Appendix

Appendix A1.1 Study Characteristics: Ross, Nunnery, & Goldfeder, 2004 (randomized controlled trial)

Characteristic	Description
Study citation	Ross, S.M., Nunnery, J., & Goldfeder, E. (2004). *A randomized experiment on the effects of Accelerated Reader/Reading Renaissance in an urban school district: Preliminary evaluation report*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.
Participants	In each of 11 schools, a minimum of two teachers at the same grade level volunteered to be randomly assigned either to implement *Accelerated Reader/Reading Renaissance* or to serve as a comparison teacher. Although participants were in grades K to 6, only students in grades K to 3 are relevant for this review. For grades K to 2, 32 teachers (642 students) were randomly assigned to an intervention or comparison group. The analysis sample included 394 students in grades K to 2 for whom pre- and posttest scores were available. There was no attrition of classrooms, but there was considerable student-level attrition in some grades, and the authors established equivalence of pretest scores for intervention and comparison students in the post-attrition sample. For third grade, 13 teachers (268 students) were randomly assigned to an intervention or a comparison group. There was no attrition of classrooms for the third grade sample, but approximately one-third of the students were missing either a pre- or posttest score, and 178 students are included in the analysis. Pretest scores were used as a covariate in outcome analyses. More than 80 percent of the students were eligible for free or reduced price lunch, and approximately 3 percent were identified as having a learning disability.[1]
Setting	Students attended 11 schools in Memphis, Tennessee.
Intervention	Teachers assigned to the intervention group implemented the *Accelerated Reader/Reading Renaissance* program (the computer software and the professional development on best practices for *Accelerated Reader/Reading Renaissance*). The authors report that the study occurred over an eight-month period during the 2002–2003 school year—the first year of program implementation.
Comparison	All schools in the study used the same commercially available basal reading program. Participating schools were implementing sustained silent reading programs to support fluency, comprehension, and vocabulary development. Comparison teachers were told that the *Accelerated Reader/Reading Renaissance* program would be available to them in the following school year.
Primary outcomes and measurement	The STAR Early Literacy Test was administered to students in kindergarten to second grade in September (pretest) and April (posttest). The STAR Reading Test was administered to third graders at the same time points. For a more detailed description of these outcome measures, see Appendix A2.2–2.3.
Staff/teacher training	The developer of the program, Renaissance Learning, trained teachers assigned to the intervention group to implement *Accelerated Reader/Reading Renaissance*. In addition, at least once a month throughout the year, Renaissance Learning consultants met with teachers in order to provide technical assistance and feedback on implementation.

1. These demographic characteristics pertain to the entire K–6 grade sample, not only to the K–3 sample on interest for this review.

Appendix A1.2 Study Characteristics: Bullock, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Bullock, J. C. (2005). Effects of the *Accelerated Reader* on reading performance of third, fourth, and fifth-grade students in one western Oregon elementary school. University of Oregon; 0171 Advisor: Gerald Tindal. DAI, 66 (07A), 56-2529.
Participants	The study examined students in grades 3–5, but the WWC analysis focused on third graders, as specified in the Beginning Reading protocol. Ninety-one percent of the students in the study school were white, and 61 percent qualified for free or reduced-price lunches. The third grade sample included 32 students, two classrooms, one school, and one school district. Experimental and control groups were created by first blocking on grade level, teacher, and reading ability. 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.[1] The author demonstrated the pretest equivalence of the intervention and control groups, and there was no attrition of students or classrooms between pretest and posttest.
Setting	The study was conducted in one elementary school near Eugene, Oregon.
Intervention	The intervention condition consisted of implementation of the *Accelerated Reader/Reading Renaissance* program over a 10-week period. Students in the intervention group 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/Reading Renaissance* quizzes were available. When finished with a book, students completed a brief, computerized, multiple-choice quiz on the book's contents 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/Reading Renaissance* books.
Comparison	The control condition relied on the business-as-usual reading program, without the addition of *Accelerated Reader/Reading Renaissance*. 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 were required to visit the library and check out a minimum of one book a week. Control students were free to choose any books in the library and asked to keep track of the books they read.
Primary outcomes and measurement	Reading fluency was measured using the Oral Reading Fluency subtest of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) administered by a trained educational assistant. Reading comprehension was measured using the STAR Reading Test, which is administered by computer and designed to customize tests for students' individual levels. The assessment lasted 15 minutes.
Staff/teacher training	Reading classes for the intervention and control groups were taught by the school's regular teachers. No information is given about any special training provided to those teachers.

