What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES

Success for All®

Program Description[2]

Success for All (SFA)® is a whole-school reform model that includes a reading, writing, and oral language development program for students in prekindergarten through eighth grade. Classroom reading instruction is delivered in daily 90-minute blocks to students grouped by reading ability. Immediate intervention with tutors who are certified teachers is given each day to those students who are having difficulty reading at the same level as their classmates.

This intervention report focuses on the reading component of *SFA*®, which is often implemented in the context of the *SFA*® whole-school reform program. Although the whole-school reform program has key components that are implemented in each school, school sites may vary considerably in the number of personnel used to implement *SFA*®, particularly tutors and family support staff. The reading curricula are essentially the same at all schools, with each school receiving the same training, coaching support, and materials. Ratings presented in this report are not disaggregated by the variations in implementation of whole-school reforms.

Research[3]

One study of *SFA*® meets the What Works Clearinghouse (WWC) evidence standards and six studies meet WWC evidence standards with reservations. Altogether, the studies included nearly 4,000 students attending more than 70 elementary schools across the United States. The seven studies focused on students in grades K–3 who received the *SFA*® intervention for varying amounts of time.[4]

Based on these seven studies, the WWC considers the extent of evidence for *SFA*® to be medium to large for alphabetics, comprehension, and general reading achievement. No studies that meet WWC evidence standards with or without reservations addressed fluency.

Effectiveness

Success for All® was found to have positive effects on alphabetics, mixed effects for comprehension, and potentially positive effects on general reading achievement.

	Alphabetics	*Fluency*	*Comprehension*	*General reading achievement*
Rating of effectiveness	Positive effects	na	Mixed effects	Potentially positive effects
Improvement index[5]	Average: +13 percentile points Range: 0 to +32 percentile points	na	Average: +8 percentile points Range: 0 to +17 percentile points	Average: +10 percentile points Range: +2 to +19 percentile points

na = not applicable

1. This report has been updated to include reviews of 19 studies that have been identified since 2006 and to incorporate the results of a reexamination of the Madden et al. (1993) study included in the previous version of this report, described in footnote 8. Of the additional studies, 15 are not within the scope of the protocol and 4 are within the scope of the protocol but do not meet evidence standards. A complete list and disposition of all studies reviewed are provided in the references.
2. The descriptive information for this program was obtained from a publicly available source: the program's website (www.successforall.net, downloaded February 23, 2009). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
3. The studies in this report were reviewed using WWC Evidence Standards, Version 1.0 (see the WWC Standards).
4. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
5. These numbers show the average and range of improvement indices for all findings across the studies.

Additional program information

Developer and contact

Developed by Robert Slavin and Nancy Madden in conjunction with the Johns Hopkins University, *Success for All*® is distributed by the Success for All Foundation, Inc., 200 W. Towsontown Boulevard, Baltimore, MD 21204-5200. Email: sfainfo@successforall.org. Web: http://www.successforall.net. Telephone: (800) 548-4998 ext. 2372.

Scope of use

SFA® is used by schools in 48 states, Guam, and the Virgin Islands. According to the Success for All Foundation, more than 1,300 schools in more than 500 districts have used the *SFA*® whole-school reform program. Schools in Israel, Canada, Mexico, and Australia have implemented adapted versions of *SFA*®.

Teaching

During the regular daily 90-minute reading period, students are grouped into reading classes of 15–20 students who are all performing at the same reading level (regardless of age or grade level). Regrouping allows teachers to teach the whole class without having to break the class into multiple smaller reading groups.

Reading teachers at every grade level begin the period by reading children's literature to students. Teachers discuss the story with students to enhance the students' understanding of the story and the story structure and to increase their listening and speaking vocabulary. In kindergarten and first grade, teachers emphasize the development of language skills and use phonetically regular storybooks and instruction to focus on phonemic awareness, auditory discrimination, and sound blending. In the second through fifth grades, teachers use school- or district-provided reading materials, either basal or trade books, in a structured set of interactive activities in which students read, discuss, and write about the books. At this stage, teachers emphasize cooperative learning activities built around partner reading. Students work on identifying characters, settings, and problem solutions in narratives. Students receive direct instruction in reading comprehension skills.

