

Appendix

Appendix A1.1 Study characteristics: Horsfall & Santa, 1994 (random assignment study)

Characteristic	Description
Study citation	Horsfall, S., & Santa, C. (1994). Project CRISS: <i>Validation report for the Program Effectiveness Panel</i> . Unpublished manuscript.
Participants	Sixteen intact classrooms of students in grades 4, 6, 8, and 11 participated in the study during the 1991–92 school year; however, only analysis of grades 4 and 6 were included in this review. ¹ Teachers within each of three schools were randomly assigned either to <i>Project CRISS®</i> or to the control condition. Teachers assigned to the intervention received <i>CRISS</i> training; control group teachers did not. Within each grade level, 4 and 6, there were three classrooms assigned to <i>Project CRISS®</i> and three classrooms assigned to the control group. Four or five students in each class were excluded from analyses due to attrition; there was no attrition of teachers. In all, the analysis sample consisted of 120 students attending six <i>Project CRISS®</i> classrooms and 111 students attending six control group classrooms.
Setting	The study took place across three different settings: (1) Kalispell School District, MT, a rural district in northwestern Montana that serves primarily white students; (2) Putnam County School District, FL, a district in central Florida that serves a population composed of white (77%), black (20%), and Hispanic students; and (3) Stafford School District, VA, a district in suburban Washington, DC that serves primarily white students.
Intervention	Intervention group students received <i>Project CRISS®</i> strategies as part of their regular instruction for approximately 18 weeks during one semester.
Comparison	Control group students received regular instruction and were not given <i>Project CRISS®</i> strategies.
Primary outcomes and measurement	For both the pretest and posttest, students took the staff-developed “free recall” tests that require students to remember details from a passage read the day before. For a more detailed description of this outcome measure, see Appendix A2.
Staff/teacher training	Teachers assigned to the intervention received <i>Project CRISS®</i> training. Districts selected a local facilitator to coordinate the program. The facilitator organized a 12-hour training conducted over two consecutive days. During this training, trainers modeled <i>Project CRISS®</i> strategies, and teachers were given the opportunity to apply each of the <i>Project CRISS®</i> strategies to their own curriculum materials. After teachers completed the training, the facilitator worked with project staff to set up a follow-up session three to six months after the completion of the final training session. Teachers frequently met to share <i>Project CRISS®</i> ideas before or after school or during duty-free periods. In addition, <i>Project CRISS®</i> trainers provided follow-up assistance for teachers through on-site visits, demonstration lessons, newsletters, and a computer network.

1. For the 8th and 11th grade samples, there was only one teacher assigned to the treatment group and one teacher assigned to the control group for each grade level. For this reason, the results from the 8th and 11th grade analyses could be confounded with factors unrelated to the *CRISS* treatment and thus are not included in the WWC's rating of effectiveness.

Appendix A1.2**Study characteristics: James-Burdumy et al., 2009 (randomized controlled trial)**

Characteristic	Description
Study citation	James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., et al. (2009). <i>Effectiveness of selected supplemental reading comprehension interventions: Impacts on a first cohort of fifth-grade students</i> (NCEE 2009–4032). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
Participants	The study, which explored the impact of <i>Project CRISS®</i> as well as three other reading comprehension curricula (<i>ReadAbout</i> , <i>Read for Real</i> , and <i>Reading for Knowledge</i>), included 6,350 5th-grade students from 89 schools in ten school districts. Districts that had at least 12 Title I schools and who were not implementing any of the four selected curricula were recruited into the study. Within each school district, schools were randomly assigned either to one of the four intervention conditions or to the control group. Eligible students attended study schools and were enrolled in grade 5 when baseline tests were administered or transferred in after baseline and before January 1, 2007. Multiage grade levels and non-mainstreamed special education students were excluded from the sample. The analysis that is included in this review focused on the effect of <i>Project CRISS®</i> and examined a sample of 1,155 students attending 17 <i>Project CRISS®</i> schools and 1,183 students attending 21 control schools.
Setting	The study took place in ten geographically diverse school districts in eight states (Arizona, California, Florida, Georgia, Louisiana, Oregon, Texas, and Wisconsin). To be eligible for the study, school districts needed to have (1) at least 12 schools that received Title I funds, (2) at least 40% of students eligible for the federal free or reduced-price lunch program, and (3) at least 60 5th-grade students per school. The school districts in the study were significantly larger, more disadvantaged, and more urban than the average U.S. school district.
Intervention	Intervention group students received <i>Project CRISS®</i> strategies as part of their regular instruction. Instructional components included: (1) use of student and teacher editions of <i>Learning How to Learn</i> , which provided detailed lesson plans, learning, and practice through use of <i>Tough Terminators</i> , a science trade book; (2) use of a variety of graphic organizers and note-taking, discussion, vocabulary, and writing strategies; and (3) application of strategies to regular science and social studies texts. <i>Project CRISS®</i> teachers, on average, were observed engaging in 78% of teaching practices important to intervention implementation. The study reported students' reading comprehension outcomes after nine months of program implementation.
Comparison	Control group schools did not have access to any of the four curricula being tested. Control group teachers could, however, use other supplemental reading programs.
Primary outcomes and measurement	For the pretest, students took the passage comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE) and the Test of Silent Contextual Reading Fluency (TOSCRF). For the posttest, all students took the passage comprehension subtest of the GRADE. Students were also randomly assigned to take one of two reading comprehension assessments developed by the Educational Testing Service (ETS) for this study; these tests focused on either science or social studies. For a more detailed description of these outcome measures, see Appendix A2.
Staff/teacher training	<i>Project CRISS®</i> teachers received 18 hours of initial training, including 12 hours on using the strategies in the teacher's guide and six hours on using the student text and workbook. Teachers received a training manual, a teacher's guide, a student text, and a wrap-around edition of the student workbook. In addition, teachers received six hours of follow-up training. Trainers also visited schools monthly to observe teachers and provide feedback. The developer also encouraged teachers to use bi-weekly study teams in which teachers review and discuss their use of <i>CRISS</i> strategies.

