

# What Works Clearinghouse



## Cooperative Integrated Reading and Composition<sup>®</sup>

### Program Description<sup>1</sup>

*Cooperative Integrated Reading and Composition*<sup>®</sup> is a reading and writing program for students in grades 2 through 6. It has three principal elements: story-related activities, direct instruction in reading comprehension, and integrated language arts/writing. Daily lessons provide students with an opportunity to practice comprehension and reading skills in pairs and small groups. Pairs of students read to each other; predict how stories

will end; summarize stories; write responses to questions posed by the teacher; and practice spelling, decoding, and vocabulary. Within cooperative teams of four, students work to understand the main idea of a story and work through the writing activities linked to the story. A Spanish version of the program is available for grades 2 through 5.

### Research<sup>2</sup>

Two studies of *Cooperative Integrated Reading and Composition*<sup>®</sup> that fall within the scope of the Adolescent Literacy review protocol meet What Works Clearinghouse (WWC) evidence standards with reservations. The two studies included approximately 1,460 students in grades 2 through 6 who attended nine schools<sup>3</sup> located in two school districts in the United States.<sup>4</sup>

Based on these two studies, the WWC considers the extent of evidence for *Cooperative Integrated Reading and Composition*<sup>®</sup>

on adolescent learners to be medium to large for the comprehension and general literacy achievement domains. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Cooperative Integrated Reading and Composition*<sup>®</sup> on adolescent learners in the alphabetic or reading fluency domains.

1. The descriptive information for this program was obtained from a publicly available source: the WWC *Beginning Reading* intervention report (July 2007). The WWC requests developers to review the program description sections for accuracy from their perspective. The program description was updated by the developer in December 2008. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by March 2010.
2. The studies in this report were reviewed using WWC Evidence Standards, Version 2.0 (see the WWC Procedures and Standards Handbook, Chapter III), as described in protocol Version 2.0.
3. The Adolescent Literacy topic area reviews studies of interventions administered to students in grades 4–12 (or 9–18 years of age). For studies that include samples of students that span both the Adolescent Literacy (grades 4–12) and Beginning Reading (grades K–3) topic areas and cannot be disaggregated by grade level, the Adolescent Literacy topic area also reviews any studies that include 5th-grade students or higher. For example, this report includes a combined sample of students from grades 2–6 (Jewell, 1994; Stevens & Slavin, 1995).
4. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

**Effectiveness** *Cooperative Integrated Reading and Composition*<sup>®</sup> was found to have potentially positive effects on comprehension and general literacy achievement for adolescent learners.

	<i>Alphabetics</i>	<i>Reading fluency</i>	<i>Comprehension</i>	<i>General literacy achievement</i>
<b>Rating of effectiveness</b>	na	na	Potentially positive effects	Potentially positive effects
<b>Improvement index<sup>5</sup></b>	na	na	Average: +7 percentile points	Average: +2 percentile points
	na	na	Range: +3 to +11 percentile points	Range: -3 to +8 percentile points

na = not applicable

### Additional program information

#### Developer and contact

*Cooperative Integrated Reading and Composition*<sup>®</sup> was developed in 1983 by Robert Slavin and Nancy Madden at the Center for Social Organization of Schools at The Johns Hopkins University. *Cooperative Integrated Reading and Composition*<sup>®</sup> is distributed by the Success for All Foundation, Inc. Address: 200 W. Towsontown Blvd., Baltimore, MD 21204-5200. Email: sfainfo@successforall.org. Web site: <http://www.successforall.net/Programs/readingwings.html>. Telephone: (800) 548-4998 ext. 2372.

#### Scope of use

*Cooperative Integrated Reading and Composition*<sup>®</sup> was first used as part of a cooperative elementary whole-school reform model. The program was later reformulated as *Reading Roots* (for beginning readers) and *Reading Wings* (for upper elementary students) and is a component of both the *Success for All* comprehensive school reform model and a stand-alone reading program.

#### Teaching

The program uses daily 90-minute lessons to focus on story-related activities, direct instruction in reading comprehension, and integrated reading and language arts activities. In a team setting, mixed-ability students work together to read, discuss their reading to clarify unknown vocabulary, reread for fluency, understand the main idea, comprehend stories, and work through the writing process linked to the texts that the students are reading (including drafting, revising, and editing each other's writing). Students are rewarded on the basis of the team's performance to provide motivation to work together and help each other.

Teacher training includes a two-day session that covers word structure and phonics, vocabulary development, fluency, and comprehension skills, as well as program management and cooperative learning strategies. Technical support by phone or on-site visits is also provided.

#### Cost<sup>6</sup>

The cost of the program is approximately \$150 per student for training and materials, depending on school size and the number of schools within a district that are participating.

5. These numbers show the average and range of student-level improvement indices for all findings across the studies.

6. The prices reported here are from the July 2007 *Beginning Reading* intervention report.

**Research** Fifty-two studies reviewed by the WWC investigated the effects of *Cooperative Integrated Reading and Composition*<sup>®</sup> on adolescent learners. Two studies (Jewell, 1994; Stevens & Slavin, 1995) are quasi-experimental designs that meet WWC evidence standards with reservations. The remaining 50 studies do not meet either WWC evidence standards or eligibility screens.

