

## Open Court Reading®

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

*Open Court Reading*® is a reading program for grades K–6 published by McGraw-Hill Education that is designed to teach decoding, comprehension, inquiry, and writing in a three-part logical progression. Part One of each unit, Preparing to Read, focuses on phonemic awareness, sounds and letters, phonics, fluency, and word knowledge. Part Two, Reading and Responding, emphasizes reading literature for understanding, comprehension, inquiry, and practical reading applications. Part Three, Language Arts, focuses on writing, spelling, grammar, usage, mechanics, and basic computer skills. In 2007, McGraw-Hill Education revised *Open Court Reading*® and changed the name to *Imagine It!*®. The studies featured in this report evaluate the use of *Open Court Reading*® in grades K–3.

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

The What Works Clearinghouse (WWC) identified two studies of *Open Court Reading*® that both fall within the scope of the Beginning Reading topic area and meet WWC group design standards. One study meets WWC group design standards without reservations, and one study meets WWC group design standards with reservations. Together, these studies included 1,113 beginning readers in grades 1–3 in six states.

The WWC considers the extent of evidence for *Open Court Reading*® on the reading skills of beginning readers to be small for two outcome domains—general reading achievement and comprehension. There were no studies that meet standards in the two other domains, so this intervention report does not report on the effectiveness of *Open Court Reading*® for those domains. (See the Effectiveness Summary on p. 4 for more details of effectiveness by domain.)

### Effectiveness

*Open Court Reading*® was found to have potentially positive effects on general reading achievement and comprehension for beginning readers.

**Table 1. Summary of findings<sup>3</sup>**

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
General reading achievement	Potentially positive effects	+12	na	1	434	Small
Comprehension	Potentially positive effects	+10	na	1	679	Small

na = not applicable

Report Contents	
Overview	p. 1
Program Information	p. 2
Research Summary	p. 3
Effectiveness Summary	p. 4
References	p. 6
Research Details for Each Study	p. 20
Outcome Measures for Each Domain	p. 23
Findings Included in the Rating for Each Outcome Domain	p. 24
Endnotes	p. 27
Rating Criteria	p. 28
Glossary of Terms	p. 29

### Program Information

#### Background

*Open Court Reading*® is published by McGraw-Hill Education (formerly SRA/McGraw-Hill). The program was originally developed in the 1960s and has undergone several revisions and updates over the years. In 2007, McGraw-Hill Education revised the program and changed the name to *Imagine It!*®. Address: McGraw-Hill Education, P.O. Box 182605, Columbus, OH 43218. Website: <https://www.mheonline.com/>. Telephone: (800) 338-3987.

#### Program details

*Open Court Reading*® materials are divided by grade and include the Reading, Phonemic Awareness and Phonics Kit (K); Sounds and Letters Workbook (K); Language Arts Skills Workbook (K); Big Books and Little Books (K–1); Language Arts Big Book (K–1); Pre-Decodable and Decodable Texts (K–3); Part 1 Lesson Cards (K–3); Desk Strips (K–3); Unit Assessment (K–6); Transparencies (K–6); Writer’s Workbook (K–6); Challenge Workbooks (K–6); Reteach Workbooks (K–6); Intervention Support (K–6); Phonics Skills Workbook (1); First and Second Readers (1–2); Reading and Phonics Package (1–3); Student Anthologies (1–6); Comprehension and Language Arts Workbook (1–6); Spelling and Vocabulary Skills Workbook (1–6); Inquiry Journal (2–6); and Language Arts Handbook (2–6). The Teacher’s Edition (K–6) contains information on providing systematic, explicit skills instruction centered on literature selections. Lesson plans indicate the goals and objectives for each lesson and provide detailed suggestions for implementation.

*Open Court Reading*® was revised and renamed *Imagine It!*® in 2007. Program revisions include increased instruction in vocabulary, writing, and inquiry; stronger support for English learners; and enhanced technology components.

#### Cost

The *Open Court Reading*® curriculum includes grade-specific student textbooks, workbooks, decodable books, and anthologies. *Open Court Reading*® Online Professional Development provides support for teachers, principals, reading specialists, and coaches. For details on specific product pricing, contact McGraw-Hill Education, the program publisher.

## Research Summary

The WWC identified 185 studies that investigated the effects of *Open Court Reading*® on the reading skills of beginning readers.

The WWC reviewed 29 of those studies against group design standards. One study (Borman, Dowling, & Schneck, 2008) is a randomized controlled trial that meets WWC group design standards without reservations, and one study (Skindrud & Gersten, 2006) uses a quasi-experimental design that meets WWC group design standards with reservations. Those two studies are summarized in this report. Twenty-seven studies do not meet WWC group design standards. One study does not meet WWC single-case design standards. The remaining 155 studies do not meet WWC eligibility criteria for review in this topic area. Citations for all 185 studies are in the References section, which begins on p. 6.

**Table 2. Scope of reviewed research**

<b>Grade</b>	1, 2, 3
<b>Delivery method</b>	Whole class
<b>Program type</b>	Curriculum

### Summary of study meeting WWC group design standards without reservations

Borman et al. (2008) conducted a randomized controlled trial that examined the effects of *Open Court Reading*® on first- through fifth-grade students attending five schools in five states during the 2005–06 school year. At each school, classrooms were randomly assigned within each grade either to implement *Open Court Reading*® or to serve as the comparison group. The WWC based its effectiveness rating on findings from 679 students from grades 1–3 who participated in the study; 300 students in the *Open Court Reading*® group and 379 in the comparison group.<sup>4</sup> The study reported student outcomes after approximately 7 months of program implementation.

### Summary of study meeting WWC group design standards with reservations

Skindrud and Gersten (2006) examined the effects of *Open Court Reading*® on second- through fourth-grade students attending 12 schools in the Sacramento City Unified School District during the 1997–98 and 1998–99 school years. Four schools implementing *Success for All*® were matched to eight schools that implemented *Open Court Reading*®. Schools were matched by poverty level as measured by the percent of students eligible for free or reduced-price meals and percent of students on Aid to Families with Dependent Children. The WWC based its effectiveness rating on findings from 434 students from grades 2–3 who participated in the study; 292 in the *Open Court Reading*® group and 142 in the comparison group.<sup>5</sup> The study reported student outcomes at two points in time: at the end of second grade and at the end of third grade. Findings at the end of third grade reflect maximum exposure to the intervention by students and were used to determine the rating of effectiveness.<sup>6</sup>

**Effectiveness Summary**

The WWC review of *Open Court Reading*® for the Beginning Reading topic area includes student outcomes in four domains: general reading achievement, comprehension, alphabetics, and reading fluency. The two studies of *Open Court Reading*® that meet WWC group design standards reported findings in two of the four domains: (a) general reading achievement and (b) comprehension. The findings below present the authors’ estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Open Court Reading*® on beginning readers. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 28.

**Summary of effectiveness for the general reading achievement domain**

One study that meets WWC group design standards with reservations reported findings in the general reading achievement domain.

Skindrud and Gersten (2006) found a statistically significant positive effect of *Open Court Reading*® on the Stanford Achievement Test, 9th Edition (SAT-9). However, a correction for clustering was needed and resulted in a WWC-computed *p*-value of .30; therefore, the WWC does not find the result to be statistically significant. The WWC characterizes these study findings as a substantively important positive effect.

Thus, for the general reading achievement domain, one study showed substantively important positive effects. This results in a rating of potentially positive effects, with a small extent of evidence.

**Table 3. Rating of effectiveness and extent of evidence for the general achievement domain**

Rating of effectiveness	Criteria met
<b>Potentially positive effects</b> <i>Evidence of a positive effect with no overriding contrary evidence.</i>	In the one study that reported findings, the estimated impact of the intervention on outcomes in the <i>general reading achievement</i> domain was positive and substantively important.
Extent of evidence	Criteria met
<b>Small</b>	One study that included 434 students in 12 schools reported evidence of effectiveness in the <i>general reading achievement</i> domain.

