms of letters to elicit the correct sounds for confusing letters, letter combinations, or silent letters; *Horizons Fast Track* uses underlining and color changes. Teachers implemented *Horizons Fast Track* in 30- to 40-minute sessions, five days a week over the year, following the scripted procedure and repeating lessons when necessary.

Outcomes and measurement

The study authors administered several reading measures at pretest and posttest. Alphabetics was measured by the Letter-Word Identification and Word Attack subtests of the Woodcock-Johnson Psycho-Educational Battery–Revised and the North Carolina Literacy Assessment. Reading comprehension was measured by the Passage Comprehension subtest of the Woodcock-Johnson Psycho-Educational Battery–Revised. The authors combined Letter-Word Identification and Passage Comprehension to form a Broad Reading Score and combined Letter-Word Identification and Word Attack to form a Basic Reading Score. These combined measures were not examined in the WWC analysis. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Prior to starting the study, teachers had been trained in *Reading Mastery Fast Cycle* by SRA/McGraw-Hill (and had four years experience with the program). Teachers were trained to implement *Horizons Fast Track* by SRA/McGraw-Hill prior to the start of the school year.

Appendix A.2: Research details for Herrera et al., 1997

Herrera, J. A., Logan, C. H., Cooker, P. G., Morris, D. P., & Lyman, D. E. (1997). Phonological awareness and phonetic-graphic conversion: A study of the effects of two intervention paradigms with learning disabled children. Learning disability or learning difference? *Reading Improvement, 34*(2), 71–86.

Table A2. Summary of findings **Meets WWC evidence standards without reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Alphabetics	11 classrooms/83 students	–13	No
Reading fluency	11 classrooms/83 students	–12	No
Writing	11 classrooms/83 students	–25	No

Setting

The study was conducted in 11 classrooms serving grades 3–5 in Florida's Orange County Public Schools.

Study sample

The sample for this study included a total of 83 learning disabled students from 11 classrooms serving grades 3–5. All students had previously been assessed and diagnosed as learning disabled. The 11 classrooms were randomly assigned either to the *Reading Mastery Program* (six classrooms) or the *Stabilized Learning System* (five classrooms). (The authors did not report the number of students in each condition; the WWC estimated an average class size of 7.5 and multiplied this by the number of classes in each condition to estimate the number of students in each condition.) Although the classrooms were randomly assigned, large baseline differences favored the *Reading Mastery Program* group on each of the five outcome measures (ranging from 0.09 to 0.78 standard deviations). There was no attrition of classrooms; the study provides no information on student-level attrition.

Intervention group

The *Reading Mastery Program* used an explicit learning approach that involved teaching specific thinking strategies and memorization of facts and rules. The intervention was implemented in 45-minute sessions, five days a week, during the 68-day study period. While the *Reading Mastery Program* was the business-as-usual reading curriculum for the school district and the comparison condition for the study authors, it was considered the intervention for this WWC review.

Comparison group

The *Stabilized Learning System* used an implicit learning approach that involved reading activities where students' primary focus is on the activity and not a cognitive analysis of the reading process. This intervention was implemented in 45-minute sessions, five days a week, during the 68-day study period. While the *Stabilized Learning System* was the intervention of interest for the study authors, it was considered the comparison condition for this WWC review

Outcomes and measurement

The Orange County Curriculum Based Assessment (OCCBA) was administered at pretest (prior to the start of the intervention) and posttest (by 15 days after the conclusion of the intervention). Alphabetics was assessed by the Sight Vocabulary Accuracy subtest of the OCCBA. Reading fluency was assessed by the Sight Vocabulary Fluency subtest, the Familiar Passage Reading subtest, and the Unfamiliar Passage Reading subtest of the OCCBA. Writing was assessed by the Writing Vocabulary subtest of the OCCBA. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

The *Reading Mastery Program* was the business-as-usual curriculum in Orange County, and all teachers had received relevant training. Teachers who implemented the *Stabilized Learning System* received 60 hours of training as part of this study. Individuals who administered the pretest and posttest assessments were trained and certified by the Orange County Public Schools. Teachers were not aware of the design or purpose of the study.

