

Appendix

Appendix A1 Study characteristics: Reading Plus, 2008

Characteristic	Description
Study citation	Reading Plus. (2008). <i>Reading improvement report: Miami-Dade regions II and III</i> . Huntington Station, NY: Taylor Associates/Communications, Inc.
Participants	This quasi-experimental study was conducted in 98 schools and included students in grades 5 to 9 who had valid 2006 and 2007 Florida Comprehensive Assessment Test (FCAT) reading scores. Students who completed one or more <i>Reading Plus</i> ® lessons during the 2006–07 school year formed the intervention group, and students who completed no <i>Reading Plus</i> ® lessons during the same period constituted the comparison group. Although <i>Reading Plus</i> ® was used by students in various grades and student populations within each of the 98 schools, ¹ this review focuses on students who scored at achievement level 1 or 2 (non-proficient) on the 2006 reading portion of the FCAT. The analysis sample consisted of 6,070 low-achieving students in the <i>Reading Plus</i> ® group and 7,058 low-achieving students in the comparison group.
Setting	The study was conducted in 98 schools located in two regions that are part of the Miami-Dade School District in Florida. The participating schools served more than 90% minority students and 22% English language learners, and 15% of the student population received special education services.
Intervention	<i>Reading Plus</i> ® is a computer-based program that is implemented online. It encompasses the following activities: instant word recognition practice with less familiar words, reading practice with text difficulty matched to each student's reading ability, exposure to high-utility words within assigned texts, attention building activities, sustained reading time matched with each student's attention span, vocabulary building, word acquisition activities, and practice with 25 comprehension skills. The participating schools implemented <i>Reading Plus</i> ® to varying degrees. Most schools followed a schedule of either two 45-minute sessions per week or three 30-minute sessions per week for approximately six months. The average number of lessons completed by students participating in the study was 33.
Comparison	Comparison students completed no <i>Reading Plus</i> ® lessons. Most nonparticipating students used Scholastic's <i>Read 180</i> and/or Renaissance Learning's <i>Accelerated Reader</i> .
Primary outcomes and measurement	For both the pretest and the posttest, students took the Florida Comprehensive Assessment Test (FCAT). Reading comprehension was measured using the developmental scaled scores on the reading section of the FCAT. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	Intervention teachers were trained on the use and intent of <i>Reading Plus</i> ® during the fall of 2006. No other information is available about the training.

1. Analyses for the whole sample, for each grade (5, 6, 7, 8, and 9), and for subpopulations (by ethnicity, for the average/high achieving subgroup), and for students receiving 40 or more *Reading Plus*® lessons, do not meet WWC evidence standards because the intervention and comparison groups were not shown to be equivalent at baseline.

Appendix A2 Outcome measure for the comprehension domain

Outcome measure	Description
Reading comprehension construct	
Florida Comprehensive Assessment Test (FCAT) reading portion	The reading portion of this standardized test is a group-administered, criterion-referenced test consisting of six to eight informational and literary reading passages (Florida Department of Education, 2005). ¹ In grades 3 through 10, students respond to between 6 and 11 multiple-choice items for each passage and are assessed across four content clusters: (1) reading comprehension in the areas of words and phrases in context, (2) main idea, (3) comparison and cause and effect, and (4) reference and research. In grades 4, 8, and 10, open-ended questions are included (as cited in Reading Plus, 2008; Schatschneider et al., 2004). ²

1. Florida Department of Education. (2005, September). *Florida Comprehensive Assessment Test summary of tests and design*. Retrieved August 21, 2008, from <http://fc05.fldoe.org/pdf/fc05designsummary.pdf>
2. Schatschneider, C., Buck, J., Torgesen, J. K., Wagner, R. K., Hassler, L., Hecht, S., and Powell-Smith, K. (2004). *A multivariate study of factors that contribute to individual differences in performance on the Florida Comprehensive Reading Assessment Test* (Technical Report No. 5). Tallahassee: Florida Center for Reading Research.

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 Plus [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			Reading Plus [®] group	Comparison group				
Reading Plus, 2008^{8,9}								
FCAT reading portion	Grades 5–9 (levels 1–2)	13,128	1,554.54 (275.07)	1,538.59 (295.19)	15.95	0.06	Statistically significant	+2
Domain average for comprehension (Reading Plus, 2008)¹⁰						0.06	Statistically significant	+2

FCAT= Florida Comprehensive Assessment Test

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain.
2. The *Reading Plus*[®] group mean outcome values for Reading Plus (2008) are the unadjusted comparison group posttest means plus the difference in mean gains between the intervention and comparison groups.
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 indicates 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 Reading Plus (2008), no corrections for clustering or multiple comparisons were needed.
9. For Reading Plus (2008), the mean outcomes and standard deviations were received through communication with the authors.
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 Plus® 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 comprehension outcome domain, the WWC rated *Reading Plus*® 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 Plus*® 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 statistically significant 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.

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, which did not meet WWC evidence standards for a strong design, 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 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	na	na	na	na
Comprehension	1	98	13,128	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.