Appendix

Appendix A1.1 Study characteristics: Bullock, 2005

Characteristic	Description
Study citation	Bullock, J. C. (2005). Effects of the *Accelerated Reader* on the reading performance of third, fourth, and fifth-grade students in one western Oregon elementary school (Doctoral dissertation, University of Oregon). *Dissertation Abstracts International, 66*(07A), 56–2529.
Participants	The study examined students in grades 3 to 5. For this review, the WWC analysis focused on fourth and fifth graders, as specified in the Adolescent Literacy review protocol.[1] Ninety-one percent of the students in the study school were white, and 61% qualified for free or reduced-price lunch. The fourth-grade sample included 45 students from two classrooms, and the fifth-grade sample included 37 students from two classrooms. Within each classroom, students were rank ordered by baseline reading fluency scores and were divided into two groups based on whether their rank was an odd or even number. A coin flip decided the assignment of each group to intervention or control status.[2] There was no attrition of students or classrooms between pretest and posttest.
Setting	The study took place in one elementary school near Eugene in western Oregon.
Intervention	Students in the intervention group participated in the *Accelerated Reader*™ program over a 10-week period. These students were provided with a minimum of 90 minutes per week of independent reading time during class and were required to visit the library and check out a minimum of one book a week. Books had to be drawn from the subset of library books for which *Accelerated Reader*™ quizzes were available. When they finished a book, students completed a brief, computerized, multiple-choice quiz on the book's content and received points based on the level of the book read and the number of questions answered correctly. During the weekly library visit, intervention teachers and the library specialist verified that intervention students had access to appropriate *Accelerated Reader*™ books.
Comparison	The control condition relied on the business-as-usual reading program throughout the 10 week study, without the addition of *Accelerated Reader*™. As was the case for the intervention group, students in the control group were provided with a minimum of 90 minutes per week of independent reading time during class and 30 minutes per week of library time. Control students were asked to keep track of the books they read.
Primary outcomes and measurement	For both the pre- and posttest, students took the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency subtest; the Standardized Test and Assessment in Reading (STAR); and the 4J Vocabulary assessment.[3] For a more detailed description of these outcome measures, see Appendices A2.1–A2.2.
Staff/teacher training	The author does not describe the training provided to study teachers.

1. Grade 3 students are excluded from the review because they fall outside the grade range of the Adolescent Literacy topic area; they will be included in the *Accelerated Reader*™ intervention report for the Beginning Reading topic area.
2. The author of the study describes the design as quasi-experimental. However, because the groups were assigned randomly to the treatment and control conditions, the WWC classified the study as a randomized controlled trial.
3. Only results for fourth-grade students were available on the 4J Vocabulary assessment due to errors in data collection for grades 3 and 5.