#### Meets evidence standards with reservations

The studies included in this report evaluated *Cooperative Integrated Reading and Composition*<sup>®</sup> either as part of a cooperative elementary whole-school reform model (Stevens & Slavin, 1995) or as a stand-alone reading program (Jewell, 1994).

Stevens and Slavin (1995) conducted a quasi-experiment that examined the effects of the *Cooperative Integrated Reading and Composition*<sup>®</sup> curriculum on students in grades 2 through 6 attending five elementary schools in Maryland. Investigators matched two treatment schools with three comparison schools based on academic achievement, ethnicity, and socioeconomic background. Treatment schools implemented the Cooperative Elementary School model, a whole-school reform model that uses cooperative learning strategies across multiple content areas, in which *Cooperative Integrated Reading and Composition*<sup>®</sup> was taught as a language arts/reading curriculum. Comparison schools implemented some components of the Cooperative Elementary School model but not the *Cooperative Integrated Reading and Composition*<sup>®</sup> curriculum. The WWC based its effectiveness ratings on findings from comparisons of 411 students in two schools who received *Cooperative Integrated Reading and Composition*<sup>®</sup> and 462 comparison group students in three schools who received regular instruction. The

study reported students' outcomes after one and two years of program implementation.<sup>7</sup>

Jewell (1994) conducted a quasi-experiment that examined the effects of *Cooperative Integrated Reading and Composition*<sup>®</sup> on students in grades 2 through 6 attending four elementary schools in one school district in the United States. Teachers volunteered to participate in the study; however, a number of these teachers were placed in the comparison group because they elected not to implement *Cooperative Integrated Reading and Composition*<sup>®</sup>. The analytical sample included 15 classrooms whose students received *Cooperative Integrated Reading and Composition*<sup>®</sup> and 15 comparison classrooms whose students received the regular district-adopted reading and language arts program. The study reported students' outcomes after seven to eight months of program implementation.

#### Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.<sup>8</sup>

The WWC considers the extent of evidence for *Cooperative Integrated Reading and Composition*<sup>®</sup> to be medium to large for comprehension and general literacy achievement for adolescent learners. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Cooperative Integrated Reading and Composition*<sup>®</sup> in the alphabets or reading fluency domains for adolescent learners.

7. Two-year findings are considered for the effectiveness rating because these findings reflect the maximum exposure to the program. One-year findings are not included in this rating but are reported in Appendices A4.1 and A4.2.
8. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept—external validity, such as the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *Cooperative Integrated Reading and Composition*<sup>®</sup> is in Appendix A6.

## Effectiveness Findings

The WWC review of interventions for Adolescent Literacy addresses student outcomes in four domains: alphabetics, reading fluency, comprehension, and general literacy achievement. The studies included in this report cover two domains: comprehension and general literacy achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Cooperative Integrated Reading and Composition*® on adolescent learners.<sup>9</sup>

*Comprehension.* Stevens and Slavin (1995) reported, and the WWC confirmed, statistically significant positive effects of *Cooperative Integrated Reading and Composition*® on the Reading Comprehension and Vocabulary subtests of the California Achievement Test (CAT) for students in grades 2 through 6. Jewell (1994) did not find statistically significant effects of *Cooperative Integrated Reading and Composition*® on the Reading Comprehension and Vocabulary subtests of the Gates–MacGinitie Reading Test for students in grades 2 through 6, and the WWC-calculated average effect across the two measures was not statistically significant or large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25).<sup>10</sup> Thus, for the comprehension domain, one study showed statistically significant positive effects, and one study showed indeterminate effects.

*General literacy achievement.* Stevens and Slavin (1995) reported, and the WWC confirmed, a statistically significant positive effect of *Cooperative Integrated Reading and Composition*® on the Language Expression subtest of the CAT for students in grades 2 through 6 but did not find statistically significant effects on the CAT Language Mechanics subtest. Jewell (1994) did not find statistically significant effects of *Cooperative Integrated Reading and Composition*® on the Reading Proficiency subtest of the Bass Academic Skills Sample for students in grades 2 through 6. According to WWC calculations, the effect was not large enough to be considered substantively important according to WWC criteria. Thus, for the general literacy achievement domain, one study showed statistically significant positive effects, and one study showed indeterminate effects.

### Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings, the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Procedures and Standards Handbook, Appendix E).

9. 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 Stevens and Slavin (1995), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. As the authors used hierarchical linear modeling (HLM) analyses, which accounted for multi-level data (of students nested within classrooms and schools), correction for clustering was not needed. In the case of Jewell (1994), no corrections for clustering or multiple comparisons were needed.
10. The WWC computes an average effect size as a simple average of the effect sizes across all individual findings within the study domain.