**Summary of effectiveness for the comprehension domain**

One study that meets WWC group design standards without reservations reported findings in the comprehension domain.

Borman et al. (2008) reported positive effects of *Open Court Reading*® on the Reading Composite score of the Comprehensive Test of Basic Skills, 5th Edition (CTBS/5) Terra Nova test for students in grades 1, 2, and 3. Although the authors did not calculate the statistical significance of the effects for the sample that aligns with the WWC’s Beginning Reading protocol, consisting of students in grades 1 through 3 only, the average effect size was large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least .25). The WWC characterizes these study findings as a substantively important positive effect.

Thus, for the comprehension domain, one study showed substantively important positive effects. This results in a rating of potentially positive effects, with a small extent of evidence.

**Table 4. Rating of effectiveness and extent of evidence for the comprehension domain**

Rating of effectiveness	Criteria met
<b>Potentially positive effects</b> <i>Evidence of a positive effect with no overriding contrary evidence.</i>	In the one study that reported findings, the estimated impact of the intervention on outcomes in the <i>comprehension</i> domain was positive and substantively important.
Extent of evidence	Criteria met
<b>Small</b>	One study that included 679 students in five schools reported evidence of effectiveness in the <i>comprehension</i> domain.

### References

#### Study that meets WWC group design standards without reservations

Borman, G. D., Dowling, N. M., & Schneck, C. (2008). A multisite cluster randomized field trial of Open Court Reading. *Educational Evaluation & Policy Analysis, 30*(4), 389–407.

#### Study that meets WWC group design standards with reservations

Skindrud, K., & Gersten, R. (2006). An evaluation of two contrasting approaches for improving reading achievement in a large urban district. *Elementary School Journal, 106*(5), 389–407.

#### Studies that do not meet WWC group design standards

Algozzine, B., Wang, C., White, R., Cooke, N., Marr, M. B., Algozzine, K., ... Duran, G. Z. (2012). Effects of multi-tier academic and behavior instruction on difficult-to-teach students. *Exceptional Children, 79*(1), 45–64. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.

Carpenter, Z. A. (2005). *Effects of Fast ForWord on reading comprehension for elementary students*. Cheney: Eastern Washington University. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.

Chard, D. J. (1997). *Final evaluation report AY 1996–97: Early reading intervention project, Springfield Public Schools, Springfield, Massachusetts*. Austin: University of Texas at Austin. Retrieved from <http://www.eduplace.com> The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Foorman, B. R., Chen, D. T., Carlson, C., Moats, L., Francis, D. J., & Fletcher, J. M. (2003). The necessity of the alphabetic principle to phonemic awareness instruction. *Reading and Writing, 16*(4), 289–324. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the effects are not reported separately for the intervention.

Foorman, B. R., Fletcher, J. M., Francis, D. J., Mehta, P., & Schatschneider, C. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*(1), 37–55. The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Foorman, B. R., Francis, D. J., Beeler, T., Winikates, D., & Fletcher, J. M. (1997). Early interventions for children with reading problems: Study design and preliminary findings. *Learning Disabilities: A Multi-Disciplinary Journal, 8*(1), 63–71. The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Haager, D., Dhar, R., Moulton, M., & Varma, S. (2005). *Reading First year 3 evaluation report*. Retrieved from <http://www.eddata.com> The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.

#### **Additional sources:**

Haager, D., Dhar, R., Moulton, M., & McMillan, S. (2006). *Reading First year 4 evaluation report*. Retrieved from <http://www.eddata.com>

Haager, D., Dhar, R., Moulton, M., & McMillan, S. (2008). *Reading First year 5 evaluation report*. Retrieved from <http://www.eddata.com>

Hayes, L. L. (2004). *A comparison of two systematic approaches to phonics and spelling instruction in beginning reading: A basal phonics program and word study* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3131455) The study does not meet WWC group design standards

because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

- Kamps, D., Abbott, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., ... Walton, C. (2007). Use of evidence-based, small-group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly*, 30(3), 153–168. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.
- Kamps, D., Abbott, M., Greenwood, C., Wills, H., Veerkamp, M., & Kaufman, J. (2008). Effects of small-group reading instruction and curriculum differences for students most at risk in kindergarten. *Journal of Learning Disabilities*, 41(2), 101–114. The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Kerr, J. M. (2001). *The development of phonological awareness in African American inner-city kindergarten students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3009025) The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.
- McRae, D. J. (2002). *Test score gains for Open Court schools in California*. DeSoto, TX: SRA/McGraw-Hill. The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Moustafa, M., & Land, R. (2001). The effectiveness of Open Court on improving the reading of economically disadvantaged children. *AERA Yearbook*, 44–53. <http://files.eric.ed.gov/fulltext/ED459421.pdf>. The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- O'Brien, D. M., & Ware, A. M. (2002). Implementing research-based reading programs in the Fort Worth Independent School District. *Journal of Education for Students Placed at Risk*, 7(2), 167–195. The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- O'Connor, R. E., Fulmer, D., Harty, K. R., & Bell, K. M. (2005). Layers of reading intervention in kindergarten through third grade: Changes in teaching and student outcomes. *Journal of Learning Disabilities*, 38(5), 440–455. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.
- O'Connor, R. E., Harty, K. R., & Fulmer, D. (2005). Tiers of intervention in kindergarten through third grade. *Journal of Learning Disabilities*, 38(6), 532–538. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.
- Parkman, L. L. (2011). *An investigation of the impact of early intervention reading programs on the academic achievement of third-grade students in reading* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3421347) The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Schacter, J. (2003). Preventing summer reading declines in children who are disadvantaged. *Journal of Early Intervention*, 26(1), 47–58. The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Schacter, J., & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading*, 28(2), 158–169. The study does not meet WWC group design standards because the

measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.

- Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Santoro, L. E., ... Kaufman, N. K. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design specificity. *Journal of Learning Disabilities, 40*(4), 331–347. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the effects are not reported separately for the intervention.
- Stockard, J. (2010). Promoting reading achievement and countering the “fourth-grade slump”: The impact of Direct Instruction on reading achievement in fifth grade. *Journal of Education for Students Placed at Risk, 15*(3), 218–240. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—the intervention was combined with another intervention.
- Stockard, J. (2011). Direct Instruction and first grade reading achievement: The role of technical support and time of implementation. *Journal of Direct Instruction, 11*, 31–50. The study does not meet WWC group design standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.
- Stockard, J., & Engelmann, K. (2010). The development of early academic success: The impact of Direct Instruction's Reading Mastery. *Journal of Behavior Assessment & Intervention in Children, 1*(1), 2–24. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.
- Webster, R. E., & Braswell, L. A. (1991). Curriculum bias and reading achievement test performance. *Psychology in the Schools, 28*(3), 193–199. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.
- Wills, H., Kamps, D., Abbott, M., Bannister, H., & Kaufman, J. (2010). Classroom observations and effects of reading interventions for students at risk for emotional and behavioral disorders. *Behavioral Disorders, 35*(2), 103–119. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.
- Wilson, G. P., Martens, P., Arya, P., & Altwerger, B. (2004). Readers, instruction, and the NRP. *Phi Delta Kappan, 86*(3), 242–246. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.
- Wiltz, N., & Wilson, G. P. (2006). An inquiry into children's reading in one urban school using SRA Reading Mastery (Direct Instruction). *Journal of Literacy Research, 37*(4), 493–528. The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention—there was only one unit assigned to one or both conditions.

### Study that does not meet WWC pilot single-case design standards

- Lane, K. L., Little, M. A., Redding-Rhodes, J., Phillips, A., & Welsh, M. T. (2007). Outcomes of a teacher-led reading intervention for elementary students at risk for behavioral disorders. *Exceptional Children, 74*(1), 47–70. The study does not meet WWC pilot single-case design standards because it does not have at least three attempts to demonstrate an intervention effect at three different points in time.