Appendix A1.2 Study characteristics: Nunnery & Ross, 2007

Characteristic	Description
Study citation	Nunnery, J. A., & Ross, S. M. (2007). The effects of the School Renaissance program on student achievement in reading and mathematics. *Research in the Schools, 14*(1), 40–59.
Participants	The intervention group consisted of 11 schools that implemented *Accelerated Reader*™. Two steps were used to identify comparison schools. The first step was taken to narrow the pool of potential comparison schools. In this step, the researchers used data from the Academic Excellence Indicator System (AEIS), which identifies—for each school in Texas (including the 11 treatment schools in this study)—40 demographically similar schools based on the percentage of African-American students, Hispanic students, white students, economically disadvantaged students, limited English proficient students, and student mobility. In the second step, from the group of 40 potential comparison schools identified for each treatment school, the most similar school not using *Accelerated Reader*™ was selected according to the schools' base-year accountability rating (low performing, acceptable, recognized, exemplary) and base-year percentage of economically disadvantaged students. One of the selected comparison schools declined to participate, and another two did not have appropriate grade-level scores for use in the study. These three comparison schools were replaced from the pool of similar schools. The analytic sample consisted of students in grades 5 and 8 who had three consecutive years of data between school years 1998/99 and 2001/02 (cohort 1 students had data from the 1998/99 through 2000/01 school years, and cohort 2 students had data from the 1999/2000 through 2001/02 school years).[1] The cohort 1 grade 5 analysis sample included 442 intervention students from nine schools who received *Accelerated Reader*™ in the 1999/2000 and 2000/01 school years and 470 nonparticipants from nine matched elementary schools. The cohort 2 grade 5 analysis sample consisted of 437 students from nine schools who received *Accelerated Reader*™ in the 1999/2000, 2000/01, and 2001/02 school years and 454 nonparticipants from nine matched elementary schools. The cohort 2 grade 8 analysis sample consisted of 482 students in two schools who received *Accelerated Reader*™ in the 1999/2000, 2000/01, and 2001/02 school years and 510 nonparticipants from two matched middle/junior high schools. Outcomes were measured at the end of the second year of intervention implementation for cohort 1 and at the end of the third year of intervention implementation for cohort 2.
Setting	The study took place in 18 elementary and 4 middle/junior high schools from nine districts in Texas. All 11 intervention schools were located in a suburban school district.
Intervention	According to study authors, *Accelerated Reader*™ was the primary reading curriculum in intervention schools. The study did not provide details on how the intervention was implemented.
Comparison	The comparison schools did not implement *Accelerated Reader*™ during the school years under study. No information is available on the reading curricula used in these schools.
Primary outcomes and measurement	For both pre-[2] and posttests, the authors used the Texas Assessment of Academic Skills (TAAS), Reading subtest. For a more detailed description of this outcome measure, see Appendix A2.2.
Staff/teacher training	No information on staff or teacher training was provided in the study.

1. Cohort 1 also included grade 8 students. However, for this group of students, the intervention and comparison groups were not shown to be equivalent at baseline. Therefore, cohort 1 grade 8 students were excluded from the review.
2. Although the baseline period was the 1998/99 school year, the authors used reading test score data from the 1999/2000 school year as a covariate for cohort 2 students. Grade 5 students in cohort 2 were in second grade during the 1998/99 school year, and second grade scores were not available to the authors; therefore, third grade reading test score data from the 1999/2000 school year were used as a covariate. The authors did not report the reason that 1999/2000 reading test score data were used as a covariate for grade 8 cohort 2 students. Because the authors used reading test score data from the 1999/2000 school year as a covariate for cohort 2 students, the pretest data for this cohort may reflect some effect of the first year of program implementation.

Appendix A2.1 Outcome measures for the reading fluency domain

Outcome measure	Description
Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency subtest	The DIBELS Oral Reading Fluency measure is a standardized test of reading accuracy and speed, based on the number of words read correctly in one minute from connected text. Hesitations of more than three seconds, omitted words, and word substitutions are counted as errors, whereas prompt self-corrections are regarded as accurate (as cited in Bullock, 2005).