**The WWC found *Cooperative Integrated Reading and Composition*® to have potentially positive effects on comprehension and general literacy achievement for adolescent learners**

**Improvement index**

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see WWC Procedures and Standards Handbook, Appendix F). The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.

The average improvement index for comprehension is +7 percentile points, with a range of +3 to +11 percentile points across findings from two studies. The average improvement index for general literacy achievement is +2 percentile points, with a range of -3 to +8 percentile points across findings from two studies.

**Summary**

The WWC reviewed 52 studies of *Cooperative Integrated Reading and Composition*® for adolescent learners. Two studies meet WWC evidence standards with reservations; the remaining 50 studies do not meet either WWC evidence standards or eligibility screens. Based on these studies, the WWC found potentially positive effects on comprehension and general literacy achievement for adolescent learners. The conclusions presented in this report may change as new research emerges.

**References**

**Meets WWC evidence standards with reservations**

Jewell, M. E. (1994). The effects of classroom-based follow-up assistance on mainstream reading and language arts instruction (Doctoral dissertation, University of Washington, 1994). *Dissertation Abstracts International*, 55(11A), 107-3473.

Stevens, R. J., & Slavin, R. E. (1995). The Cooperative Elementary School: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal*, 32(2), 321-351.

**Additional source:**

Stevens, R. J., & Slavin, R. E. (1992). *The Cooperative Elementary School: Effects on students' achievement, attitudes, and social relations*. Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students.

**Studies that fall outside the Adolescent Literacy review protocol or do not meet WWC evidence standards**

American Federation of Teachers. (1998). *Building on the best, learning from what works: Seven promising reading and English language arts programs*. Washington, DC: Author.

The study is ineligible for review because it is not a primary

analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

August, D. (2002). *Transitional programs for English language learners: Contextual factors and effective programming*. Baltimore, MD: Center for Research on the Education of Students Placed At Risk (CRESPAR). The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2002). *Comprehensive school reform and student achievement: A meta-analysis*. Baltimore, MD: CRESPAR/Johns Hopkins University. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Borman, G. D., Rachuba, L., Datnow, A., Alberg, M., MacIver, M., Stringfield, S., et al. (2000). *Four models of school improvement. Successes and challenges in reforming low-performing, high poverty Title 1 schools* (Report No. 48). Baltimore, MD: Center for Research on the Education of Students Placed At Risk (CRESPAR). The study is ineligible for review because it does not use a comparison group design or a single-case design.

## References (continued)

- Bramlett, R. K. (1994). Implementing cooperative learning: A field study evaluating issues for school-based consultants. *Journal of School Psychology, 32*(1), 67–84. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Additional source:**  
Bramlett, R. K. (1992, August). *Cooperative learning: A field study with implications for school psychologists*. Paper presented at the annual convention of the American Psychological Association, Washington, DC.
- Briggs, K. L., & Clark, C. (1997). *Reading programs for students in the lower elementary grades: What does the research say?* Austin, TX: Texas Center for Educational Research. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Chambers, B., Abrami, P. C., Tucker, B. J., Cheung, A., & Gifford, R. (2005). *Computer-assisted tutoring in Success for All: Reading outcomes for first graders*. Baltimore, MD: Success for All. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P. C., Tucker, B. J., Cheung, A., et al. (2005). *Technology infusion in Success for All: Reading outcomes for first-graders*. Baltimore, MD: Success for All. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Costa, S. R. (1994). Limited reading proficient students in two types of cooperative learning groups for reading instruction (Doctoral dissertation, Fordham University, 1994). *Dissertation Abstracts International, 55*(11A), 385–3460. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Crowe, E. C., Connor, C. M., & Petscher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. *Journal of School Psychology, 47*(3), 187–214. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Dillon, I. C. (2000). A comparative study of compensatory education outcomes of at-risk students (Doctoral dissertation, Texas Southern University, 2000). *Dissertation Abstracts International, 62*(12A), 93–4110. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Duran, R. P., & Szymanski, M. H. (1993). *Construction of learning and interaction of language minority children in cooperative learning* (Report No. 45). Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Education Commission of the States. (1999). *Cooperative Integrated Reading and Composition (CIRC)*. Denver, CO: Author. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Educational Research Service. (1990). *What we know about: Cooperative learning*. Arlington, VA: Author. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Golmic, B. A. (1992). Facilitating change: An ethnographic study to identify interventions applied during the implementation of a *Cooperative Integrated Reading and Composition* program (Doctoral dissertation, Indiana University of Pennsylvania, 1992). *Dissertation Abstracts International, 53*(10A), 235–3441. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Grek, M. L., & Robinson, C. (2002). *Success for All*. Tallahassee, FL: Florida Center for Reading Research. The study is

ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Gumperaz, J. J., Cook-Gumperaz, J., & Szymanski, M. H. (1999). *Collaborative practices in bilingual cooperative learning classrooms*. Santa Cruz, CA: Center for Research on Education, Diversity and Excellence. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Jenkins, J. R., Jewell, M., O'Connor, R. E., Jenkins, L. M., & Troutner, N. M. (1994). Accommodations for individual differences without classroom ability groups: An experiment in school restructuring. *Exceptional Children*, 60(4), 344–358. The study does not meet WWC evidence standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

**Additional source:**

Jenkins, J. R., Jewell, M., Leicester, N., Jenkins, L., & Troutner, N. M. (1991). Development of a school building model for educating students with handicaps and at-risk students in general education classrooms. *Journal of Learning Disabilities*, 24(5), 311–320.