### Studies that are ineligible for review using the Beginning Reading Evidence Review Protocol

- Adams, M. J., Bereiter, C., McKeough, A., Case, R., Roit, M., Hirschberg, J., & Treadway, G. H., Jr. (2002). *Open Court Reading*. Columbus, OH: McGraw-Hill. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Ajayi, L. (2005). Teachers' needs and predesigned instructional practices: An analysis of a reading/language arts coursebook for a second grade class. *Reading Improvement, 42*(4), 200–211. The study is ineligible for review because it does not examine the effectiveness of an intervention.



- Allen, M. M., Ukrainetz, T. A., & Carswell, A. L. (2012). The narrative language performance of three types of at-risk first-grade readers. *Language, Speech, and Hearing Services in Schools, 43*(2), 205–221. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.
- Allor, J. H., Mathes, P. G., Roberts, J. K., Cheatham, J. P., & Champlin, T. M. (2010). Comprehensive reading instruction for students with intellectual disabilities: Findings from the first three years of a longitudinal study. *Psychology in the Schools, 47*(5), 445–466. The study is ineligible for review because it does not use a sample aligned with the protocol.
- Altwerger, B. (2005). *Reading for profit: How the bottom line leaves kids behind*. Portsmouth, NH: Heinemann. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Altwerger, B., Arya, P., Jin, L., Jordan, N. L., Laster, B., Martens, P., & Wiltz, N. (2004). When research and mandates collide: The challenges and dilemmas of teacher education in the era of NCLB. *English Education, 36*(2), 119. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Alvarez, L., & Corn, J. (2008). Exchanging assessment for accountability: The implications of high-stakes reading assessments for English learners. *Language Arts, 85*(5), 354–365. The study is ineligible for review because it does not use a sample aligned with the Beginning Reading evidence review protocol.
- American Federation of Teachers. (1999). *Building on the best, learning from what works: Five promising remedial reading intervention programs*. Washington, DC: Author. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Association for Supervision and Curriculum Development, Council of Chief State School Officers, & McGraw-Hill Education. (2005). *Results with Open Court*. Columbus, OH: McGraw-Hill Education. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Barrett, A. J. (2007). *Diffusion of innovation: A study of the impact of professional development practices on coherent implementation of Open Court Reading in the Los Angeles Unified School District* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3295740) The study is ineligible for review because it does not include a student outcome.
- Barrett, T. J. (1995, November). *A comparison of two approaches to first grade phonics instruction in the Riverside Unified School District*. Paper presented at the annual meeting of the California Educational Research Association, Lake Tahoe, CA. The study is ineligible for review because it does not occur within the time frame specified in the protocol.
- Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review, 40*(1), 23–38. The study is ineligible for review because it does not include a student outcome.
- Begeny, J. C., Ross, S. G., Greene, D. J., Mitchell, R. C., & Whitehouse, M. H. (2012). Effects of the Helping Early Literacy with Practice Strategies (HELPS) reading fluency program with Latino English language learners: A preliminary evaluation. *Journal of Behavioral Education, 21*(2), 134–149. The study is ineligible for review because it does not use a sample aligned with the protocol.
- Bodilly, S. J., Chun, J., Ikemoto, G., & Stockly, S. (2004). *Challenges and potential of a collaborative approach to education reform*. Santa Monica, CA: RAND. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Bohrstedt, G., Poduska, J., Garet, M., Kellam, S., & Myers, D. (2007, April). *Designing and implementing randomized field trials of interventions in schools*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Brenner, D., & Hiebert, E. H. (2010). If I follow the teachers' editions, isn't that enough? Analyzing reading volume in six core reading programs. *The Elementary School Journal, 110*(3), 358–363. The study is ineligible for review because it does not include a student outcome.

- 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. <http://files.eric.ed.gov/fulltext/ED420046.pdf>. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Campbell, M. L., Helf, S., & Cooke, N. L. (2008). Effects of adding multisensory components to a supplemental reading program on the decoding skills of treatment resisters. *Education & Treatment of Children*, 31(3), 267–295. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Cannon, J., & Karoly, L. A. (1997) *Who is ahead and who is behind? Gaps in school readiness and student achievement in the early grades for California's children*. Santa Monica, CA: RAND Corporation. Retrieved from <http://www.rand.org> The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Chambers, B., Cheung, A., Slavin, R. E., Smith, D., & Laurenzano, M. (2010). *Effective early childhood education programs: A systematic review*. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education. <http://files.eric.ed.gov/fulltext/ED527643.pdf>. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Chao, Y. J., & Lo, H. (2011). Students' perceptions of wiki-based collaborative writing for learners of English as a foreign language. *Interactive Learning Environments*, 19(4), 395–411. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Cheatham, J. P., & Allor, J. H. (2012). The influence of decodability in early reading text on reading achievement: A review of the evidence. *Reading and Writing*, 25(9), 2223–2246. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Coles, G. (2000). "Direct, explicit, and systematic"—Bad reading science. *Language Arts*, 77(6), 543–545. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Colvin, L. (2002). *Getting it right and making a difference: The instructional leadership strategies and programmatic materials that raise student achievement in low-socioeconomic urban elementary schools* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3071243) 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.
- Cowan, B. L. (1987). *A comparison of the vocabulary scores of fourth-grade students receiving basal reader instruction with computer-assisted reinforcement, with paper-and-pencil reinforcement, and with no reinforcement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8810933) 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.
- 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 is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Dormashev, Y. (2010). Flow experience explained on the grounds of an activity approach to attention. In B. Bruya (Ed.), *Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 287–333). Cambridge, MA: MIT Press. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Eastman, D. N. (2012). *Investigating the influence of the Open Court language arts curriculum on standardized test scores* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3514595) The study is ineligible for review because it does not use a comparison group design or a single-case design.

- Eckhoff, B. L. (1985). *How basal reading texts affect children's writing* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8523323) The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Erneling, C. E. (2010). *Towards discursive education: Philosophy, technology, and modern education*. New York: Cambridge University Press. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start children. *American Educational Research Journal*, 48(3), 763–793. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Florida Center for Reading Research. (2004). *Open Court Reading pre-K*. Tallahassee, FL: Author. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Foorman, B. R., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology*, 31(1), 1–29. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Ge, X., & Hardre, P. L. (2010). Self-processes and learning environment as influences in the development of expertise in instructional design. *Learning Environments Research*, 13(1), 23–41. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Gersten, R., Dimino, J., Jayanthi, M., Kim, J. S., & Santoro, L. E. (2010). Teacher study group: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal*, 47(3), 694–739. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Gillam, R. B., & Carlile, R. M. (1997). Oral reading and story retelling of students with specific language impairment. *Language, Speech, and Hearing Services in Schools*, 28(1), 30–42. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Griffin, N. C., Silk, Y., Chow, K. A., & Chai, Y. (2011). *Evaluation of the artist-teacher collaborative program: Summary of year 1 findings*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing. <http://files.eric.ed.gov/fulltext/ED520426.pdf>. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Henderson, H. (2006). *Let's kill Dick & Jane: How the Open Court Publishing Company fought the culture of American education*. South Bend, IN: St. Augustine's Press. The study is ineligible for review because it does not include a student outcome.
- Heydon, R., & Stooke, R. (2012). Border work: Teachers' expressions of their literacy-related professional development needs in a First Nations school. *Teaching and Teacher Education*, 28(1), 11–20. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Hiebert, E. H. (2010). Understanding the word-level features of texts for students who depend on schools to become literate. In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 207–231). New York: Guilford Press. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Hiebert, E. H., & Fisher, C. W. (2007). Critical word factor in texts for beginning readers. *Journal of Educational Research*, 101(1), 3–11. The study is ineligible for review because it does not include a student outcome.
- Hiebert, E. H., Martin, L. A., & Menon, S. (2005). Are there alternatives in reading textbooks? An examination of three beginning reading programs. *Reading & Writing Quarterly*, 21(1), 7–32. The study is ineligible for review because it does not include a student outcome.