1. The author of the study calls the design 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.

Appendix A2.1 Outcome measures in the reading fluency domain

Outcome measure	Description
Oral Reading Fluency subtest of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS)[1]	This is a standardized, individually-administered, one-minute oral reading fluency assessment designed to evaluate a student's accuracy and fluency with connected text (as cited in Bullock, 2005). It is designed to identify children who may need additional instructional support and monitor progress towards instructional goals.

1. The DIBELS is distributed by the Center on Teaching and Learning at the University of Oregon in Eugene, OR. Information on the DIBELS can be found at http://dibels.uoregon.edu. The website notes that the DIBELS oral fluency measure correlates .78 with the DIBELS measure of reading comprehension and .69 with the DIBELS measure of vocabulary.

Appendix A2.2 Outcome measures in the comprehension domain

Outcome measure	Description
STAR Reading Test[1]	This test is a computer-adaptive, norm-referenced test that measures student reading comprehension. It is designed for students who have at least a 100-word reading vocabulary and can be used with all students in grades 1–12. Students read passages of text and fill in key missing words from a set of options (modified cloze procedure). The assessment is designed for repeated administration throughout the school year to monitor progress (as cited in Ross, Nunnery, & Goldfeder, 2004).

1. This test was developed by Renaissance Learning, the developer of *Accelerated Reader*. According to research conducted by Renaissance Reading, STAR Reading Test scale scores are correlated with other standardized reading tests (such as, depending on the grade and time point, .67 to .85 for the California Achievement Test; .62 to .89 for the Gates McGinitie Test; and .71 for the Degrees of Reading Power test). (See: Nebelsick-Gullett, L. Review of STAR Reading, version 2.2. In B.S. Plake, J.C. Impara, & R.A. Spies (Eds.), *The fifteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved March 02, 2007, from Buros Institute of Mental Measurements website: http://www.unl.edu/buros/.)

Appendix A2.3 Outcome measures in the general reading achievement domain

Outcome measure	Description
STAR Early Literacy Test[1]	This test measures seven domains: general readiness, graphophonemic knowledge, phonics, comprehension, phonemic awareness, structural analysis, and vocabulary. It is a computer-adaptive audio test (students wear headphones and the test is read to them). The test can be administered to non-readers and to students who do not have a high enough reading vocabulary (100 words) to take the STAR Reading Test on their own. The assessment is designed for repeated administration throughout the school year to monitor progress (as cited in Ross, Nunnery, & Goldfeder, 2004).

1. This test was developed by Renaissance Learning, the developer of *Accelerated Reader*. According to research conducted by Renaissance Reading, the STAR Early Literacy Test is correlated with other standardized reading tests (average correlations range from .57 to .64 with the Brigance K & 1 Screen for Kindergarten and First Grade, the DIAL, the Iowa Tests of Basic Skills, and the Stanford Achievement Test). (See: Graham, T. [2003]. Review of STAR Literacy. In B.S. Plake, J.C. Impara, & R.A. Spies [Eds.], *The fifteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements. Retrieved March 02, 2007, from Buros Institute of Mental Measurements website: http://www.unl.edu/buros/.)