Manarino-Leggett, P., & Salomon, P. A. (1989, April). *Cooperation vs. competition: Techniques for keeping your classroom alive but not endangered*. Paper presented at the annual meeting of the International Reading Association, New Orleans, LA. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Mitchell, D. R. (2008). *What really works in special and inclusive education: Using evidence-based teaching strategies*. New York: Routledge. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Mitchell, S., & Wile, N. (2001). *2001 literacy program evaluation: A report of the evaluation of literacy programs in elementary and middle schools*. Portland, OR: Portland Public Schools, Research and Evaluation Department. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Nath, L. R. (1996). A peer tutoring training model for cooperative groupings: Is the effectiveness of cooperative groupings enhanced by students obtaining peer tutoring skills? (Doctoral dissertation, University of Memphis, 1996). *Dissertation Abstracts International*, 57(12A), 127–5051. The study is ineligible for review because it does not use a comparison group design or a single-case design.

O'Connor, R. E., & Jenkins, J. R. (1993). *Cooperative learning as an inclusion strategy: The experience of children with disabilities*. Atlanta, GA: American Educational Research Association. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Porter, P. (1999). Cooperative education—the key to bilingual success? *SEDLetter*, 11(1). The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% general education students.

Prado-Olmos, P., Garcia, R. G., & Duran, R. P. (1991, April). *Cooperative learning for bilingual students: A case study of a CIRC implementation*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Prado-Olmos, P. L., Smith, M. E. F., & Szymanski, M. (1993, April). *Students “DO” process: Bilingual students’ interactions in a small cooperative reading group*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

- Promising Practices Network. (2004). Cooperative Integrated Reading and Composition. Retrieved November 23, 2009, from <http://www.promisingpractices.net/>. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Rapp, J. C. (1992). The effect of cooperative learning on selected student variables (*Cooperative Integrated Reading and Composition*) on academic achievement in reading comprehension, vocabulary and spelling and on student self-esteem (Doctor of education dissertation, Washington State University, 1991). *Dissertation Abstracts International*, 52(10-A), 3516. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Richman, B. S. (1989). *Involvement in learning for low-achieving students*. Philadelphia, PA: Research for Better Schools, Inc. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Risner, G. P., Nicholson, J. I., & Webb, B. (1994, November). *Levels of comprehension promoted by the Cooperative Integrated Reading and Composition (CIRC) program*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Nashville, TN. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Robinson, A. (1991). *Cooperative learning and the academically talented student. Research-based decision making series*. Storrs, CT: National Research Center on the Gifted and Talented. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Schundler, E. T. (1992). *The effect of cooperative learning on comprehension: An analysis of the effect of a modified CIRC instructional approach and cooperative learning partnerships on reading comprehension*. Unpublished master's thesis, Kean University, Union, NJ. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.
- Schutt, A. M. (1990). *Cooperative Integrated Reading and Composition: An evaluative study of the CIRC writing component in Tift County elementary schools*. Unpublished educational specialists' thesis, Valdosta State College, GA. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Skeans, S. E. S. (1991). The effects of *Cooperative Integrated Reading and Composition*, fidelity of implementation, and teacher concerns on student achievement (Doctoral dissertation, Texas A&M University, 1991). *Dissertation Abstracts International*, 53(02A), 211–455. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Slavin, R. E. (1990). Learning together. *American School Board Journal*, 177(8), 22–23. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Slavin, R. E. (1991). *Student team learning: A practical guide to cooperative learning* (3rd ed.). Washington, DC: National Education Association. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Slavin, R. E., & Cheung, A. (2003). *Effective reading programs for English language learners. A best-evidence synthesis* (Report No. 66). Baltimore, MD: Center for Research on the Education of Students Placed At Risk (CRESPAR). The study is ineligible for review because it does not use a sample aligned with the protocol—the sample includes less than 50% general education students.
- Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2009). *Effective programs for struggling readers: A best evidence synthesis*. Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education. The study is ineligible



for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Slavin, R. E., & Madden, N. A. (1999). *Disseminating Success for All: Lessons for policy and practice* (Report No. 30). Baltimore, MD: Center for Research on the Education of Students Placed At Risk (CRESPAR). The study is ineligible for review because it does not examine the effectiveness of an intervention.

Slavin, R. E., & Madden, N. A. (1999). *Roots & Wings: A comprehensive approach to elementary school reform*. Baltimore, MD: Success for All Foundation. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Slavin, R. E., & Madden, N. A. (1999). *Roots & Wings: Effects of whole-school reform on student achievement* (Report No. 36). Baltimore, MD: Center for Research on the Education of Students Placed At Risk (CRESPAR). The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

**Additional source:**

Slavin, R. E., & Madden, N. A. (2000). Roots & Wings: Effects of whole-school reform on student achievement. *Journal of Education for Students Placed at Risk*, 5(1), 109–136.