- Hong, H., Chen, F., Chai, C. S., & Chan, W. (2011). Teacher-education students' views about knowledge building theory and practice. *Instructional Science*, 39(4), 467–482. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Howland, B. A. (2009). *Effects of individualized professional development on the theoretical understandings and instructional practices of teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3355457) The study is ineligible for review because it does not include a student outcome.
- Hoxby, C. M., & Murarka, S. (2008). New York City charter schools: How well are they teaching their students? *Education Next*, 8(3), 54–61. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Humphrey, D., & Wechsler, M. (2007). Insights into alternative certification: Initial findings from a national study. *The Teachers College Record*, 109(3), 483–530. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Iman, J. A. (2009). *The influence of direct vocabulary instruction in reading proficiency in kindergarten and first grade* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3370191) The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Izumi, L. T., Coburn, K. G., & Cox, M. (2002). *They have overcome: High-poverty, high-performing schools in California*. San Francisco: Pacific Research Institute for Public Policy. <http://files.eric.ed.gov/fulltext/ED469963.pdf>. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Jolley, R. P. (2010). *Children and pictures: Drawing and understanding*. Hoboken, NJ: Wiley-Blackwell. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Jordan, N. L. (2005). Basal readers and reading as socialization: What are children learning? *Language Arts*, 82(3), 204–213. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Kallas, L. A. (2002). *Does a popular reading program truly meet the needs of second language learners?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 11411341) The study is ineligible for review because it does not use a sample aligned with the protocol.
- Kamps, D. M., & Greenwood, C. R. (2005). Formulating secondary-level reading interventions. *Journal of Learning Disabilities*, 38(6), 500–509. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Kek, M. Y. C. A., & Huijser, H. (2011). The power of problem-based learning in developing critical thinking skills: Preparing students for tomorrow's digital futures in today's classrooms. *Higher Education Research & Development*, 30(3), 329–341. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Kent, S. C., Wanzek, J., & Al Otaiba, S. (2012). Print reading in general education kindergarten classrooms: What does it look like for students at-risk for reading difficulties? *Learning Disabilities Research & Practice*, 27(2), 56–65. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Kim, Y. (2012). The relations among L1 (Spanish) literacy skills, L2 (English) language, L2 text reading fluency, and L2 reading comprehension for Spanish-speaking ELL first grade students. *Learning and Individual Differences*, 22(6), 690–700. The study is ineligible for review because it does not use a sample aligned with the Beginning Reading evidence review protocol.
- Kim, Y., Al Otaiba, S., Puranik, C., Folsom, J. S., Greulich, L., & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences*, 21(5), 517–525. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- King, R., & Torgesen, J. K. (2006). Improving the effectiveness of reading instruction in one elementary school: A description of the process. In P. Blaunstein & R. Lyon (Eds.), *It doesn't have to be this way* (pp. 1–26). Lanham, MD: Scarecrow Press. The study is ineligible for review because it does not use a comparison group design or a single-case design.

- Lauer, P., Akiba, M., Wilkerson, S., Aphthorp, H., Snow, D., & Martin-Glenn, M. (2004). *Effectiveness of out-of-school-time strategies in assisting low-achieving students in reading and mathematics: A research synthesis*. Washington, DC: Institute of Education Sciences. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- LeapFrog SchoolHouse joins print-based curriculum as ERF choice. (2007). *Electronic Education Report*, 14(18), 5–6. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Lee, S. K., Ajayi, L., & Richards, R. (2007). Teachers' perceptions of the efficacy of the Open Court program for English proficient and English language learners. *Teacher Education Quarterly*, 34(3), 19–33. <http://files.eric.ed.gov/fulltext/EJ795167.pdf>. The study is ineligible for review because it does not include a student outcome.
- Levin, J., Haertel, E., Kirst, M., & Williams, T. (2006). *Similar students, different results: Why do some schools do better? Additional findings: Elementary school curriculum program and API: A more detailed examination*. Mountain View, CA: EdSource. <http://files.eric.ed.gov/fulltext/ED491575.pdf>. 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.
- Lo, Y., Cooke, N. L., & Starling, A. L. P. (2011). Using a repeated reading program to improve generalization of oral reading fluency. *Education & Treatment of Children*, 34(1), 115–140. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Macken-Horarik, M., & Morgan, W. (2011). Towards a metalanguage adequate to linguistic achievement in post-structuralism and English: Reflections on voicing in the writing of secondary students. *Linguistics and Education*, 22(2), 133–149. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Maddahian, E. (2002). *A comparative study of second grade students' reading, language, and spelling gain scores for the Los Angeles Unified School District reading programs*. Los Angeles, CA: Los Angeles Unified School District, Program Evaluation and Research Branch. <http://files.eric.ed.gov/fulltext/ED465793.pdf>. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Manzo, K. K. (2004a). L.A. students get reading by the book. *Education Week*, 24(3), 1–18. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Manzo, K. K. (2004b). Leading commercial series don't satisfy "gold standard". *Education Week*, 24(3), 16–17. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Manzo, K. K. (2004c). Page turner. *Teacher Magazine*, 16(3), 9–10. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Manzo, K. K. (2006). Reading program benefits some Calif. schools. *Education Week*, 25(34), 9. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Martens, P., & Goodman, Y. (2005). The influence of decodable texts on readers' strategies. In B. Altwerger (Ed.), *Reading for profit: How the bottom line leaves kids behind* (pp. 78–95). Portsmouth, NH: Heinemann. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Martens, B. K., Daly, E. J., Begeny, J. C., & VanDerHeyden, A. (2011). Behavioral approaches to education. In W. Fisher, C. Piazza, & H. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 385–401). New York: Guilford Press. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- McDougal, J. L., Graney, S. B., Wright, J. A., & Ardoin, S. P. (2009). *RTI in practice: A practical guide to implementing effective evidence-based interventions in your school*. New York: John Wiley & Sons. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- McGraw-Hill Companies, Business Roundtable, & National Association of Elementary School Principals. (2002). *Results with Open Court Reading*. Columbus, OH: McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.

- McGraw-Hill Education. (2002a). *Results with Open Court Reading*. (Study: Canopy Oaks Elementary, Tallahassee, FL). New York: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED464189.pdf> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McGraw-Hill Education. (2002b). *Results with Open Court Reading*. (Study: Curtis Creek School District, Sonora, CA). New York: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED464189.pdf> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McGraw-Hill Education. (2002c). *Results with Open Court Reading*. (Study: Fort Worth Independent School District, Fort Worth, TX). New York: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED464189.pdf> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McGraw-Hill Education. (2002d). *Results with Open Court Reading*. (Study: Hartsfield Elementary School, Tallahassee, FL). New York: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED464189.pdf> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McGraw-Hill Education. (2002e). *Results with Open Court Reading*. (Study: Kelso Elementary School, Inglewood, CA). New York: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED464189.pdf> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McGraw-Hill Education. (2002f). *Results with Open Court Reading*. (Study: Lemoore Union Elementary School District, Lemoore, CA). New York: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED464189.pdf> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McGraw-Hill Education. (2002g). *Results with Open Court Reading*. (Study: Public School 161, Crown Heights, Brooklyn, NY). New York: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED464189.pdf> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McGraw-Hill Education. (2002h). *Results with Open Court Reading*. (Study: Sacramento City Unified School District, Sacramento, CA). New York: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED464189.pdf> The study is ineligible for review because it does not use a comparison group design or a single-case design.
- McRae, D. J. (2002). *Research findings 2002*. DeSoto, TX: SRA/McGraw-Hill. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Miners, Z. (2007). Open Court Reading program: A Florida district NCLB success. *District Administration*, 43(3), 24. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Mitchell, S., Wile, N., & Portland Public Schools Research and Evaluation Dept. (2002). *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. <http://files.eric.ed.gov/fulltext/ED466634.pdf>. The study is ineligible for review because it does not include a student outcome.
- Mora, E. R. (2002). *Spanish speakers learning to read in English-only classrooms: Language policy beginning reading instruction, instructional strategies, and perseverance* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3045603) The study is ineligible for review because it does not use a sample aligned with the Beginning Reading evidence review protocol.
- Morris, J. B. (2002). *The role of literacy coaches in implementing research based reading programs at low achieving schools* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3045590) The study is ineligible for review because it does not examine the effectiveness of an intervention.
- National Center for Education Research. (2008). *Effects of preschool curriculum programs on school readiness: Report from the Preschool Curriculum Evaluation Research initiative*. Washington, DC: Institute of Education Sciences. 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.
- Newkirk, T. (2002). Reading and the limits of science. *Education Week*, 21(32), 39. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.