Appendix B: Outcome measures for each domain

Alphabetics	
Letter-word identification construct	
Woodcock-Johnson Psycho-Educational Battery–Revised (WJ-R): Letter-Word Identification subtest	The WJ-R is a nationally-standardized individually-administered battery of cognitive and achievement tests. The Letter-Word Identification subtest measures basic word-reading skills and requires the student to read aloud isolated words that range in frequency and difficulty. The reliability for the Letter-Word Identification subtest is greater than 0.93 (as cited in Cooke et al., 2004).
Orange County Curriculum Based Assessment (OCCBA): Sight Vocabulary Accuracy subtest	The OCCBA is used by the Orange County Public Schools and other independent school districts in Florida. The source for Sight Vocabulary Accuracy (word recognition) was Hillerich (1988).[5] Sight Vocabulary Accuracy was assessed by having students read starter, primary, and content-based lists of words. Scores were recorded for the number of words read correctly (as cited in Herrera et al., 1997).
Word attack construct	
WJ-R: Word Attack subtest	The WJ-R is a nationally-standardized individually-administered battery of cognitive and achievement tests. The Word Attack subtest measures the student's ability to apply phonic and structural analysis skills to pronounce unfamiliar words. Phonemic decoding skills are measured by asking students to read pseudo words. Students are aware that the words are not real. The reliability for the Word Attack subtest is greater than 0.87 (as cited in Cooke et al., 2004).
Multiple constructs	
North Carolina Literacy Assessment	The North Carolina Literacy Assessment for grades K–2 has several sections that measure letter and sound identification, book and print awareness, phonemic awareness, fluency, oral retelling, writing about reading, spelling, and writing (as cited in Cooke et al., 2004).
Reading fluency	
OCCBA: Sight Vocabulary Fluency subtest	The OCCBA is used by the Orange County Public Schools and other independent school districts throughout Florida. Sight Vocabulary Fluency was assessed by having the students read starter, primary, and content-based lists of words. Scores were based on words read correctly per minute relative to errors made (as cited in Herrera et al., 1997).
OCCBA: Familiar Passage Reading subtest	The OCCBA is used by the Orange County Public Schools and other independent school districts throughout Florida. Familiar Passage Reading was assessed in terms of words read correctly per minute relative to errors made while reading familiar passages (as cited in Herrera et al., 1997).
OCCBA: Unfamiliar Passage Reading subtest	The OCCBA is used by the Orange County Public Schools and other independent school districts throughout Florida. Unfamiliar Passage Reading was assessed in terms of words read correctly per minute relative to errors made while reading unfamiliar passages (as cited in Herrera et al., 1997).
Reading comprehension	
Reading and listening comprehension construct	
WJ-R: Passage Comprehension subtest	The WJ-R is a nationally-standardized individually-administered battery of cognitive and achievement tests. The Passage Comprehension subtest is a measure of reading comprehension at the sentence level that uses a cloze procedure. Students read a sentence or short passage and supply missing words based on the overall context. Reliability ranges from 0.87 to 0.97 (as cited in Cooke et al., 2004).
Writing	
OCCBA: Writing Vocabulary subtest	The OCCBA is used by the Orange County Public Schools and other independent school districts throughout Florida. The source for Writing Vocabulary test content was Hillerich (1978).[6] The score for the Writing Vocabulary subtest is the number of words spelled correctly (as cited in Herrera et al., 1997).

