

Appendix

Appendix A1 Study characteristics: Gunn, Biglan, Smolkowski, & Ary, 2000 (randomized controlled trial)

Characteristic	Description
Study citation	Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and Non-Hispanic students in early elementary school. <i>The Journal of Special Education, 34</i> , 90–103.
Participants	The original study involved 156 students in grades K–3. Students in kindergarten, first, and second grades were assessed during the spring prior to beginning the first year of the intervention (Time 1), assessed again one year later (Time 2), and assessed a final time the following year (Time 3). Students were selected for participation in the study on the basis of low reading achievement and aggressive tendencies. Specifically, students who scored below grade level on reading assessments and high on aggression (as rated by teachers) were included in the study to examine the effect of supplemental reading instruction on students meeting these criteria. A post hoc analysis was conducted on a small portion of these students (n=17) who were English language learners and for whom pre- and posttest data were available (there were 19 of these students at the beginning of the study). All estimates of intervention effects are based on this subsample. The English language learners were included in the process of randomly assigning all participants (limited and fluent English proficient) to a condition. All students were grouped by ethnicity and then rank-ordered by reading ability. Participants were matched, beginning with poorest readers, and randomly assigned to a condition. That is, students from each pair were randomly assigned to the intervention or comparison condition.
Setting	The study was part of a larger evaluation of a program in nine elementary schools across three school districts in Oregon.
Intervention	The intervention group received their usual reading instruction supplemented by <i>Reading Mastery</i> if they were beginning readers in grades 1 or 2. ¹ Students below grade level in grades 3 or 4 were put into an appropriate level of <i>SRA Corrective Reading</i> . ² Both programs include components that facilitate the development of beginning reading skills, but the programs differ in instructional methodology. <i>Reading Mastery</i> and <i>Corrective Reading</i> both entail explicit instruction in phonemic awareness, sound-letter correspondence, and blending. New sounds were introduced to students assigned to the <i>Corrective Reading</i> group at a faster pace than to students in the <i>Reading Mastery</i> group, and stories used for the <i>Corrective Reading</i> group were selected based on their appeal to older students. Relative to English speaking peers, English language learning students were provided additional time per lesson if assistants needed to explain English vocabulary. Most instruction was conducted in groups of two to three students, though some one-to-one instruction was provided. The program was delivered as a pull-out lasting 25–30 minutes a day.
Comparison	The comparison group of English language learning students had the same regular reading instruction but did not participate in the supplemental instruction programs.
Primary outcomes and measurement	A series of reading subtests from Woodcock-Johnson were administered four times in the course of the two-year intervention. (See Appendix A2 for more detailed descriptions of outcome measures.) Outcomes reported here are drawn from the spring of the second year (that is, after two years of the intervention; reported in Appendix A3). In addition, a follow-up assessment was conducted one year after the conclusion of the study. It is reported in Appendix A4.
Teacher training	Project assistants delivered the intervention to students, supplementing the normal reading instruction delivered by the classroom teacher. In all cases except one, instruction took place as a pull-out program. All assistants received 10 hours of preservice training in testing, student-grouping, general instructional skills, and the theoretical approach of the program. To ensure program delivery met program standards, assistants were observed weekly in the first month of the program and twice a month thereafter.

1. Students were in kindergarten, first, and second grades during Time 1 screening, prior to intervention implementation, so they were in first, second, and third grades at the start of the intervention year.
2. The English Language Learners subsample received instruction with *Reading Mastery* only. This was determined after corresponding with the first author of the study.

Appendix A2 Outcome measures in the reading achievement domain

Outcome measure	Description
Oral Reading Fluency	To calculate total number of words correctly read per minute, students read aloud three 1-minute grade-level reading samples. Mean scores were recorded; note that this measure is not a Woodcock-Johnson subtest (as cited in Gunn et al., 2000).
Woodcock-Johnson, Letter-Word Identification subtest	This is a standardized subtest from the Woodcock-Johnson Tests of Achievement that assessed a student's word reading skills. Students identified a list of letters and then read a list of words. Scores were available as raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, or grade-level equivalencies (as cited in Gunn et al., 2000).
Woodcock-Johnson, Word Attack subtest	This is a standardized subtest from the Woodcock-Johnson Tests of Achievement that is part of a broad reading cluster score. This subtest assessed the student's phonemic awareness skills. Students read a list of nonsense words. Scores were available as raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, or grade-level equivalencies (as cited in Gunn et al., 2000).
Woodcock-Johnson, Reading Vocabulary subtest	This is a standardized subtest from the Woodcock-Johnson Tests of Achievement that is part of a broad reading comprehension cluster score. This subtest assessed the student's overall skill at understanding text. Students were asked to identify antonyms, synonyms, and analogies. Scores were available as raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, or grade-level equivalencies (as cited in Gunn et al., 2000).
Woodcock-Johnson, Passage Comprehension subtest	This is a standardized subtest from the Woodcock-Johnson Tests of Achievement that is part of a broad reading comprehension cluster score. This subtest assesses the student's overall skill at understanding text. Students silently read a short passage and then filled in the missing word. Scores were available as raw scores, standard scores, Normal Curve Equivalent scores (NCES), age equivalencies, or grade-level equivalencies (as cited in Gunn et al., 2000).