- O'Connor, R. E., Bocian, K., Beebe-Frankenberger, M., & Linklater, D. L. (2010). Responsiveness of students with language difficulties to early intervention in reading. *The Journal of Special Education, 43*(4), 220–235. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Ogura, P., Coco, L., & Bulat, J. (2007). Using innovative technology to foster reading development among young children with severe cognitive impairments. *Teaching Exceptional Children Plus, 4*(1), 2–13. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Oliver, D., & Maddahian, E. (2002). *K-3 district reading plan evaluation: Year 2 report*. Los Angeles, CA: Los Angeles Unified School District, Program Evaluation and Research Branch. <http://files.eric.ed.gov/fulltext/ED472492.pdf>. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Ortiz, M., Folsom, J. S., Al Otaiba, S., Greulich, L., Thomas-Tate, S., & Connor, C. M. (2012). The componential model of reading: Predicting first grade reading performance of culturally diverse students from ecological, psychological, and cognitive factors assessed at kindergarten entry. *Journal of Learning Disabilities, 45*(5), 406–417. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Pandya, J. Z. (2012). Mandating and standardizing the teaching of critical literacy skills: A cautionary tale. *Theory Into Practice, 51*(1), 20–26. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Pascopella, A. (2004). Struggling English language learner might be at advantage. *District Administration, 40*(4), 21. The study is ineligible for review because it does not use a sample aligned with the Beginning Reading evidence review protocol.
- Pease-Alvarez, L., & Samway, K. D. (2008). Negotiating a top-down reading program mandate: The experiences of one school. *Language Arts, 86*(1), 32–41. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Pease-Alvarez, L., Samway, K., & Cifka-Herrera, C. (2010). Working within the system: Teachers of English learners negotiating a literacy instruction mandate. *Language Policy, 9*(4), 313–334. The study is ineligible for review because it does not use a sample aligned with the Beginning Reading evidence review protocol.
- Peck, S., & Serrano, A. M. (2002, April). *Open Court and English language learners: Questions and strategies*. Paper presented at the annual meeting of the American Association for Applied Linguistics, Salt Lake City, UT. <http://files.eric.ed.gov/fulltext/ED473081.pdf>. The study is ineligible for review because it does not use a sample aligned with the Beginning Reading evidence review protocol.
- Piasta, S. B., Purpura, D. J., & Wagner, R. K. (2010). Fostering alphabet knowledge development: A comparison of two instructional approaches. *Reading and Writing, 23*(6), 607–626. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Pilonieta, P. (2006). *Genre and comprehension strategies presented in elementary basal reading programs: A content analysis* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3228164) The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Pilonieta, P. (2010). Instruction of research-based comprehension strategies in basal reading programs. *Reading Psychology, 31*(2), 150–175. The study is ineligible for review because it does not include a student outcome.
- Presley, E. (2008). *A content analysis of the two state-approved kindergarten and first grade reading programs in California* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3345282) The study is ineligible for review because it does not include a student outcome.
- Puhalla, E. M. (2011). Enhancing the vocabulary knowledge of first-grade children with supplemental booster instruction. *Remedial and Special Education, 32*(6), 471–481. The study is ineligible for review because it does not examine the effectiveness of an intervention.

- Puranik, C., & Al Otaiba, S. (2012). Examining the contribution of handwriting and spelling to written expression in kindergarten children. *Reading and Writing, 25*(7), 1523–1546. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Putnam, L. R., & Farber, F. (1986). Evaluation of oral responses of urban first graders in five different reading programs. *Journal of Clinical Reading: Research and Programs, 2*(1), 18–22. The study is ineligible for review because it does not occur within the time frame specified in the protocol.
- Rivera, M. O., Al Otaiba, S., & Koorland, M. A. (2006). Reading instruction for students with emotional and behavioral disorders and at risk of antisocial behaviors in primary grades: Review of literature. *Behavioral Disorders, 31*(3), 323–339. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Rosenshine, B. (2002). Helping students from low-income homes read at grade level. *Journal of Education for Students Placed at Risk, 7*(2), 273–283. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Ross, J. A., Scott, G., & Sibbald, T. M. (2012). Student achievement outcomes comprehensive school reform: A Canadian case study. *The Journal of Educational Research, 105*(2), 123–133. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Saji, N., Imai, M., Saalbach, H., Zhang, Y., Shu, H., & Okada, H. (2011). Word learning does not end at fast-mapping: Evolution of verb meanings through reorganization of an entire semantic domain. *Cognition, 118*(1), 48–64. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Salasin, E. K. (2005). *“Someone should know where we’re going”*: Teacher negotiation of mandated literacy reform (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3190862) The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Samway, K. D., & Pease-Alvarez, L. (2005). Teachers’ perspectives on Open Court. In B. Altwerger (Ed.), *Reading for profit: How the bottom line leaves kids behind* (pp. 142–155). Portsmouth, NH: Heinemann. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Schacter, J. (1999). *Reading programs that work: A review of programs for pre-kindergarten to 4th grade*. Santa Monica, CA: Milken Family Foundation. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Schacter, J. (2001). Reading programs that work: An evaluation of kindergarten-through-third-grade reading instructional programs. *ERS Spectrum, 19*(4), 12–25. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Simmons, D. C., Coyne, M. D., Oi-man, K., McDonagh, S., Harn, B. A., & Kame’enui, E. J. (2008). Indexing response to intervention. *Journal of Learning Disabilities, 41*(2), 158–173. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Slavin, R. E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly, 43*(3), 290–322. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Slavin, R. E., Lake, C., Cheung, A., & Davis, S. (2009). *Beyond the basics: Effective reading programs for the upper elementary grades*. Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education. <http://files.eric.ed.gov/fulltext/ED527574.pdf>. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- So, H., Seah, L. H., & TohHeng, H. L. (2010). Designing collaborative knowledge building environments accessible to all learners: Impacts and design challenges. *Computers & Education, 54*(2), 479–490. The study is ineligible for review because it does not take place in the geographic area specified in the protocol.



- SRA/McGraw-Hill. (2005a). *California elementary school closes achievement gap with SRA reading programs*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2005b). *Combination of Open Court Reading and Direct Instruction equal consistently high reading scores*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2005c). *Fort Worth school district builds reading achievement, especially among economically disadvantaged students*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2005d). *Results with Open Court Reading*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006a). *Achievement gap begins to close in California school district with Open Court Reading*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006b). *Alabama's largest school district ranks in top 10 statewide after using Open Court Reading*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006c). *Chicago school's reading scores outshine district, state*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006d). *Colorado elementary school uses Open Court Reading for core reading program and direct instruction to close achievement gap*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006e). *Colorado school district earns state's top rating of excellent since 2001*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006f). *Ethnically diverse California district makes strides with Open Court reading and Spanish companion program*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006g). *Florida Exceptional Students attain reading proficiency with SRA/McGraw-Hill programs*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006h). *Georgia district uses Open Court Reading to reach record reading scores*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006i). *Horizons and Language for Learning close achievement gap among limited English proficient students*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006j). *Missouri elementary school closes achievement gap using SRA reading programs*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006k). *Nebraska Reading First school reaches states highest scores with SRA reading programs*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006l). *Open Court Reading helps Bethel School District excel*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.

