

# Appendix

## Appendix A1.1 Study characteristics: Stockard, 2010

Characteristic	Description
<b>Study citation</b>	Stockard, J. (2010). <i>Fourth graders' growth in reading fluency: A pretest-posttest randomized control study comparing Reading Mastery and Scott Foresman Basal Reading Program</i> . Eugene, OR: National Institute for Direct Instruction.
<b>Participants</b>	In this randomized study, 58 general education elementary students were assigned to a treatment or control condition, using alternative assignment with a random start technique. The author used a class list (arranged in alphabetical order) to conduct the assignment. First, the author used a random numbers table to determine where to begin in the class list. Second, the author used a coin flip to determine whether the assignment would start with the treatment or control group. Finally, after this initial assignment was determined, the author assigned each additional student in an alternating fashion to the treatment or control group. For example, in the first step, the author might have started the assignment with John Smith (based on a random numbers table). In the second step, the author might have assigned John Smith to the treatment group (based on the coin flip). Then, the author would have proceeded through the rest of the ordered class list (alternating between the control and treatment groups). Four classrooms participated in the study. Two pairs of teachers were formed and then—within these pairs—teachers were randomly assigned to the treatment or control group via a coin flip. <sup>1</sup> Students were predominantly non-Hispanic whites from middle-income families. The analysis sample consisted of 29 fourth-grade students who received <i>Reading Mastery</i> and 28 fourth-grade students in the comparison group.
<b>Setting</b>	The study was conducted in a midwestern elementary school.
<b>Intervention</b>	Beginning in the fall of 2009, students in the treatment condition received instruction for 90 minutes a day in the SRA/McGraw-Hill program, <i>Reading Mastery Signature Edition</i> . Students were exposed to <i>Reading Mastery</i> over five months.
<b>Comparison</b>	The control group received instruction for 90 minutes a day in the <i>Scott Foresman Basal Reading Program</i> , which the school had been using in prior years.
<b>Primary outcomes and measurement</b>	Data on the AIMS Web Curriculum-Based Measurement Words Read Correct were gathered in the spring of 2009 before instruction began (for use as a baseline measure), in the fall of 2009 shortly after the start of the school year, and again in winter of 2010, approximately halfway through the school year. For a more detailed description of this outcome measure, see Appendix A2.1.
<b>Staff or teacher training</b>	No information about training was provided.

1. The study author did not describe how students assigned to the treatment group were assigned to the two treatment teachers or how students assigned to the control group were assigned to the two control teachers.

## Appendix A1.2 Study characteristics: Yu & Rachor, 2000

Characteristic	Description
<b>Study citation</b>	Yu, L., & Rachor, R. (2000, April). <i>The two-year evaluation of the three-year Direct Instruction program, in an urban public school system</i> . Presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
<b>Participants</b>	This retrospective quasi-experiment included students in three elementary schools that participated in <i>Reading Mastery</i> for two consecutive years. Each of the three <i>Reading Mastery</i> schools was matched with a school with a similar level of poverty and percentage of minority students. Then, <i>Reading Mastery</i> students were matched with comparison students in the same grade on the basis of race, gender, free lunch status, and reading achievement test scores. This review focuses on findings from students who were in grades 4 and 5 <sup>1</sup> during the first year of program implementation (1997/98). The analysis sample for students who were in grade 4 in 1997/98 consisted of 71 students who received <i>Reading Mastery</i> and 71 matched comparison students. The analysis sample for students who were in grade 5 in 1997/98 consisted of 81 students in the <i>Reading Mastery</i> group and 81 matched comparison students. Of the students, more than 96% were African-American, and about 80% were eligible for free or reduced-price lunch. The study reported students' outcomes after two years of program implementation; these findings were used in the intervention ratings and can be found in Appendix A3.2. Additional findings reflecting students' outcomes after one year of program implementation can be found in Appendix A4.
<b>Setting</b>	The study was conducted in six elementary schools in a northwestern urban public school system.
<b>Intervention</b>	The complete <i>Reading Mastery</i> program was adopted, including all materials and teacher training. The program provided scripted, carefully sequenced lessons; rapid pacing; and responses of the students in unison as well as in individual turns. Students were exposed to <i>Reading Mastery</i> over two school years.
<b>Comparison</b>	The comparison group received the standard instruction provided in the regular school curriculum.
<b>Primary outcomes and measurement</b>	For both the pretest and posttest, students took the State Reading Proficiency Test and the Riverside Publishing Off Grade Reading Proficiency Test. The State Reading Proficiency Test provided data at pretest and after two years of intervention implementation for the 5th-grade group and data after one year of intervention implementation for the 4th-grade group. The Riverside Publishing Off Grade Reading Proficiency Test provided data at pretest and after two years of intervention implementation for the 4th-grade group, and data after one year of intervention implementation for the 5th-grade group. For a more detailed description of these outcome measures, see Appendix A2.2.
<b>Staff or teacher training</b>	Teachers participated in training prior to implementation and were provided with ongoing consultations from the provider for the program duration. No additional details about training were provided.