Appendix A2 Outcome measures for the comprehension domain

Outcome measure	Description
Reading comprehension construct	
Staff-developed (CRISS) free recall assessment	The outcome is a staff-developed “free recall” measure that requires students to read a passage (2–4 pages, depending on grade level) over a 40-minute period; then, 24 hours later, students write down from memory as much as they can remember from the passage. Students are scored based on the number of idea units (one point per idea) they remember. The topics varied across grade levels, and care was taken to choose readings that had a content base similar to what students would experience in their regular coursework but with actual topics that would not normally have been covered in those courses. The Grade 4 assessment covered <i>The Western Movement</i> (770 words, 2 single-spaced pages), and the Grade 6 assessment covered <i>The Mystery of Thirst</i> (920 words, 3 single-spaced pages). Inter-rater reliability ranged from 0.93 to 0.95 (as cited in Horsfall & Santa, 1994).
Group Reading Assessment and Diagnostic Evaluation (GRADE)—Passage Comprehension subtest	This standardized measure is a norm-referenced diagnostic test for all reading abilities. The Passage Comprehension subtest measures a student’s understanding of an extended text through explicit and implicit multiple choice questions requiring questioning, predicting, summarizing, and clarifying information from several paragraphs (as cited in James-Burdumy et al., 2009).
Educational Testing Service (ETS) science reading comprehension assessment	This assessment, designed by the ETS, focuses on students’ reading comprehension of science text. The test measures the ability to comprehend five science-related expository text passages based on responses to six multiple choice questions per passage. Internal consistency was reported as 0.85 (as cited in James-Burdumy et al., 2009).
ETS social studies reading comprehension assessment	This assessment, designed by the ETS, focuses on students’ reading comprehension of social studies text. The test measures the ability to comprehend five social studies-related expository text passages based on responses to six multiple choice questions per passage. Internal consistency was reported as 0.84 (as cited in James-Burdumy et al., 2009).

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

Outcome measure	Study sample	Sample size (classrooms or schools/students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation) ²		Mean difference ³ (Project CRISS®– comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Project CRISS® group	Comparison group				
Horsfall & Santa, 1994^{7,8}								
Free recall assessment	Grade 4	6 classrooms/ 118 students	8.97 (3.38)	5.32 (2.73)	3.65	1.17	Statistically significant	+38
Free recall assessment	Grade 6	6 classrooms/ 113 students	12.54 (5.35)	7.93 (4.07)	4.61	0.96	Statistically significant	+33
Average for comprehension (Horsfall & Santa, 1994)⁹						1.07	na	+36
James-Burdumy et al., 2009¹⁰								
GRADE—Passage comprehension	Grade 5	38 schools/ 2,332 students	100.48 (14.20)	101.06 (13.69)	-0.57	-0.04	ns	-2
ETS science comprehension	Grade 5	38 schools/ 1,153 students	501.44 (29.51)	500.76 (27.59)	0.69	0.02	ns	+1
ETS social studies comprehension	Grade 5	38 schools/ 1,140 students	499.64 (30.57)	500.61 (29.68)	-0.96	-0.03	ns	-1
Average for comprehension (James-Burdumy et al., 2009)⁹						-0.02	na	-1
Domain average for comprehension across all studies⁹						0.53	na	+20

ns = not statistically significant

na = not applicable

GRADE = Group Reading Assessment and Diagnostic Evaluation

ETS = Educational Testing Service

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.
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.

(continued)

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

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 Horsfall and Santa (1994), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study. In the case of James-Burdumy et al. (2009), the authors adjusted for clustering and applied a correction for multiple comparisons.
8. The *Project CRISS®* group mean outcome values for Horsfall and Santa (1994) are the unadjusted control group posttest means plus the difference in mean gains between the intervention and control groups. Control group means are unadjusted.
9. 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.
10. Sample sizes, regression-adjusted means, and standard deviations were provided to the WWC by the study authors and thus differ slightly from the information presented in the original study.

Appendix A4 Project CRISS® 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 *Project CRISS®* 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 *Project CRISS®* 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, and one study showed 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. Only one study 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	Schools	Sample size		Extent of evidence ¹
			Students	Extremely small	
Alphabetics	na	na	na	na	na
Reading fluency	na	na	na	na	na
Comprehension	2	41	2,569 ²	Medium to large	
General literacy achievement	na	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.
2. This sample size varies slightly from the sample sizes presented in Appendix A3 because in James-Burdumy et al. (2009), the total sample size includes students who had outcomes for at least one of the three tests.