- SRA/McGraw-Hill. (2006m). *Open Court Reading helps California district improve API score, reach out to parents*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006n). *Open Court Reading students outscore others in district*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006o). *St. Louis schools close achievement gap with Open Court Reading*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2006p). *Two Danville schools achieve top 10 most improved status after using Open Court Reading for nine months*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2007a). *SRA/McGraw-Hill's reading programs bring increases in Baltimore's scores*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2007b). *State reading scores improve at California elementary, Open Court Reading and REACH lead the way*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2008a). *Oakland school district receives national recognition from SRA/McGraw-Hill for achievement in reading*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- SRA/McGraw-Hill. (2008b). *Orlando elementary schools posting highest FCAT reading gains rely on SRA/McGraw-Hill literacy programs*. Columbus, OH: The McGraw-Hill Companies. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Stewart, R. M., Benner, G. J., Martella, R. C., & Marchand-Martella, N. E. (2007). Three-tier models of reading and behavior: A research review. *Journal of Positive Behavior Interventions*, 9(4), 239–253. The study is ineligible for review because it is a secondary analysis of the effectiveness of an intervention, such as a meta-analysis or research literature review.
- Stinnett, M. (2009). Research in reading. *Illinois Reading Council Journal*, 37(2), 59–64. The study is ineligible for review because it does not include a student outcome.
- Stufft, D. L., & Brogadir, R. (2011). Urban principals' facilitation of English language learning in public schools. *Education and Urban Society*, 43(5), 560–575. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Sun, Y., Zhang, J., & Scardamalia, M. (2010). Knowledge building and vocabulary growth over two years, grades 3 and 4. *Instructional Science*, 38(2), 147–171. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- The Business Roundtable Education Initiative, & the National Association of Elementary School Principals. (2005). *Open Court results*. Columbus, OH: McGraw-Hill Education. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Torgesen, J. K., & King, R. (2000). *FCRR technical report #3: Improving the effectiveness of reading instruction in one elementary school: A description of the process*. Tallahassee, FL: Florida Center for Reading Research. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Twining, L. L. (2008). *Raising student achievement at Eberman Elementary School with effective teaching strategies* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3317991) The study is ineligible for review because it does not use a comparison group design or a single-case design.

- Vail, M. C. (2006). *The utilization of the Accelerated Reading Program to minimize the discourse between the Open Court Reading Program and the cultural behavior of the student* (Unpublished doctoral dissertation). California State University, Sacramento. The study is ineligible for review because it does not use a comparison group design or a single-case design.
- Vernon-Feagans, L., Kainz, K., Amendum, S., Ginsberg, M., Wood, T., & Bock, A. (2012). Targeted reading intervention: A coaching model to help classroom teachers with struggling readers. *Learning Disability Quarterly, 35*(2), 102–114. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Walk, R. S. (2008). *Associations involving Open Court Reading in kindergarten and student performance on standardized assessments in reading in a Tennessee school system* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3305591) The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.
- Wang, C., & Algozzine, B. (2011). Rethinking the relationship between reading and behavior in early elementary school. *The Journal of Educational Research, 104*(2), 100–109. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Wehby, J. H., Falk, K. B., Barton-Arwood, S., Lane, K. L., & Cooley, C. (2003). The impact of comprehensive reading instruction on the academic and social behavior of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 11*(4), 225. The study is ineligible for review because it does not use a sample aligned with the Beginning Reading evidence review protocol.
- Whisnant, K. L. (2005). *Instructional strategies that increase literacy achievement among second and third grade Hispanic English learners* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3191764) The study is ineligible for review because it does not use a sample aligned with the Beginning Reading evidence review protocol.
- Wilson, P., Martens, P., & Arya, P. (2005). Accountability for reading and readers: What the numbers don't tell. *Reading Teacher, 58*(7), 622–631. The study is ineligible for review because it does not examine the effectiveness of an intervention.
- Yates, C. (2010). *Teaching for success: A comparison study of teachers' perceptions of different approaches to the instruction of reading* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3398703) The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.
- Young, C. A. (2008). How is contextualized spelling used to support reading in first-grade core reading programs? *Reading Improvement, 45*(1), 26–45. The study is ineligible for review because it does not include a student outcome.

Appendix A.1: Research details for Borman et al. (2008)

Borman, G. D., Dowling, N. M., & Schneck, C. (2008). A multi-site cluster randomized field trial of Open Court Reading. *Educational Evaluation and Policy Analysis, 30*(4), 389–407.

Table A1. Summary of findings

Meets WWC group standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Comprehension	679 students	+10	No

**Setting** The study initially included six schools—one each in Florida, Georgia, Idaho, Indiana, North Carolina, and Texas. Two schools were from rural areas, two from suburban areas, and two from urban areas. The Georgia school dropped out of the study.

**Study sample** McGraw-Hill Education recruited a group of schools that had not previously used *Open Court Reading*® to participate in the study. The six schools that initially participated were given free *Open Court Reading*® materials, as well as a teacher training program and implementation support. At each school, classrooms were randomly assigned within each grade either to be enrolled in *Open Court Reading*® or to serve as the comparison group.

The entire study sample consisted of 57 grade 1–5 classrooms containing a total of 1,099 students. The sample considered in this review, which aligns to the Beginning Reading review protocol, initially consisted of 44 grade 1–3 classrooms containing a total of 855 students. After attrition, the combined analysis sample consisted of 36 classrooms containing 679 students in grades 1–3; 379 students in the 20 *Open Court Reading*® classrooms and 300 students in 16 comparison classrooms. Of the participating students, more than 70% were minorities, and more than 75% were eligible for free or reduced-price lunches. Fewer than 15% were English as Second Language (ESL) students, and fewer than 10% were special education students.

**Intervention group** *Open Court Reading*® is a curriculum that includes textbooks, workbooks, decodable books, and anthologies. The curriculum consists of three main components: (a) Preparing to Read, (b) Reading and Responding, and (c) Language Arts. For this study, teachers were given a teacher’s edition of the curriculum that included scripted direct instruction lessons and diagnostic and assessment packages. The program is designed to be used for 2.5 hours per day with grades 1–2 and for 2 hours per day with grades 4–6. However, the authors report that external consultants observed that some teachers provided only 90 minutes of daily instruction. The intervention was implemented from fall to spring during the 2005–06 school year.

**Comparison group** The comparison classrooms used a “business-as-usual” curriculum and were instructed not to use *Open Court Reading*® or any of its materials. Principals mentioned that curricula currently in use by the comparison classrooms consisted of *Reading Street* by Scott Foresman, *Literacy Place* by Scholastic, *McGraw-Hill Reading* by MacMillan/McGraw-Hill, *Collections* by Harcourt, and *Trophies* by Harcourt. Consultants visited comparison classrooms and verified that they were not using *Open Court Reading*®.

**Outcomes and measurement**

For both the pretest (October 2005) and the posttest (May 2006), students took the CTBS/5 Terra Nova Reading Comprehension and Vocabulary subtests. A Reading Composite score was also reported, which is the average of these two subtest measures. For a more detailed description of these outcome measures, see Appendix B. Findings for the combined student sample on the Reading Composite score can be found in Appendix C.2. Additional findings reflecting subtest outcomes separately for grades 1, 2, and 3 can be found in Appendix D.2.

**Support for implementation**

Teachers were provided training opportunities with external consultants, which consisted of 2- to 3-day summer workshops. In addition, the consultants, who had teaching experience and detailed knowledge of *Open Court Reading*® and were trained by McGraw-Hill Education, visited and observed classrooms, and provided feedback to teachers.

**Appendix A.2: Research details for Skindrud and Gersten (2006)**

Skindrud, K., & Gersten, R. (2006). An evaluation of two contrasting approaches for improving reading achievement in a large urban district. *Elementary School Journal*, 106(5), 389–407.

**Table A2. Summary of findings**

**Meets WWC group design standards with reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General reading achievement	434 students	+12	No

**Setting**

The study was conducted in 12 schools in the Sacramento City Unified School District (SCUSD), a large urban district in northern California.