1. Findings for 3rd grade are outside the scope of the Adolescent Literacy review.

## Appendix A2.1 Outcome measures for the reading fluency domain

Outcome measure	Description
<b>AIMS Web Curriculum-Based Measurement Words Read Correct</b>	When a child reads graded passages aloud for one minute, this curriculum-based AIMS web assessment gathers data on reading fluency in the form of number of words read correctly. The assessment has well-established validity and reliability. AIMS Web is a benchmark and progress-monitoring system based on direct, frequent, and continuous student assessment (as cited in Stockard, 2010).

## Appendix A2.2 Outcome measures for the comprehension domain

Outcome measure	Description
<i>Reading comprehension construct</i>	
<b>Riverside Publishing Off Grade Reading Proficiency Test</b>	This test for 3rd- and 5th-grade students is designed to measure students' reading proficiency. Test items, which are based on fiction, nonfiction, and poetry reading passages, are designed to capture four strands of learning outcomes defined by the publisher as (1) examining meaning given a fiction or poetry text (examining meaning indicates that students are able to comprehend the overall meaning of what they have read), (2) extending meaning given a fiction or poetry text (extending meaning indicates that students can interpret what they have read and infer beyond the text), (3) examining meaning given a nonfiction text, and (4) extending meaning given a nonfiction text (as cited in Yu & Rachor, 2000).
<b>State Reading Proficiency Test</b>	This state annual reading test is administered to 4th- and 6th-grade students (as cited in Yu & Rachor, 2000). No additional information on the test was provided.

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

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> ( <i>Reading Mastery</i> – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Reading Mastery</i> group	Comparison group				
<b>Stockard, 2010<sup>7,8</sup></b>								
AIMS Web CBM Words Read Correct	Grade 4	4/57	149.0 (29.5)	134.7 (27.8)	14.30	0.49	ns	+19
<b>Domain average for reading fluency (Stockard, 2010)<sup>9</sup></b>						<b>0.49</b>	<b>ns</b>	<b>+19</b>

ns = not statistically significant

AIMS Web CBM = AIMS Web Curriculum-Based Measurement

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the reading fluency domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect-size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The *Reading Mastery* group mean outcome values for Stockard (2010) are the unadjusted control group posttest means plus the difference in mean gains between the intervention and control groups. Control group means are unadjusted.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Stockard (2010), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A3.2 Summary of study findings included in the rating for the comprehension domain<sup>1</sup>