Slavin, R. E., & Madden, N. A. (2004). *Success for All/Roots & Wings: Summary of research on achievement outcomes* (Report No. 41). Baltimore, MD: Center for Research on the Education of Students Placed At Risk (CRESPAR). The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Slavin, R. E., Madden, N. A., & Stevens, R. J. (1989). Cooperative learning models for the 3 R's. *Educational Leadership*, 47(4), 22. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Steals, M. H. (1989). Development of an adapted cooperative learning strategy in a secondary Chapter I Option 4 reading English/language arts pairing program (Doctoral dissertation, University of Pittsburgh, 1989). *Dissertation Abstracts International*, 51(05A), 203–1564. The study is ineligible for review because it does not use a comparison group design or a single-case design.

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). *Cooperative Integrated Reading and Composition: Two field experiments*. *Reading Research Quarterly*, 22(4), 433–454. The study is ineligible for review because it does not occur within the time frame specified in the protocol.

Stevens, R. J., & Slavin, R. E. (1995). Effects of a cooperative learning approach in reading and writing on academically handicapped and nonhandicapped students. *Elementary School Journal*, 95(3), 241–262. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

**Additional source:**

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1989). *A cooperative learning approach to elementary reading and writing instruction: Long-term effects* (Report No. 42). Baltimore, MD: Center for Research on Elementary and Middle Schools.

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification. *Journal of Educational Psychology*, 83(1), 8–16. The study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

**Additional source:**

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1989). *The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification* (Report No. 44). Baltimore, MD: Center for Research on Elementary and Middle Schools.

Uhry, J. K., & Clark, D. B. (2004). *Dyslexia: Theory & practice of instruction* (3rd ed.). Austin, TX: PRO-ED. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

Urdegar, S. M. (1998). *Evaluation of the Success for All program, 1997–98*. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

**Additional sources:**

Urdegar, S. M. (2000). *Evaluation of the Success for All program, 1998–99*. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research.

Urdegar, S. M. (2001). *Evaluation of the Success for All program, 1999–2000*. Miami, FL: Miami-Dade Public Schools, Office of Evaluation Research.

Yeh, Y. (2007). Elementary students' prior knowledge and the *Cooperative Integrated Reading and Composition (CIRC)* model in second-language reading comprehension (Doctoral dissertation, Fordham University, 2007). *Dissertation Abstracts International*, 68(03A), 291–867. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.

# Appendix

## Appendix A1.1 Study characteristics: Stevens & Slavin, 1995

Characteristic	Description
<b>Study citation</b>	Stevens, R. J., & Slavin, R. E. (1995). The Cooperative Elementary School: Effects on students' achievement, attitudes and social relations. <i>American Educational Research Journal</i> , 32(2), 321–351.
<b>Participants</b>	This study is a quasi-experiment conducted in five schools. Two treatment schools were selected by the investigators to implement the intervention, and three schools, matched on academic achievement, ethnicity, and socioeconomic background, were selected to serve as comparison schools. Classes in grades 2 through 6 in the treatment schools were matched with classes in the comparison schools based on pretest scores on the California Achievement Test (Reading, Language Arts, and Mathematics). The study's analytic sample included 411 students in 21 treatment classrooms and 462 students in 24 comparison classrooms. The study reported students' outcomes after two years of program implementation; these findings were used in the intervention ratings and can be found in Appendices A3.1 and A3.2. Additional findings reflecting students' outcomes after one year of program implementation can be found in Appendices A4.1 and A4.2.
<b>Setting</b>	The study was conducted in five schools in one suburban school district in Maryland. The student populations of each school ranged from 4% to 15% minority students, and from 2% to 20% students received free or reduced-price lunch.
<b>Intervention</b>	Intervention schools implemented the Cooperative Elementary School model, a whole-school reform model that uses cooperative learning strategies across multiple content areas. Teachers used peer coaching and conducted their planning in a cooperative manner. Cooperative Elementary School emphasizes teacher involvement in site-based management and parent involvement in schools. The language arts/reading curriculum within Cooperative Elementary School is <i>Cooperative Integrated Reading and Composition</i> ®. Daily lessons, which focus on story-related activities, direct instruction in reading comprehension, and integrated reading and language arts activities, incorporate team practice, peer assessment, and team/partner recognition. This program was phased in gradually during the first year of the two-year implementation.
<b>Comparison</b>	Reading activities consisted of students working in small reading groups using a basal series, workbooks, worksheets, and activities based on teacher-prepared materials. Language arts activities generally involved whole-class instruction using a published language arts series, as well as teacher-developed activities. Comparison schools did not use structured cooperative learning during classroom instruction, although occasional cooperative activities were used by some of the teachers. Comparison schools implemented some of the components of the Cooperative Elementary School model, but they did not implement <i>Cooperative Integrated Reading and Composition</i> ®.
<b>Primary outcomes and measurement</b>	For both the pretest and the posttest, students took the Vocabulary, Reading Comprehension, Language Expression, and Language Mechanics subtests of the California Achievement Test. Scores were converted to z-scores in order to conduct analyses across the grades included in the study sample (grades 2–6). For a more detailed description of these outcome measures, see Appendices A2.1 and A2.2.
<b>Staff/teacher training</b>	Intervention teachers were trained in <i>Cooperative Integrated Reading and Composition</i> ® prior to implementation. Subsequent trainings were conducted at two-month intervals during the school year. Trainers reviewed with treatment teachers a detailed manual that explains how to implement the program in the classroom. Trainers also provided simulated demonstrations of lessons. In addition, during the school year, members of the research staff observed treatment-group classes, participated in meetings with treatment-group teachers, and observed steering committee meetings in order to facilitate implementation of the program components.