Teachers in their first year of teaching *SFA*® receive a three-day summer training and 12 additional on-site support days during the school year. Additional in-service presentations covering topics such as classroom management, instructional pace, and cooperative learning are made by school facilitators and other program staff throughout the year. Facilitators organize information sessions to allow teachers to share problems and solutions, suggest changes, and discuss individual children. Twice a year, trainers provided by the developer visit and observe teachers. After the first year, training is reinforced by regular in-services, an annual *SFA*® conference, and on-site implementation support visits for school leaders and teachers. The staff development model used in whole-school *SFA*® reform emphasizes relatively brief initial training with extensive classroom follow-up, coaching, and group discussion.

Principals and facilitators receive five days of initial training in leadership, data collection and progress monitoring, classroom instructional practices, school climate, and intervention using *SFA*® strategies.

Cost

The cost of the *SFA*® whole-school reform program is approximately $80,000 in the first year, about $50,000 in the second year, and $35,000 in the third. Teacher training and ongoing support are required and are included in the cost of the program.

Research

One hundred twelve studies reviewed by the WWC investigated the effects of *SFA*®. One study (Borman, Slavin, Cheung, Chamberlain, Madden, & Chambers, 2006) is a randomized controlled trial that meets WWC evidence standards. Six other studies (Dianda & Flaherty, 1995; Madden, Slavin, Karweit, Dolan, & Wasik, 1993; Ross, Alberg, & McNelis, 1997; Ross & Casey, 1998; Ross, McNelis, Lewis, & Loomis, 1998; and Smith, Ross, Faulks, Casey, Shapiro, & Johnson, 1993) are quasi-experimental designs that meet WWC evidence standards with reservations. The remaining 105 studies do not meet either WWC evidence standards or eligibility screens.

Some studies measured the impact of *SFA*® after a cohort of students was exposed to *SFA*® for one, two, and three years.

To determine ratings, the WWC used results from the final year reported in a study for the overall domain rating, prioritizing the outcomes that reflected students' exposure to the intervention for the longest period of time available.[6] The studies in this report reflect results after (1) three years of exposure to *SFA*® (2 studies), (2) two years of exposure to *SFA*® (2 studies), and (3) one year of exposure to *SFA*® (3 studies).

Meets evidence standards

Borman et al. (2006) was a cluster randomized controlled trial that examined the effects of *SFA*® on students in grades K–2 across 14 states. The study randomly assigned 41 schools to *SFA*® and the comparison conditions and presented findings on students who had completed one, two, or three years of the program compared with students who took part in their schools' typical reading programs. The WWC based effectiveness ratings on findings from the third-year longitudinal sample of 1,425 students who began the study in kindergarten in 18 intervention and 17 comparison schools.

Meets evidence standards with reservations

Dianda and Flaherty (1995) studied the impact of *SFA*® on three different cohorts of students who started kindergarten in 1992, 1993, or 1994. Students were from six elementary schools in California. Students were grouped into four language categories; the WWC focuses only on the English-speaking group of 539 students for this review.[7] *SFA*® students were compared with students who did not use the *SFA*® program. The WWC based effectiveness ratings on findings for the three cohorts that were exposed to *SFA*® for two, three, or four years.

Madden et al. (1993) evaluated the effects of *SFA*® in Baltimore City elementary schools. The study investigated the effects of two versions of the *SFA*® program: full implementation and dropout prevention. Although these versions varied in their implementation of the whole-school reform model, the reading curricula are essentially the same at all schools, with each school receiving the same training, coaching support, and materials. Ratings presented in this report are not disaggregated by the variations in implementation of whole-school reforms.[8] The WWC based effectiveness ratings on the findings for students at the end of three years of implementation for the alphabetics and general reading achievement domains.

Ross, Alberg, and McNelis (1997) included first-grade students from 19 elementary schools implementing alternative schoolwide programs in the Northwest. The 19 schools were formed into four clusters of similar schools. For this review, the WWC reported results from students in three *SFA*® schools who were compared with the students from three schools that implemented the *Accelerated Schools* program. This subsample consisted of "cluster 2A" schools, which were neither the most disadvantaged nor the most affluent schools in the sample. This WWC review focused on the findings for 252 students at the end of the second grade, who had received one year of the *SFA*® program.