Appendix A3 Summary of study findings included in the rating for the reading achievement domain¹

Outcome measure ³	Study sample	Sample size ⁴ (students)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁶ (Reading Mastery – comparison)	Effect size ⁷	Statistical significance ⁸ (at $\alpha = 0.05$)	Improvement index ⁹
			Reading Mastery group ⁵	Comparison group				
Gunn et al., 2000 (randomized controlled trial)								
Oral Reading Fluency	Grades K–3	16	51.75 (30.07)	24.92 (17.63)	26.83	1.03	ns	+35
Letter-Word Identification	Grades K–3	17	19.63 (12.21)	14.11 (4.81)	5.52	0.55	ns	+21
Word Attack	Grades K–3	17	11.63 (10.43)	5.33 (5.50)	6.30	0.70	ns	+26
Domain average¹⁰ for reading achievement						0.76		+28

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index. Follow-up findings from the same study 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. Reading Vocabulary and Passage Comprehension were not included because of severe overall attrition on these outcome measures.
4. Small sample sizes decrease the power of the analysis to accurately detect differences. The effects from a small number of participants can be magnified and so results may not reflect the likely effect of the program, given a larger sample. These results should be interpreted with caution.
5. The WWC requested and received means and standard deviations for the English language learner subgroup because they were not reported separately in the original paper.
6. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
7. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
8. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of *Reading Mastery*, a correction for multiple comparisons was needed.
9. 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.
10. This row provides the study average, which is also the domain average in this case. 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 subgroup findings for the reading achievement domain: Follow-up data one year after the conclusion of the intervention¹

Outcome measure	Study sample	Sample size ³ (students)	Author's findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
			Reading Mastery group ⁴	Comparison group	Mean difference ⁵ (Reading Mastery – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
Gunn et al., 2002 (randomized controlled trial)								
Oral Reading Fluency	Grades K–3	16	67.38 (32.24)	60.12 (24.40)	7.26	0.24	ns	+9
Letter-Word Identification	Grades K–3	17	33.88 (27.75)	24.11 (14.24)	9.77	0.43	ns	+17
Word Attack	Grades K–3	17	27.25 (25.56)	2.89 (19.81)	24.36	1.02	Statistically significant	+35
Reading Vocabulary	Grades K–3	17	22.88 (16.40)	12.44 (11.73)	10.44	0.70	ns	+26
Passage Comprehension	Grades K–3	16	34.13 (21.54)	23.38 (11.75)	10.75	0.59	ns	+22

ns = not statistically significant

1. This appendix presents follow-up findings for measures that fall in the reading achievement domain. Immediate posttest scores were used for rating purposes and are presented in Appendix A3.
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. Small sample sizes decrease the power of the analysis to accurately detect differences. The effects from a small number of participants can be magnified and so results may not reflect the likely effect of the program, given a larger sample. These results should be interpreted with caution.
4. The WWC requested and received means and standard deviations for the English language learner subgroup because they were not reported separately in the original paper. With the exception of Oral Reading Fluency, all outcomes for this table were reported as Normal Curve Equivalent scores.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of *Reading Mastery*, no corrections were needed.
8. 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.

Appendix A5 Reading Mastery rating for the reading achievement domain

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of reading achievement, the WWC rated *Reading Mastery* as having potentially positive effects. It did not meet the criteria for positive effects because it had only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *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 reviewed by the WWC reported an average effect size that is substantively important.

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

Met. The WWC analysis found no indeterminate, statistically significant negative, or substantively important negative effects in this domain.

Other ratings considered

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

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

Not met. Reading Mastery only one study meeting WWC evidence standards.

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

Met. The WWC analysis found no statistically significant or substantively important negative effects in this domain.

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 effects. See the [WWC Intervention Rating Scheme](#) for a complete description.