Appendix C.1: Findings included in the rating for the alphabetics domain

Outcome measure	Study sample	Sample size	Mean (standard deviation) Intervention group	Mean (standard deviation) Comparison group	WWC calculations Mean difference	WWC calculations Effect size	WWC calculations Improvement index	*p*-value
Cooke et al., 2004[a]								
WJ-R Letter-Word Identification subtest	Grades 2–4	3 schools/ 30 students	82.30 (18.36)	80.00 (14.82)	2.30	0.13	+5	0.31
WJ-R Word Attack subtest	Grades 2–4	3 schools/ 30 students	85.17 (13.49)	82.64 (15.71)	2.53	0.17	+7	0.32
North Carolina Literacy Assessment	Grades 2–4	3 schools/ 30 students	40.00 (8.28)	41.43 (9.83)	–1.43	–0.15	–6	0.50
Domain average for alphabetics (Cooke et al., 2004)						**0.05**	**+2**	**Not statistically significant**
Herrera et al., 1997[b]								
OCCBA Sight Vocabulary Accuracy subtest	Grades 3–5	11 classrooms/ 83 students	102.86 (51.20)	119.50 (48.96)	–16.64	–0.33	–13	0.02
Domain average for alphabetics (Herrera et al., 1997)						**–0.33**	**–13**	**Not statistically significant**
Domain average for alphabetics across all studies						**–0.13**	**–5**	**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 change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. 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. WJ-R = Woodcock-Johnson Psycho-Educational Battery–Revised. OCCBA = Orange County Curriculum Based Assessment.

[a] For Cooke et al. (2004), the *p*-value was reported by the study authors. Corrections for clustering and multiple comparisons were needed but did not affect significance levels. The WWC calculated the intervention group mean using a difference-in-differences approach (see the *WWC Handbook*) by adding the impact of the program (difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. The WWC calculated and reported effect sizes using Hedges' *g* rather than Cohen's *d* which is reported by the study authors.

[b] For Herrera et al. (1997), the *p*-value was reported by the study authors. In this study, means and standard deviations are presented in a table that states the means are "adjusted." If the reported means were actually adjusted, that would indicate that the *Reading Mastery Program* had a positive effect on most outcomes. However, the study text indicates that the *Reading Mastery Program* had a negative effect on all outcomes. The large pretest differences favoring the *Reading Mastery Program* imply that the reported means are actually unadjusted, and the WWC assumes that the means are unadjusted. The study authors did not reply to WWC contacts. Corrections for clustering (using the reported means) and multiple comparisons were needed and resulted in significance levels that differ from those in the original study. The WWC calculated the intervention group mean using a difference-in-differences approach (see the *WWC Handbook*) by adding the impact of the program (difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means.

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	
Herrera et al., 1997[a]								
OCCBA Sight Vocabulary Fluency subtest	Grades 3–5	11 classrooms/ 83 students	31.48 (26.21)	41.63 (21.18)	–10.15	–0.42	–16	0.02
OCCBA Familiar Passage Reading subtest	Grades 3–5	11 classrooms/ 83 students	44.67 (36.70)	52.93 (21.09)	–8.26	–0.27	–11	0.08
OCCBA Unfamiliar Passage Reading subtest	Grades 3–5	11 classrooms/ 83 students	34.67 (30.64)	41.13 (21.68)	–6.46	–0.24	–9	0.08
Domain average for reading fluency (Herrera et al., 1997)						**–0.31**	**–12**	**Not statistically significant**
Domain average for reading fluency across all studies						**–0.31**	**–12**	**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 change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. 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 the study's domain average was determined by the WWC. na = not applicable. OCCBA = Orange County Curriculum Based Assessment.

[a] For Herrera et al. (1997), the *p*-value was reported by the study authors. In this study, means and standard deviations are presented in a table that states the means are "adjusted." If the reported means were actually adjusted, that would indicate that the *Reading Mastery Program* had a positive effect on most outcomes. However, the study text indicates that the *Reading Mastery Program* had a negative effect on all outcomes. The large pretest differences favoring the *Reading Mastery Program* imply that the reported means are actually unadjusted, and the WWC assumes that the means are unadjusted. The study authors did not reply to WWC contacts. Corrections for clustering (using the reported means) and multiple comparisons were needed and resulted in significance levels that differ from those in the original study. The WWC calculated the intervention group mean using a difference-in-differences approach (see the *WWC Handbook*) by adding the impact of the program (difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means.