## Appendix A1.2 Study characteristics: Jewell, 1994

Characteristic	Description
<b>Study citation</b>	Jewell, M. E. (1994). The effect of classroom-based follow-up assistance on mainstream reading and language arts instruction (Doctoral dissertation, University of Washington, 1994). <i>Dissertation Abstracts International</i> , 55(11A), 107–3473.
<b>Participants</b>	This study is a quasi-experiment that initially included a sample of 51 second- to sixth-grade classrooms assigned to one of three conditions: (1) comparison; (2) treatment, receiving <i>Cooperative Integrated Reading and Composition</i> <sup>®</sup> training in the summer preceding implementation; and (3) treatment, receiving the program training as well as follow-up support during the school year. <sup>1</sup> The treatment classrooms were matched with comparison classrooms on the Gates–MacGinitie pretest scores. This review focuses on comparisons of the 15 classrooms taught by teachers who received either program training or program training with follow-up support, and the 15 classrooms in the comparison group.
<b>Setting</b>	The study took place in four schools in one district in the United States. The participating elementary schools served 9% to 27% minority students, and less than 15% of the student population received special education services.
<b>Intervention</b>	There were two forms of the <i>Cooperative Integrated Reading and Composition</i> <sup>®</sup> intervention: (1) training only and (2) training plus follow-up support. The intervention group participated in teacher-led basal-related activities, partner reading, story-related writing, reading of words aloud, word meaning activities, story retelling, spelling, direct instruction in reading comprehension, home reading, integrated language arts and writing, weekly tests, and cooperative learning groups of four students. The program was implemented in intervention classrooms for seven to eight months.
<b>Comparison</b>	Comparison group teachers continued to teach in accordance with their own style and used the regular district-adopted reading and language arts program (basal materials). All comparison schools used the same reading and language arts curricula: <i>Houghton Mifflin Reading</i> (Durr et al., 1989) <sup>2</sup> and the <i>Silver Burdett &amp; Ginn English</i> series (Ragno, Toth, & Gray, 1988). <sup>3</sup>
<b>Primary outcomes and measurement</b>	For both the pretest and the posttest, students took the Gates–MacGinitie Reading Test Comprehension and Vocabulary subtests and the Basic Academic Skills Sample Reading Proficiency subtest. For a more detailed description of these outcome measures, see Appendices A2.1 and A2.2.
<b>Staff/teacher training</b>	The study's investigator provided <i>Cooperative Integrated Reading and Composition</i> <sup>®</sup> training and follow-up assistance. The training took place during five 4-hour sessions (spanning one week) during the summer preceding the study. Teachers were provided with lesson plans that were aligned to the district's basal series curriculum. Teachers practiced the <i>Cooperative Integrated Reading and Composition</i> <sup>®</sup> components and received feedback from the investigator and peers. All <i>Cooperative Integrated Reading and Composition</i> <sup>®</sup> -trained teachers participated in two follow-up meetings during the school year. Teachers assigned to receive classroom-based follow-up assistance also had the investigator observe lessons (on average, about 10 observations), provide feedback, demonstrate teaching procedures, and make recommendations for future lessons.

1. Results from these analyses are not included in this report because the treatment groups and the comparison group were not equivalent at baseline.
2. Durr, W. K., Pikulski, J. J., Bean, R. M., Cooper, J. D., Glaser, N. A., Greenlaw, M. J., & Schoephoerster, H. (1989). *Houghton Mifflin Reading*. Boston: Houghton Mifflin.
3. Ragno, N. N., Toth, M. D., & Gray, B. G. (1988). *Silver Burdett & Ginn English*. Morristown, NJ: Silver Burdett & Ginn.