Appendix A2.2 Outcome measures for the comprehension domain

Outcome measure	Description
Vocabulary development construct	
4J Vocabulary	4J Vocabulary is a curriculum-based assessment[1] which consists of 90 vocabulary words selected from a list of words in *World Book*. Each of the words has three possible synonym answer choices: (1) the correct response, (2) one near-response, and (3) one far-response. Items were field tested and normed with oral reading fluency measures. This measure was administered at the beginning and end of the 10-week study (as cited in Bullock, 2005).
Reading comprehension construct	
Standardized Test and Assessment in Reading (STAR)	This is an individually administered, nationally normed, computer-adaptive cloze assessment of a K–12 student's level of reading achievement that takes about 10 minutes to complete. Developed by Renaissance Learning, the developer of *Accelerated Reader*™, STAR measures a student's reading ability and reading level for diagnosis and progress monitoring. The test includes exercises such as selecting a word from the list to best complete a given sentence. The test is standardized, and scale scores exhibit moderate to strong correlation to other standardized reading tests (as cited in Bullock, 2005).
Texas Assessment of Academic Skills (TAAS), Reading subtest	The TAAS was the state-administered benchmark test in Texas for grades 3 to 8 and 10 until replaced by the Texas Assessment of Knowledge and Skills in 2003. Reading test objectives are consistent across grades and include mastery in identifying word meaning, supporting ideas, summarization, relationships and outcomes, inferences and generalizations, point of view, propaganda, and fact and opinion. The reading test consists of approximately 50 multiple-choice questions about passages of various length and style (as cited in Nunnery & Ross, 2007, http://ritter.tea.state.tx.us/student.assessment/resources/guides/tli.html, http://www.education.com/reference/article/Ref_Explanation_TASS, and http://ritter.tea.state.tx.us/student.assessment/resources/guides/interpretive/2009_Interpretive_Guide_TAAS.pdf).

1. Duesbery, L., Alonzo, J., Bettesworth, L., Yovanoff, P., & Tindal, G. (2003). *Predicting middle school reading achievement using practical curriculum based measures of reading*. Eugene, OR: University of Oregon.

Appendix A3.1 Summary of study findings included in the rating for the reading fluency domain[1]

Outcome measure	Study sample	Sample size (students)	Author's findings from the study: Mean outcome (standard deviation)[2] — *Accelerated Reader*™ group	Author's findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] (*Accelerated Reader*™ – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Bullock, 2005[7]								
DIBELS Oral Reading Fluency[8]	Grade 4	45	132.70 (42.20)	119.30 (39.20)	13.40	0.32	ns	+13
DIBELS Oral Reading Fluency[8]	Grade 5	37	135.60 (50.50)	134.60 (39.30)	1.00	0.02	ns	+1
Domain average for reading fluency[9]						**0.17**	**na**	**+7**

ns = not statistically significant
na = not applicable
DIBELS = Dynamic Indicators of Basic Early Literacy Skills

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 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 Bullock (2005), no corrections for clustering or multiple comparisons were needed.
8. The intervention and comparison group means are posttest scores reported by the authors in the article.
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[1]

Outcome measure	Study sample	Sample size (clusters/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Accelerated Reader*™ group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] (*Accelerated Reader*™ – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Bullock, 2005[7]								
STAR, Reading[8]	Grade 4	45	472.00 (249.90)	473.60 (163.70)	–1.60	–0.01	ns	0
STAR, Reading[8]	Grade 5	37	564.00 (263.10)	510.40 (153.10)	53.60	0.25	ns	+10
4J Vocabulary[8]	Grade 4	42	63.50 (16.20)	64.10 (14.20)	–0.60	–0.04	ns	–2
Average for comprehension (Bullock, 2005)[9]						**0.11**	**na**	**+4**
Nunnery and Ross, 2007[7]								
TAAS, Reading[10]	Grade 5, cohort 1	18/912	88.44 (18.11)	89.45 (18.11)	–1.01	–0.06	ns	–2
TAAS, Reading[10]	Grade 5, cohort 2	18/891	91.53 (15.64)	90.64 (15.64)	0.89	0.06	ns	+2
TAAS, Reading[10]	Grade 8, cohort 2	4/992	90.67 (16.38)	88.56 (16.38)	2.11	0.13	ns	+5
Average for comprehension (Nunnery & Ross, 2007)[9]						**0.04**	**na**	**+2**
Domain average for comprehension across all studies[9]						**0.08**	**na**	**+3**

ns = not statistically significant
na = not applicable
STAR = Standardized Test and Assessment in Reading
TAAS = Texas Assessment of Academic Skills

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain. End of first year of intervention findings from Nunnery and Ross (2007) 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. For Nunnery and Ross (2007), the pooled standard deviation across two conditions is reported for each group.
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.