Ross and Casey (1998) examined the effects of *SFA*® in three schools in Ft. Wayne, Indiana, by comparing them with five schools that implemented "locally developed programs." The WWC focused on students who started the program in kindergarten at two *SFA*® schools. The WWC based effectiveness ratings on the findings for 288 students at the end of first grade who received two years of *SFA*®.

Ross et al. (1998) included 97 first-grade students from four elementary schools located in Little Rock, Arkansas. Two schools that implemented the *SFA*® were compared with two matched comparison schools that did not receive the

6. *SFA*® is designed to teach children to read at grade level by third grade, and the third year of program implementation is regarded as the full "dose" of *Success for All*® (Borman et al., 2006).
7. The WWC Beginning Reading topic focuses only on students learning to read in English (see Beginning Reading Protocol).
8. A previous version of this report separated the two implementations. It was later determined that both used the same reading curriculum. This updated report does not distinguish between the two implementations.

Research *(continued)*

intervention. The WWC based effectiveness ratings on findings at the end of the second grade after students received one year of *SFA*® implementation.

Smith et al. (1993) evaluated *SFA*® in two elementary schools in Ft. Wayne, Indiana, by comparing them with similar students in two matched comparison schools who did not receive *SFA*®. The WWC based effectiveness ratings on findings for 286 students spread across kindergarten and first grade who had received one year of *SFA*® implementation.

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the What Works Clearinghouse Extent of Evidence Categorization Scheme). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.[9]

The WWC considers the extent of evidence for *SFA*® to be medium to large for alphabetics, comprehension, and general reading achievement. No studies that meet WWC evidence standards with or without reservations addressed fluency.

Effectiveness

Findings

The WWC review of interventions for Beginning Reading addresses student outcomes in four domains: alphabetics, fluency, comprehension, and general reading achievement.[10] The studies included in this report cover three domains: alphabetics, comprehension, and general reading achievement. Alphabetics includes five constructs: phonemic awareness, phonological awareness, print awareness, letter knowledge, and phonics. Comprehension includes two constructs: reading comprehension and vocabulary development. General reading achievement includes outcome measures that do not explicitly differentiate among different reading domains (for example, a summary standardized test score). The findings below present the authors' estimates and WWC calculated estimates of the size and the statistical significance of the effects on students.[11] The results are presented by domain for each of the *SFA*® studies that meets the WWC evidence standards with or without reservations.

Alphabetics

In the alphabetics domain, all seven studies addressed phonics outcomes and one of these studies also measured students' letter knowledge skills.

Three years of program implementation:

- Borman et al. (2006) examined scores on the Woodcock Reading Mastery Test (WRMT) and reported statistically significant positive effects for two phonics subtests: Word Identification and Word Attack. The WWC analysis confirmed the statistical significance of these effects.
- Madden et al. (1993) found statistically significant positive effects on the phonics measure (the Woodcock Language Proficiency Battery [WLPB] Word Attack subtest) for students who began receiving the intervention in preschool and statistically significant positive effects on the WLPB Letter-Word Identification subtest for those who began in kindergarten. The WWC confirmed statistically significant positive effects on

9. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept, external validity, such as the students' demographics and the types of settings in which studies took place, are not taken into account for the categorization.
10. For definitions of the domains, see the Beginning Reading Protocol.
11. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of Borman et al. (2006), a correction for multiple comparisons was needed. In the six other studies, corrections for clustering and multiple comparisons were needed.

the phonics measure for the preschool cohort but found that none of the combined effects across schools for any other measures was statistically significant. The average effect size across these outcomes was substantively important according to WWC criteria (that is, an effect size of at least 0.25).

Two years of program implementation:

- Dianda and Flaherty (1995) reported effect sizes but did not report on the statistical significance of the effect of *SFA*® on two phonics measures: the WLPB Letter-Word Identification subtest and the Word Attack subtest. According to WWC calculations, there were no statistically significant effects of *SFA*®, but the average effect size across the two measures was positive and large enough to be considered substantively important.
- Ross and Casey (1998) reported no statistically significant effect of *SFA*® for one phonics measure (WRMT Word Identification subtest) but found a statistically significant positive effect for the other phonics measure (WRMT Word Attack subtest). In WWC computations, neither of the effects was statistically significant, and the average effect was not large enough to be considered substantively important.