**Study sample**

Under California’s interpretation of Reading First, all 59 elementary schools in SCUSD were required to implement one of two models of reading instruction, *Success for All (SFA)*® or *Open Court Reading*®. In the fall of 1997, four schools implemented *SFA*®. A matched sample of *Open Court Reading*® schools were created by rank-ordering SCUSD schools by poverty level (measured by the percent of students eligible for free or reduced-price meals and percent of students on Aid to Families with Dependent Children), and selecting two comparison schools for each *SFA*® school—those ranked just above and just below each *SFA*® school. The study included two cohorts of students: students in Cohort 1 began using the reading programs in grade 2, while students in Cohort 2 began in grade 3. A total of 936 students in Cohort 1 and Cohort 2 participated in the study, including students continually enrolled at study schools from fall 1997 to spring 1999 who completed all study tests and did not repeat a grade. The WWC based its effectiveness rating on findings from 434 Cohort 1 students who participated in the study; 292 in the *Open Court Reading*® group and 142 in the comparison group—these students were followed from second to third grade. Results for the Cohort 2 students are not included in this report because, based on information obtained from the authors, that sample of students was not equivalent on key characteristics at baseline.

### Intervention group

Students in the intervention group received reading instruction using *Open Court Reading*<sup>®</sup>, a systematic approach to teaching alphabets, print knowledge, and phonemic awareness. For this study, the district used the 1996 version of the curricula, *Open Court Collections for Young Scholars*. Two hours of daily whole-class reading instruction was followed by 30 minutes of small-group instruction and/or independent work. All study students received a condensed selection of instructional content to “catch-up” students to *Open Court Reading*<sup>®</sup> content that they had not received in prior years (since they began using the curriculum in either second or third grade).

### Comparison group

Students in the comparison group received reading instruction through *SFA*<sup>®</sup>. Students were put into homogeneous groups, across classrooms and grades, based on reading skills. They received 90 minutes of reading instruction daily, outside of their homerooms. *SFA*<sup>®</sup> also prescribes additional writing instruction outside of these groups. The *SFA*<sup>®</sup> training consultants monitored implementation fidelity and observed additional writing instruction in all study schools during both study years. The authors noted that teachers in *SFA*<sup>®</sup> schools frequently included additional spelling and grammar, along with writing instruction, outside of the 90-minute reading block. A core reading curriculum is only prescribed in grades K–1; in grades 2–6, the schools can choose their own reading curricula. The authors state that the materials and guidelines for instruction (*Reading Roots* for grade 1, and *Reading Wings* for grades 2–4), as well as the professional development, tutoring, and the *SFA*<sup>®</sup> school facilitator and regional consultant oversight procedures, all followed those outlined by the developers of the curriculum.

### Outcomes and measurement

The outcome measure was the Reading subtest from the SAT-9, administered in both spring 1998 and spring 1999. The authors converted all measures to normal curve equivalent scores. For a more detailed description of this outcome measure, see Appendix B. The Language subtest from the SAT-9 was reported by the authors; however, this outcome measure is not included in this report because it is not an eligible outcome under the Beginning Reading evidence review protocol. The intermediate findings (after 1 year of implementation) for second graders are reported in Appendix D.1.

### Support for implementation

At *Open Court Reading*<sup>®</sup> schools, teachers received 4 days of basic grade-level training in year 1, followed by 4 days of advanced grade-level training in year 2. Each *Open Court Reading*<sup>®</sup> school received a reading coach (either full-time or part-time, depending on school size). Curriculum experts met monthly with reading coaches and administrators to refine instruction and supervision and to solve problems. Reading coaches collected implementation information but were prohibited from sharing the information with the study authors; the district-level reading coordinator indicated that although some schools had implementation problems at the beginning of the study, these were resolved by the second study year.

At *SFA*<sup>®</sup> schools, training and technical assistance was provided by *SFA*<sup>®</sup> consultants from a regional *SFA*<sup>®</sup> office. The *SFA*<sup>®</sup> consultants assessed implementation fidelity and rated it as a typical level of implementation when compared with national implementation averages.

### Appendix B: Outcome measures for each domain

General reading achievement	
<i>Stanford Achievement Test, 9th Edition (SAT-9)</i>	The outcome measure was the Reading subtest from the SAT-9, administered in both spring 1998 and spring 1999. The authors converted all assessment scores to normal curve equivalent scores (as cited in Skindrud & Gersten, 2006).
Comprehension	
<i>Comprehensive Test of Basic Skills, 5th Edition (CTBS/5) Terra Nova Reading Composite score</i>	This assessment consists of two subtests, Reading Comprehension and Vocabulary, and combines selected response items with constructed-response items that allow students to produce short and extended responses. The Reading Composite score is a simple average of the CTBS/5 Reading Comprehension and Vocabulary subtests described below (as cited in Borman et al., 2008).
Reading comprehension construct	
<i>CTBS/5 Terra Nova Reading Comprehension subtest</i>	This assessment combines selected-response items with constructed-response items that allow students to produce short and extended responses. The Reading Comprehension subtest items focus on five objectives: (a) oral comprehension of passages read aloud, (b) basic understanding of literal meanings of passages, (c) analyzing text, (d) evaluating and extending meaning, and (e) identifying reading strategies (as cited in Borman et al., 2008).
Vocabulary development construct	
<i>CTBS/5 Terra Nova Vocabulary subtest</i>	This assessment combines selected-response items with constructed-response items that allow students to produce short and extended responses. The Vocabulary subtest focuses on three objectives: (a) understanding word meaning, (b) identifying multi-meaning words, and (c) inferring words in context (as cited in Borman et al., 2008).

**Appendix C.1: Findings included in the rating for the general reading achievement domain**

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Skindrud &amp; Gersten, 2006<sup>a</sup></b>								
<i>Stanford Achievement Test, 9th Edition (SAT-9)</i>	Grade 3	12 schools/ 434 students	43.90 (16.50)	38.60 (18.50)	5.30	.31	+12	< .01
<b>Domain average for general reading achievement (Skindrud &amp; Gersten, 2006)</b>						<b>.31</b>	<b>+12</b>	<b>Not statistically significant</b>
<b>Domain average for general reading achievement across all studies</b>						<b>.31</b>	<b>+12</b>	<b>na</b>

**Table Notes:** For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

<sup>a</sup> For Skindrud and Gersten (2006), the p-value presented here was reported in the original study. A correction for clustering was needed and resulted in a WWC-computed p-value of .30 for the SAT-9; therefore, the WWC does not find the result to be statistically significant. The reported group means are based on an analysis of covariance (ANCOVA), which adjusted for pretest. This study is characterized as having a substantively important positive effect because the domain average effect size is larger than .25. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 26.

**Appendix C.2: Findings included in the rating for the comprehension domain**

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Borman et al., 2008<sup>a</sup></b>								
<i>Comprehensive Test of Basic Skills, 5th Edition (CTBS/5) Reading Composite score</i>	Grade 1–3	36 classrooms/ 679 students	603.07 (47.63)	590.98 (45.00)	12.09	.26	+10	nr
<b>Domain average for comprehension (Borman et al., 2008)</b>						<b>.26</b>	<b>+10</b>	<b>Not statistically significant</b>
<b>Domain average for comprehension across all studies</b>						<b>.26</b>	<b>+10</b>	<b>na</b>

**Table Notes:** For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. nr = not reported. na = not applicable.

<sup>a</sup> For Borman et al. (2008), a correction for clustering was needed but did not affect whether the contrast was found to be statistically significant. The WWC aggregated means and pooled standard deviations for grades 1–3 to align to the Beginning Reading topic area protocol. The authors presented grade-level means and standard deviations but did not report a p-value for the comparison of grades 1–3, and the WWC-calculated p-value for this comparison was larger than .05. The effect size in the table is based on the grade-level means and standard deviations in Table 3 of Borman et al. (2008). This study is characterized as having a substantively important positive effect because the domain average effect size is larger than .25. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 26.