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain[1] *(continued)*

7. 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 Bullock (2005), no corrections for clustering or multiple comparisons were needed. In the case of Nunnery and Ross (2007), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
8. The intervention and comparison group means are posttest scores reported by the authors in the article.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. To prevent double counting within grade, the grade 4 effect in Bullock (2005) was calculated as a simple average of two effect sizes (for STAR Reading and 4J Vocabulary). The average effect size for the study was then calculated as a simple average of the grade 4 effect and grade 5 effect. The average improvement indices are calculated from the average effect sizes.
10. The intervention and comparison group means are calculated from author-reported untransformed scores by aggregating data across schools. The intervention group means are the comparison group means plus the difference in mean gains between the intervention and comparison groups. Because the authors used transformed scores to induce normality of the student test score distribution, the significance levels may differ from those reported in the original study.

Appendix A4 Summary of end of first year of intervention findings for the comprehension domain[1]

Outcome measure	Study sample	Sample size (clusters/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Accelerated Reader*™ group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] (*Accelerated Reader*™ – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Nunnery and Ross, 2007[7]								
TAAS, Reading[8]	Grade 4, cohort 1	18/912	87.27 (17.83)	87.64 (17.83)	–0.37	–0.02	ns	–1
TAAS, Reading[8]	Grade 4, cohort 2	18/891	89.01 (16.02)	87.77 (16.02)	1.24	0.08	ns	+3
TAAS, Reading[8]	Grade 7, cohort 2	4/992	88.38 (18.54)	87.27 (18.54)	1.11	0.06	ns	+2

ns = not statistically significant

TAAS = Texas Assessment of Academic Skills

1. This appendix presents findings from the end of the first year of intervention implementation for measures that fall in the comprehension domain. Findings from the end of the second and third year of intervention implementation 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. For Nunnery and Ross (2007), the pooled standard deviation across two conditions is reported for each group.
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 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 Nunnery and Ross (2007), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
8. The intervention and comparison group means are calculated from author-reported untransformed scores by aggregating data across schools. The intervention group means are the comparison group means plus the difference in mean gains between the intervention and comparison groups. Because the authors used transformed scores to induce normality of the student test score distribution, the significance levels may differ from those reported in the original study.

Appendix A5.1 *Accelerated Reader*™ 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.[1]

For the outcome domain of reading fluency, the WWC rated *Accelerated Reader*™ 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 statistically significant or substantively important effects, 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 statistically significant or substantively important 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.

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 statistically significant or 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.

  Not met. No studies showed statistically significant or substantively important negative effects. One study showed indeterminate effects, and no studies showed statistically significant or substantively important positive 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 statistically significant or substantively important effects, either positive or negative.

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 statistically significant or substantively important effects, and one study showed indeterminate effects.

(continued)

Appendix A5.1 *Accelerated Reader*™ rating for the reading fluency 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 statistically significant or substantively important effects, 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 statistically significant or substantively important effects, 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 statistically significant negative effects.

AND

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

  Met. No studies showed statistically significant or substantively important positive 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 *Accelerated Reader*™ 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.[1]

For the outcome domain of comprehension, the WWC rated *Accelerated Reader*™ 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. None of the studies showed statistically significant or substantively important effects, 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 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.
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 statistically significant or 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.
Not met. No studies showed statistically significant or substantively important negative effects, and two studies showed indeterminate effects, while no studies showed statistically significant or substantively important positive 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 statistically significant or substantively important effects, either positive or negative.
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 statistically significant or substantively important effects, and two studies showed indeterminate effects.

(continued)

Appendix A5.2 *Accelerated Reader*™ 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 statistically significant or substantively important effects, 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 statistically significant or substantively important effects, 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 statistically significant negative effects.

AND

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

  Met. No studies showed statistically significant or substantively important positive 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 A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	na	na	na	na
Reading fluency	1	1	82	Small
Comprehension	2	23	2,877	Medium to large
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.