Outcome measure	Study sample	Sample size (schools/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> ( <i>Reading Mastery</i> –comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Reading Mastery</i> group	Comparison group				
<b>Yu &amp; Rachor, 2000<sup>7,8</sup></b>								
Riverside Publishing Off Grade Reading Proficiency Test	Grade 4	6/142	209.01 (20.16)	205.46 (18.42)	3.55	0.18	ns	+7
State Reading Proficiency Test	Grade 5	6/162	201.15 (25.30)	213.80 (23.42)	-12.65	-0.52	ns	-20
<b>Domain average for comprehension (Yu &amp; Rachor, 2000)<sup>9</sup></b>						<b>-0.17</b>	<b>na</b>	<b>-7</b>

ns = not statistically significant

na = not applicable

1. This appendix reports two-year findings considered for the effectiveness rating and the average improvement indices for the comprehension domain. One-year findings from Yu and Rachor (2000) are not included in these ratings but are reported in Appendix A4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect-size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The *Reading Mastery* group mean outcome values for Yu and Rachor (2000) differ from those presented in the paper. The WWC calculated the program group mean using a difference-in-differences approach (see WWC Handbook)—calculating the program means by adding the impact of the program (i.e., difference in mean gains between the intervention and control groups) to the unadjusted control group posttest means.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Yu and Rachor (2000), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A4 Summary of one-year implementation findings for the comprehension domain<sup>1</sup>

Outcome measure	Study sample	Sample size (schools/students)	Authors' findings from the study					
			Mean outcome (standard deviation) <sup>2</sup>		WWC calculations			
			Reading Mastery group	Comparison group	Mean difference <sup>3</sup> (Reading Mastery – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
Yu & Rachor, 2000 <sup>7,8</sup>								
State Reading Proficiency Test	Grade 4	6/142	207.13 (12.28)	200.92 (12.18)	6.21	0.51	ns	+19
Riverside Publishing Off Grade Reading Proficiency Test	Grade 5	6/162	207.42 (21.33)	204.95 (24.86)	2.47	0.11	ns	+4

ns = not statistically significant

1. This appendix presents one-year findings for measures that fall in the comprehension domain. Two-year findings were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect-size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The *Reading Mastery* group and control group mean outcome values for Yu and Rachor (2000) are the unadjusted one-year posttest means.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Yu and Rachor (2000), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

## Appendix A5.1 *Reading Mastery* rating for the reading fluency domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the reading fluency outcome domain, the WWC rated *Reading Mastery* as having potentially positive effects for adolescent learners. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, or negative effects) were not considered, as *Reading Mastery* was assigned the highest applicable rating.

### Rating received

**Potentially positive effects:** Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

**Met.** One study showed substantively important positive effects.

### AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Met.** No studies showed statistically significant or substantively important negative effects. No studies showed indeterminate effects.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No studies showed statistically significant positive effects.

### AND

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

**Met.** No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A5.2 Reading Mastery rating for the comprehension domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup> For the comprehension outcome domain, the WWC rated *Reading Mastery* as having no discernible effects for adolescent learners.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** No studies showed a statistically significant or substantively important effect, either positive or negative.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No studies showed a statistically significant positive effect.

### AND

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

**Met.** No studies showed a statistically significant or substantively important negative effect.

**Potentially positive effects:** Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

**Not met.** No studies showed a statistically significant or substantively important positive effect.

### AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** No studies showed a statistically significant or substantively important negative effect. One study showed indeterminate effects.

**Mixed effects:** Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

**Not met.** No studies showed a statistically significant or substantively important positive effect, and no studies showed a statistically significant or substantively important negative effect.

### OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

**Not met.** No studies showed a statistically significant or substantively important effect. One study showed indeterminate effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

(continued)



## Appendix A5.2 Reading Mastery rating for the comprehension domain (continued)

**Potentially negative effects:** Evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies showing a statistically significant or substantively important *positive* effect.

**Not met.** No studies showed a statistically significant or substantively important effect, either positive or negative.

**OR**

- Criterion 2: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Not met.** No studies showed a statistically significant or substantively important effect, either positive or negative.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a *strong* design.

**Not met.** No studies showed a statistically significant negative effect.

**AND**

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

**Met.** No studies showed a statistically significant or substantively important positive effect.

## Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence <sup>1</sup>
		Schools	Students	
Alphabetics	na	na	na	na
Reading fluency	1	1	57	Small
Comprehension	1	6	304	Small
General literacy achievement	na	na	na	na

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.” For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.