Appendix D.1: Description of supplemental findings for the general reading achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Skindrud &amp; Gersten, 2006<sup>a</sup></b>								
<i>Stanford Achievement Test, 9th Edition (SAT-9)</i>	Grade 2	12 schools/ 434 students	44.30 (17.10)	37.20 (16.80)	7.10	.42	+16	< .01
<i>SAT-9</i>	Bottom quartile— grade 2	12 schools/ 108 students	33.60 (13.70)	25.80 (5.90)	7.80	.66	+24	nr
<i>SAT-9</i>	Bottom quartile— grade 3	12 schools/ 108 students	34.60 (13.10)	25.40 (14.20)	9.20	.68	+25	nr

**Table Notes:** The supplemental findings presented in this table are additional findings from studies in this report that do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported.

<sup>a</sup>For Skindrud and Gersten (2006), the p-values presented here were reported in the original study. Note that the authors did not conduct univariate statistical tests for all reported outcomes. For example, the two bottom quartile reading outcomes (in grade 2 and grade 3) were jointly significant at  $p < .001$ . The WWC does not find the results to be statistically significant after the correction for clustering and multiple comparisons adjustment were performed. A correction for clustering was needed and resulted in a WWC-computed p-value of .30, .054, and .047, respectively. A correction for multiple comparisons was needed for the two bottom-quartile outcomes and resulted in a WWC-computed critical p-value of .03, which is smaller than the corresponding p-value of .047 for the SAT-9 outcome in grade 3. The reported group means are ANCOVA-adjusted. The effect sizes reported here differ from those reported in the original study due to differences in the effect-size formulas used; WWC uses Hedges' g statistic, while the study appears to use the Cohen's d statistic to calculate effect sizes.

Appendix D.2: Description of supplemental findings for the comprehension domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Borman et al., 2008<sup>a</sup></b>								
<i>Comprehensive Test of Basic Skills, 5th Edition (CTBS/5), Reading Composite score</i>	Grade 1	16 classrooms/ 304 students	575.79 (37.19)	567.81 (42.26)	7.98	.20	+8	nr
<i>CTBS/5, Terra Nova Vocabulary subtest</i>	Grade 1	16 classrooms/ 304 students	563.59 (46.01)	551.72 (49.93)	11.87	.25	+10	nr
<i>CTBS/5, Terra Nova Reading Comprehension subtest</i>	Grade 1	16 classrooms/ 304 students	587.46 (36.15)	583.35 (42.30)	4.11	.10	+4	nr
<i>CTBS/5, Reading Composite score</i>	Grade 2	11 classrooms/ 207 students	610.01 (37.50)	599.97 (35.10)	10.04	.27	+11	nr
<i>CTBS/5, Terra Nova Vocabulary subtest</i>	Grade 2	11 classrooms/ 207 students	596.71 (43.41)	590.41 (42.10)	6.30	.15	+6	nr
<i>CTBS/5, Terra Nova Reading Comprehension subtest</i>	Grade 2	11 classrooms/ 207 students	622.74 (38.51)	608.99 (39.18)	13.75	.35	+14	nr
<i>CTBS/5, Reading Composite score</i>	Grade 3	9 classrooms/ 168 students	642.44 (45.35)	623.63 (35.42)	18.81	.45	+18	nr

<i>CTBS/5, Terra Nova Vocabulary subtest</i>	Grade 3	9 classrooms/ 168 students	633.18 (48.64)	616.46 (40.25)	16.72	.37	+14	nr
<i>CTBS/5, Terra Nova Reading Comprehension subtest</i>	Grade 3	9 classrooms/ 168 students	651.17 (47.14)	630.23 (35.44)	20.94	.49	+19	nr

**Table Notes:** The supplemental findings presented in this table are additional findings from studies in this report that do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. nr = not reported.

<sup>a</sup> For Borman et al. (2008), corrections for clustering and multiple comparisons were needed. The authors did not report *p*-values for the grade-specific contrasts; rather, they described the effect sizes for these contrasts. The effect size in the table is based on the grade-level means and standard deviations in Table 3 of Borman et al. (2008). WWC calculations show no statistically significant differences between the intervention and comparison groups for all of these outcome measures (*p*-values > .05).

### Endnotes

<sup>1</sup> The descriptive information for this program was obtained from a publicly available source: the program's website (<https://www.mheonline.com>, downloaded February 2014). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in March 2014, and the WWC incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.

<sup>2</sup> The literature search reflects documents publicly available by December 2013. The previous report was released in August 2008. This report has been updated to include reviews of 68 studies that have been released since 2008, and 91 studies that were released prior to 2008 but were not included in the earlier report. Of the additional studies, 134 were not within the scope of the review protocol for the Beginning Reading topic area, and 25 were within the scope of the review protocol for the Beginning Reading topic area but did not meet WWC group design standards. A complete list and disposition of all studies reviewed are provided in the references. One new study (Borman et al., 2008) *meets WWC group design standards without reservations*. One study from the 2008 report (Skindrud & Gersten, 2006) received a revised rating in this report of *meets WWC group design standards with reservations*, where it had previously received the rating of *does not meet WWC group design standards*. In the version 1.0 standards used to review the 2008 version of the intervention report, a statistically significant ( $p < .05$ ) difference in key baseline differences was sufficient to have a quasi-experiment receive a rating of *does not meet WWC group design standards*. However, in the WWC's version 3.0 standards, if baseline differences between intervention and comparison groups are between .05 and .25 standard deviations, the study can still meet standards after a proper statistical adjustment in the impact analysis. The studies in this report were reviewed using the Standards from the WWC Procedures and Standards Handbook (version 3.0), along with those described in the Beginning Reading review protocol (version 2.1). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

<sup>3</sup> For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 28. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

<sup>4</sup> Results for grades 4 and 5 (Borman et al., 2008) are reported in the WWC Adolescent Literacy *Open Court Reading*® intervention report.

<sup>5</sup> The study (Skindrud & Gersten, 2006) was conducted over 2 school years and analyzed two separate cohorts of students; Cohort 1 students began in grade 2, and Cohort 2 students began in grade 3. The sample of students in Cohort 1 meets the WWC baseline equivalence standard and is included in this report. The sample of students in Cohort 2 does not meet the WWC baseline equivalence standard and is excluded from this report.

<sup>6</sup> The findings considered for the effectiveness rating reflect the maximum exposure of students to the program. For example, in the second year of *Open Court Reading*® implementation, third graders (from Cohort 1) had been exposed to the program over a period of 2 school years (when they were in the second and third grades). The corresponding intermediate findings (after 1 year of implementation) for second graders from the same Cohort 1 are reported in Appendix D.2 and were not used in the effectiveness rating.

### Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014, October). *Beginning Reading intervention report: Open Court Reading*®. Retrieved from <http://whatworks.ed.gov>

## WWC Rating Criteria

### Criteria used to determine the rating of a study

Study rating	Criteria
<b>Meets WWC group design standards without reservations</b>	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
<b>Meets WWC group design standards with reservations</b>	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

### Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
<b>Positive effects</b>	Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
<b>Potentially positive effects</b>	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
<b>Mixed effects</b>	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
<b>Potentially negative effects</b>	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
<b>Negative effects</b>	Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
<b>No discernible effects</b>	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

### Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
<b>Medium to large</b>	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
<b>Small</b>	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

### Glossary of Terms

<b>Attrition</b>	Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.
<b>Clustering adjustment</b>	If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.
<b>Confounding factor</b>	A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.
<b>Design</b>	The design of a study is the method by which intervention and comparison groups were assigned.
<b>Domain</b>	A domain is a group of closely related outcomes.
<b>Effect size</b>	The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.
<b>Eligibility</b>	A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.
<b>Equivalence</b>	A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.
<b>Extent of evidence</b>	An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 28.
<b>Improvement index</b>	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from -50 to +50.
<b>Multiple comparison adjustment</b>	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
<b>Quasi-experimental design (QED)</b>	A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.
<b>Randomized controlled trial (RCT)</b>	A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.
<b>Rating of effectiveness</b>	The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 28.
<b>Single-case design</b>	A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.
<b>Standard deviation</b>	The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.
<b>Statistical significance</b>	Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ( $p < .05$ ).
<b>Substantively important</b>	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the [WWC Procedures and Standards Handbook \(version 3.0\)](#) for additional details.