## Appendix A2.1 Outcome measures for the comprehension domain

Outcome measure	Description
<i>Reading comprehension construct</i>	
<b>California Achievement Test (CAT)—Reading Comprehension subtest</b>	The CAT is a norm- and criterion-referenced annual test. The Reading Comprehension subtest is administered to grades 1 through 12 and focuses on students' use of reading comprehension strategies. Passages reflect a wide range of narrative, expository, contemporary, and traditional texts. The test measures information recall, meaning construction, form analysis, and meaning evaluation of seven different selections (as cited in Stevens & Slavin, 1995).
<b>Gates–MacGinitie Reading Tests—Comprehension subtest</b>	The Comprehension subtest of the Gates–MacGinitie Reading Test measures each student's ability to read and understand different types of prose. The test contains 11 passages of various lengths and subjects and 48 questions (as cited in Jewell, 1994).
<i>Vocabulary development construct</i>	
<b>California Achievement Test (CAT)—Vocabulary subtest</b>	The CAT is a norm- and criterion-referenced annual test. The Vocabulary subtest contains 20 items measuring same-meaning, opposite-meaning words; multi-meaning words; words in context; and the meaning of affixes (as cited in Stevens & Slavin, 1995).
<b>Gates–MacGinitie Reading Tests—Vocabulary subtest</b>	The Vocabulary subtest of the Gates–MacGinitie Reading Test measures a student's reading vocabulary. The test contains 45 questions that measure word knowledge by asking students to choose one word or phrase that means most nearly the same as a presented word. The vocabulary test takes 20 minutes to administer (as cited in Jewell, 1994).

## Appendix A2.2 Outcome measures for the general literacy achievement domain

Outcome measure	Description
<b>California Achievement Test (CAT)—Language Mechanics subtest</b>	The Language Mechanics and Language Expression subtests of the CAT work together to measure a broad range of language and writing skills, including the ability to apply standard usage and writing conventions and to develop effective sentences and paragraphs. The Language Mechanics subtest contains 20 items that measure skills in the mechanics of capitalization and punctuation. Editing skills are measured in the context of passages presented in various formats (for grades 4–12) (as cited in Stevens & Slavin, 1995).
<b>California Achievement Test (CAT)—Language Expression subtest</b>	The Language Mechanics and Language Expression subtests of the CAT work together to measure a broad range of language and writing skills, including the ability to apply standard usage and writing conventions and to develop effective sentences and paragraphs. The Language Expression subtest contains 20 items that measure skills in language usage and sentence structure. The items measure skills in the use of various parts of speech, formation and organization of sentences and paragraphs, and writing for clarity (for grades 4–12) (as cited in Stevens & Slavin, 1995).
<b>Basic Academic Skills Sample—Reading Proficiency subtest</b>	The Basic Academic Skills Sample is a group-administered, curriculum-based assessment of students' reading proficiency, measured via maze procedure using three passages averaging 263 words. In this maze procedure, every seventh word in the passage is replaced by a multiple choice question item containing the correct word and two distractors. Students' scores reflect the number of correct multiple choice selections, given one minute per passage (as cited in Jewell, 1994).

## Appendix A3.1 Summary of study findings included in the rating for the comprehension domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Cooperative Integrated Reading and Composition® – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			Cooperative Integrated Reading and Composition® group	Comparison group				
<b>Stevens &amp; Slavin, 1995<sup>7,8</sup></b>								
CAT—Reading Vocabulary	Grades 2–6	45/873	0.10 (0.98)	–0.11 (1.01)	0.21	0.21	Statistically significant	+8
CAT—Reading Comprehension	Grades 2–6	45/873	0.15 (0.98)	–0.13 (1.01)	0.28	0.28	Statistically significant	+11
<b>Average for comprehension (Stevens &amp; Slavin, 1995)<sup>9</sup></b>						<b>0.25</b>	<b>Statistically significant</b>	<b>+10</b>
<b>Jewell, 1994<sup>7,10</sup></b>								
GMRT—Vocabulary subtest	Grades 2–6	30 classrooms	50.48 (21.06)	49.07 (21.06)	1.41	0.07	ns	+3
GMRT—Comprehension subtest	Grades 2–6	30 classrooms	52.20 (21.06)	49.16 (21.06)	3.04	0.14	ns	+6
<b>Average for comprehension (Jewell, 1994)<sup>9</sup></b>						<b>0.11</b>	<b>ns</b>	<b>+4</b>
<b>Domain average for comprehension across all studies<sup>9</sup></b>						<b>0.18</b>	<b>na</b>	<b>+7</b>

ns = not statistically significant

na = not applicable

CAT = California Achievement Test

GMRT = Gates–MacGinitie Reading Test

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.
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.

(continued)

## Appendix A3.1 Summary of study findings included in the rating for the comprehension domain *(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 Stevens and Slavin (1995), correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. A correction for clustering was not needed, as the authors used HLM analyses, which accounted for multi-level data (of students nested within classrooms and schools). In the case of Jewell (1994), no corrections for clustering or multiple comparisons were needed, as analyses were performed at the classroom level and findings were not statistically significant.
8. The intervention group and comparison group mean outcome values for Stevens and Slavin (1995) are the HLM-fitted two-year posttest means.
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. The group mean values reported for Jewell (1994) are the pretest group means plus the gain scores. Standard deviations of 21.06 are from the normative student sample.