Appendix C.3: Findings included in the rating for the reading 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	
Cooke et al., 2004[a]								
WJ-R Passage Comprehension subtest	Grades 2–4	3 schools/ 30 students	85.29 (21.96)	85.57 (18.27)	–0.28	–0.01	–1	0.92
Domain average for reading comprehension (Cooke et al., 2004)						**–0.01**	**–1**	**Not statistically significant**
Domain average for reading comprehension across all studies						**–0.01**	**–1**	**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 change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. 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 the study's domain average was determined by the WWC. na = not applicable. WJ-R = Woodcock-Johnson Psycho-Educational Battery–Revised.

[a] For Cooke et al. (2004), the *p*-value was reported by the study authors. A correction for clustering was needed but did not affect significance levels. The WWC calculated the intervention group mean using a difference-in-differences approach (see the *WWC Handbook)* by adding the impact of the program (difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. The WWC calculated and reported effect sizes using Hedges' *g* rather than Cohen's *d* (as used by the study authors).

Appendix C.4: Findings included in the rating for the writing domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Herrera et al., 1997[a]								
OCCBA Writing Vocabulary subtest	Grades 3–5	11 classrooms/ 83 students	44.90 (22.71)	60.53 (22.23)	–15.63	–0.69	–25	0.01
Domain average for writing (Herrera et al., 1997)						**–0.69**	**–25**	**Statistically significant**
Domain average for writing across all studies						**–0.69**	**–25**	**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 change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. 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 the study's domain average was determined by the WWC. na = not applicable. OCCBA = Orange County Curriculum Based Assessment.

[a] For Herrera et al. (1997), the *p*-value was reported by the study authors. In this study, means and standard deviations are presented in a table that states the means are "adjusted." If the reported means were actually adjusted, that would indicate that the *Reading Mastery Program* had a positive effect on most outcomes. However, the study text indicates that the *Reading Mastery Program* had a negative effect on all outcomes. The large pretest differences favoring the *Reading Mastery Program* imply that the reported means are actually unadjusted, and the WWC assumes that the means are unadjusted. The study authors did not reply to WWC contacts. Corrections for clustering (using the reported means) and multiple comparisons were needed and resulted in significance levels that differ from those in the original study. The WWC calculated the intervention group mean using a difference-in-differences approach (see the *WWC Handbook*) by adding the impact of the program (difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means.

Endnotes

[1] The descriptive information for this program was obtained from a publicly available source: the program's website (https://www.mheonline.com/discipline/tags/1/3/, downloaded June 2011; updated January 2012). The WWC requests developers to review the program description sections for accuracy from their perspective. The program description was provided to the developer in July 2011; however, the WWC received no response. 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 August 2011.

[2] The studies in this report were reviewed using WWC Evidence Standards, version 2.1, as described in the Students with Learning Disabilities review 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. 18. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

[4] The two studies reviewed in this report use different nomenclature when referring to *Reading Mastery*. Cooke et al. (2004) refer to the "*Reading Mastery Fast Cycle*" program, while Herrera et al. (1997) refer to the "*Reading Mastery Program*". We use these program labels throughout the report when discussing the two studies. All discussion that is not specific to one of these studies uses the generic label of *Reading Mastery*.

[5] Hillerich, R. (1988). A little dab'l do ya? *Reading Central*. January 23.

[6] Hillerich, R. (1978). *A writing vocabulary of elementary children*. Springfield, IL: Charles C. Thomas.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2012, July). *Students with Learning Disabilities intervention report: Reading Mastery*. 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. 18.

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. 18.

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.