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

Outcome measure	Study sample	Sample size (teachers/ 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 (randomized controlled trial)[7]								
DIBELS Oral Reading Fluency Test	Grade 3	2/32	116.30 (40.90)	112.80 (55.40)	3.50	0.07	ns	+3
Domain average for reading fluency (Bullock, 2005)[8]						**0.07**	**ns**	**+3**

ns = not statistically significant

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 Technical Details of WWC-Conducted Computations.
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, 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations.
8. 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 (teachers/ 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]
Ross, Nunnery, & Goldfeder, 2004 (randomized controlled trial)[7]								
STAR Reading Test	Grade 3	13/178	389.5 (139.6)	336.8 (198.3)	52.70	0.31	ns	+12
Average for comprehension (Ross, Nunnery, & Goldfeder, 2004)[8]						**0.31**	**ns**	**+12**
Bullock, 2005 (randomized controlled trial)								
STAR Reading Test	Grade 3	2/32	412.40 (149.50)	462.30 (182.40)	–49.90	–.30	ns	–12
Average for comprehension (Bullock, 2005)[8]						**–.30**	**ns**	**–12**
Domain average for comprehension across all studies[8]						**.00**	**na**	**0**

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension 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. Ross, Nunnery, & Goldfeder (2004) adjusted posttest scores for pretest differences between study groups.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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, 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Ross, Nunnery, & Goldfeder (2004), a correction for clustering was needed, so the significance levels may differ from those reported in the original studies.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A3.3 Summary of study findings included in the rating for the general reading achievement domain[1]

Outcome measure	Study sample	Sample size (teachers/ 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]
Ross, Nunnery, & Goldfeder, 2004 (randomized controlled trial)[7]								
STAR Early Literacy test	Kindergarten	7/92	644.40 (114.40)	569.20 (94.10)	75.20	0.69	ns	+25
STAR Early Literacy test	Grade 1	9/97	733.60 (96.20)	698.00 (97.80)	35.60	0.36	ns	+14
STAR Early Literacy test	Grade 2	16/205	791.70 (72.10)	772.70 (82.20)	19.00	0.25	ns	+10
Average for general reading achievement (Ross, Nunnery, & Goldfeder, 2004)[8]						**0.43**	**Statistically significant**	**+16**

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the general reading achievement 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. The authors adjusted posttest scores for pretest differences between study groups.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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, 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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Ross, Nunnery, & Goldfeder (2004), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4.1 *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.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. *Accelerated Reader* has no studies that showed statistically significant or substantively important effects on reading fluency.

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. *Accelerated Reader* had no studies that showed statistically significant or substantively important positive effects.

AND

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

  Met. *Accelerated Reader* had no studies that 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. *Accelerated Reader* had no studies that 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.

  Met. *Accelerated Reader* had no studies that showed statistically significant or substantively important negative effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through EITHER of the following.

- 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. *Accelerated Reader* had no studies that showed statistically significant or substantively important positive effects and no studies that showed statistically significant or substantively important negative effects.

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. *Accelerated Reader* had no studies that showed statistically significant or substantively important positive effects and no studies that 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 Intervention Rating Scheme.

Appendix A4.2 *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 mixed effects.

Rating received

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.

  Met. *Accelerated Reader* had one study that showed a substantively important (not statistically significant) positive effect and one study that showed a substantively important (not statistically significant) negative effect on comprehension.

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. *Accelerated Reader* had no studies showing an indeterminate effect on comprehension.

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. *Accelerated Reader* had only one study that showed a substantively important (not statistically significant) positive effect.

AND

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

  Not met. *Accelerated Reader* had one study that showed a substantively important (not statistically significant) 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.

  Met. *Accelerated Reader* had one study that showed a substantively important (not statistically significant) 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. *Accelerated Reader* had one study that showed a substantively important (not statistically significant) negative effect.

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 Intervention Rating Scheme.

Appendix A4.3 *Accelerated Reader* rating for the general reading achievement 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 general reading achievement, the WWC rated *Accelerated Reader* as having potentially positive effects.

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. *Accelerated Reader* had one study showing a statistically significant positive effect on general reading achievement.

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. *Accelerated Reader* had no studies showing negative or indeterminate effects on general reading achievement.

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. *Accelerated Reader* had only one study that showed a substantively important (not statistically significant) positive effect.

AND

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

  Met. *Accelerated Reader* had no studies showing 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 Intervention Rating Scheme.

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	0	na	na	na
Fluency	1	1	32	Small
Comprehension	2	12	210	Medium to large
General reading achievement	1	12	426	Small

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."