## Appendix A3.2 Summary of study findings included in the rating for the general literacy achievement domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Cooperative Integrated Reading and Composition® – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
		Cooperative Integrated Reading and Composition® group	Comparison group					
<b>Stevens &amp; Slavin, 1995<sup>7,8</sup></b>								
CAT—Language Mechanics subtest	Grades 2–6	45/873	0.05 (0.97)	–0.05 (1.02)	0.10	0.10	ns	+4
CAT—Language Expression subtest	Grades 2–6	45/873	0.11 (0.96)	–0.10 (1.03)	0.21	0.21	Statistically significant	+8
<b>Average for general literacy achievement (Stevens &amp; Slavin, 1995)<sup>9</sup></b>						<b>0.16</b>	<b>ns</b>	<b>+6</b>
<b>Jewell, 1994<sup>7,10</sup></b>								
BASS—Reading Proficiency subtest	Grades 2–6	30 classrooms	100.72 (15.00)	101.86 (15.00)	–1.14	–0.08	ns	–3
<b>Average for general literacy achievement (Jewell, 1994)<sup>9</sup></b>						<b>–0.08</b>	<b>ns</b>	<b>–3</b>
<b>Domain average for general literacy achievement across all studies<sup>9</sup></b>						<b>0.04</b>	<b>na</b>	<b>+2</b>

ns = not statistically significant

na = not applicable

CAT = California Achievement Test

BASS = Basic Academic Skills Sample

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the general literacy 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.
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 Stevens and Slavin (1995), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. A correction for clustering was not needed, as the authors used HLM analyses, which accounted for multi-level data (of students nested within classrooms and schools). In the case of Jewell (1994), no corrections for clustering or multiple comparisons were needed.
8. The intervention group and comparison group mean outcome values for Stevens and Slavin (1995) are the HLM-fitted two-year posttest means.
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.
10. The group mean values reported for Jewell (1994) are the pretest group means plus the gain scores. Standard deviations of 15.00 are from the normative student sample.



## Appendix A4.1 Summary of one-year implementation findings for the comprehension domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Cooperative Integrated Reading and Composition® – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			Cooperative Integrated Reading and Composition® group	Comparison group				
<b>Stevens &amp; Slavin, 1995<sup>7,8</sup></b>								
CAT—Reading Vocabulary	Grades 2–6	45/873	0.08 (0.99)	–0.09 (1.01)	0.17	0.17	Statistically significant	+7
CAT—Reading Comprehension	Grades 2–6	45/873	0.08 (1.01)	–0.05 (0.99)	0.13	0.13	ns	+5

ns = not statistically significant

CAT = California Achievement Test

1. This appendix presents one-year findings for measures that fall in the comprehension domain. Two-year findings were used for rating purposes and are presented in Appendix A3.1
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 Stevens and Slavin (1995), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. A correction for clustering was not needed, as the authors used HLM analyses, which accounted for multi-level data (of students nested within classrooms and schools).
8. The intervention group and comparison group mean outcome values for Stevens and Slavin (1995) are the HLM-fitted one-year posttest means.

## Appendix A4.2 Summary of one-year implementation findings for the general literacy achievement domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>3</sup> (Cooperative Integrated Reading and Composition® – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			Cooperative Integrated Reading and Composition® group	Comparison group				
<b>Stevens &amp; Slavin, 1995<sup>7,8</sup></b>								
CAT—Language Mechanics subtest	Grades 2–6	45/873	0.00 (0.99)	0.01 (1.00)	–0.01	–0.01	ns	–0.4
CAT—Language Expression subtest	Grades 2–6	45/873	0.04 (1.01)	–0.04 (0.99)	0.08	0.08	ns	+3

ns = not statistically significant

CAT = California Achievement Test

1. This appendix presents one-year findings for measures that fall in the general literacy achievement domain. Two-year findings 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.
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 Stevens and Slavin (1995), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. A correction for clustering was not needed, as the authors used HLM analyses, which accounted for multi-level data (of students nested within classrooms and schools).
8. The intervention group and comparison group mean outcome values for Stevens and Slavin (1995) are the HLM-fitted one-year posttest means.

## Appendix A5.1 Cooperative Integrated Reading and Composition® 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.<sup>1</sup>

For the outcome domain of comprehension, the WWC rated *Cooperative Integrated Reading and Composition*® 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 *Cooperative Integrated Reading and Composition*® 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, 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.** One study 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 A5.2 Cooperative Integrated Reading and Composition® rating for the general literacy 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.<sup>1</sup>

For the outcome domain of general literacy achievement, the WWC rated *Cooperative Integrated Reading and Composition*® 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 *Cooperative Integrated Reading and Composition*® 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, 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.** One study 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 A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence <sup>1</sup>
		Schools	Students	
Alphabetics	na	na	na	na
Reading fluency	na	na	na	na
Comprehension	2	9	1,460 <sup>2</sup>	Medium to large
General literacy achievement	2	9	1,460 <sup>2</sup>	Medium to large

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 number is an estimate, as the exact number of students is not available for Jewell (1994).