

Appendix

Appendix A1 Study characteristics: Kemple et al., 2008

Characteristic	Description
Study citation	<p>Kemple, J. J., Corrin, W., Nelson, E., Salinger, T., Herrmann, S., & Drummond, K. (2008). <i>The Enhanced Reading Opportunities study: Early impact and implementation findings</i> (NCEE 2008–4015). Washington, DC: Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.</p> <p>Additional source: Corrin, W., Somers, M., Kemple, J. J., Nelson, E., & Sepanik, S. (2008). <i>The Enhanced Reading Opportunities study: Findings from the second year of implementation</i> (NCEE report no. 2009–4036). Washington, DC: Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.</p>
Participants	<p>The study is a randomized controlled trial that used a two-stage random assignment design. First, 34 eligible schools were randomly assigned to implement one of two supplemental literacy programs: 17 schools were assigned to <i>Reading Apprenticeship</i>[®] and 17 to <i>Xtreme Reading</i>.¹ Second, in each of the 17 schools assigned to <i>Reading Apprenticeship</i>[®], 9th-grade students were randomly assigned to <i>Reading Apprenticeship</i>[®] or to the control group. Eligible students were defined as those who were reading at least two years below grade level. The study includes two cohorts of 9th-grade students: Cohort 1 was formed in the 2005/06 school year and consisted of 686 ninth-grade students who received <i>Reading Apprenticeship</i>[®] and 454 ninth-grade students in the control group who did not. Cohort 2 was formed in the 2006/07 school year and consisted of 645 ninth-grade students who received <i>Reading Apprenticeship</i>[®] and 470 ninth-grade students in the control group who did not. Overall and differential attrition rates of student attrition were low for Cohort 1 (30% and 6%, respectively) and Cohort 2 (36% and 3%, respectively).</p>
Setting	<p>The study was conducted in 17 schools located in 10 school districts across the United States.</p>
Intervention	<p>The intervention group received the <i>Reading Apprenticeship</i>[®] <i>Academic Literacy</i> course as a supplemental intervention that replaced a 9th-grade elective class, rather than a core academic class. Therefore, students in the intervention group continued to attend their regular English language arts classes. <i>Reading Apprenticeship</i>[®] <i>Academic Literacy</i> includes a detailed curriculum that is guided by the concept of “flexible fidelity” (i.e., teachers have flexibility in how they implement the curriculum in their day-to-day activities). Students attended <i>Reading Apprenticeship</i>[®] classes for about 11 hours per month, on average. The study reported students’ outcomes after 7.5 to 9 months of program implementation.</p>
Comparison	<p>The control group received the standard instruction provided in the regular school curriculum and continued their participation in any regularly scheduled elective class (such as career/technical education, art, physical education, health, or foreign language).</p>
Primary outcomes and measurement	<p>For both the pretest and posttest, students took the reading comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE). For the posttest, students also took the GRADE vocabulary subtest. For a more detailed description of these outcome measures, see Appendix A2.</p>
Staff/teacher training	<p>Teachers took part in professional development activities prior to the start of the school year and on an ongoing basis throughout the school year. Training included one 5-day summer training institute as well as two 2-day booster sessions during the school year. They also received three 2-day coaching visits during the year and had access to a special online listserv.</p>

1. Comparison between the *Xtreme Reading* group and the control group is outside the scope of this review. The study also presented impact findings for the subgroups of students defined by their baseline reading comprehension test scores, whether they were overage for the ninth grade, and whether a language other than English was spoken in their homes. These subgroup analyses were presented for the combined intervention group that included both literacy programs: *Xtreme Reading* and *Reading Apprenticeship*[®]. These findings are also outside the scope of this review because the measures of effectiveness cannot be attributed solely to *Reading Apprenticeship*[®].

Appendix A2 Outcome measures for the comprehension domain

Outcome measure	Description
<i>Reading comprehension construct</i>	
Group Reading Assessment and Diagnostic Evaluation (GRADE): reading comprehension subtest	GRADE is a norm-referenced reading assessment that can be used with students in pre-K through adulthood. The GRADE has four subtests: (1) vocabulary, (2) sentence comprehension, (3) passage comprehension, and (4) listening comprehension. The reading comprehension subtest score is a composite of the sentence comprehension and passage comprehension subtest scores (as cited in Kemple et al., 2008).
<i>Vocabulary development construct</i>	
GRADE: vocabulary subtest	GRADE is a norm-referenced reading assessment that can be used with students in pre-K through adulthood. The vocabulary subtest assesses students' decoding and vocabulary knowledge (as cited in Kemple et al., 2008).

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

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome ² (standard deviation) ³		Mean difference ⁴ (Reading Apprenticeship [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			Reading Apprenticeship [®] group	Comparison group				
Kempe et al., 2008⁸								
GRADE: comprehension subtest	Grade 9, cohort 1	1140	89.88 (10.35)	88.94 (10.35)	0.94	0.09	ns	+4
GRADE: vocabulary subtest	Grade 9, cohort 1	1140	93.33 (10.18)	92.85 (10.19)	0.48	0.05	ns	+2
GRADE: comprehension subtest	Grade 9, cohort 2 ⁹	1115	90.43 (10.02)	89.05 (10.11)	1.38	0.14	Statistically significant	+5
GRADE: vocabulary subtest	Grade 9, cohort 2 ⁹	1115	93.59 (10.04)	94.02 (9.31)	-0.43	-0.04	ns	-2
Domain average for comprehension (Kempe et al., 2008)¹⁰						0.06	na	+2

ns = not statistically significant

na = not applicable

GRADE = Group Reading Assessment and Diagnostic Evaluation

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain.
2. The intervention group values are the comparison group means plus the regression-adjusted impacts presented in the study.
3. 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.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. 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.
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 Kempe et al. (2008), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
9. Findings for Cohort 2 ninth-grade students are reported in Corrin et al. (2008). See References section for more information.
10. 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 Reading Apprenticeship® 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 *Reading Apprenticeship*® as having potentially positive effects for adolescent learners. The remaining ratings (mixed, no discernible effects, potentially negative, or negative) were not considered, as *Reading Apprenticeship*® 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 a 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.

Met. No study showed a statistically significant or substantively important negative effect or 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. One study of *Reading Apprenticeship*® showed a statistically significant positive effect.

AND

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

Met. No study showed a statistically significant or substantively important 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 Procedures and Standards Handbook, Appendix E.

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
Reading fluency	0	na	na	na
Comprehension	1	17	2255	Small
General literacy achievement	0	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.