One year of program implementation:

- Ross, Alberg, and McNelis (1997) did not find a statistically significant effect of *SFA*® for one phonics measure (the WRMT Word Identification subtest) but did find a statistically significant positive effect for the other phonics measure (WRMT Word Attack subtest). The WWC analyses showed that neither of the effects was statistically significant. In addition, the average effect size across the two outcomes was neither statistically significant nor large enough to be considered substantively important.
- Ross et al. (1998) found no statistically significant effects of *SFA*® on the two phonics outcomes: WRMT Word Identification and Word Attack subtests. The WWC analyses also found that no effects were statistically significant, but the average effect size across outcomes was positive and large enough to be considered substantively important.
- Smith et al. (1993) reported no statistically significant effect of *SFA*® on the letter knowledge construct (WRMT Letter Identification subtest) but found statistically significant positive effects for the two phonics outcomes (WRMT Word Identification and Word Attack subtests) for first-grade students. For kindergarten students, the authors found statistically significant positive effects for the WRMT Letter Identification and the Word Identification subtests. The WWC calculations found that although none of these effects was statistically significant, the average effect size across outcomes was positive and large enough to be substantively important.

Overall, in the alphabetics domain, two studies showed statistically significant positive effects. Three studies showed substantively important positive effects, and two studies showed indeterminate effects.[12]

Comprehension

In the comprehension domain, six studies addressed reading comprehension outcomes, and one of these studies also measured students' vocabulary development skills.

Three years of program implementation:

- Borman et al. (2006) reported and the WWC confirmed a statistically significant positive effect of *SFA*® on the WRMT Passage Comprehension subtest.

Two years of program implementation:

- Dianda and Flaherty (1995) did not report on the statistical significance of the effect of *SFA*® on the WLPB Passage Comprehension subtest. The WWC found no statistically significant effect, but the positive effect was large enough to be considered substantively important according to WWC criteria.
- Ross and Casey (1998) reported no statistically significant effect of *SFA*® on the WRMT Passage Comprehension

12. Indeterminate effects are defined as effects that are not statistically significant and with effect sizes smaller than 0.25.

subtest. In addition, the WWC found that the effect size was positive, but not substantively important.

One year of program implementation:

- Ross, Alberg, and McNelis (1997) reported no statistically significant effect on the WRMT Passage Comprehension subtest. The WWC found that the effect size was positive, but not substantively important.
- Ross et al. (1998) reported a positive but not statistically significant effect of *SFA*® on the WRMT Passage Comprehension subtest. The effect size was not large enough to be considered substantively important according to WWC criteria.
- Smith et al. (1993) reported no statistically significant effect of *SFA*® on the vocabulary development measure (Peabody Picture Vocabulary Test) for kindergarteners. For first-graders, the study authors found a statistically significant positive effect on the WRMT Passage Comprehension subtest. The WWC analysis found that none of the effects was statistically significant; and the average effect size across all outcomes was not large enough to be considered substantively important.

For the comprehension domain, one study reported a statistically significant positive effect and had a strong design. One study showed substantively important positive effects, and four studies showed indeterminate effects.

General reading achievement

Six studies examined outcomes in the general reading achievement domain.

Three years of program implementation:

- Dianda and Flaherty (1995) examined the effects of *SFA*® on the combined measure of WLPB and Durrell Oral Reading subtest for three cohorts of students after two to four years of program implementation. The authors did not report on the statistical significance of the findings. The WWC effect size computations found that although none of the effects was statistically significant, the mean effect size across all outcomes was positive and large enough to be considered substantively important.
- Madden et al. (1993) found statistically significant positive effects of *SFA*® on the Durrell Oral Reading subtest for students who began in kindergarten and first grade. The WWC computations found that none of the positive effects combined across schools was statistically significant, but the mean effect across grade levels was large enough to be considered substantively important.

Two years of program implementation:

- Ross and Casey (1998) reported a positive but not statistically significant effect of *SFA*® on the Durrell Oral Reading subtest. The effect size was not large enough to be considered substantively important according to WWC criteria.

One year of program implementation:

- Ross, Alberg, and McNelis (1997) reported a positive but not statistically significant effect of *SFA*® on the Durrell Oral Reading subtest. The effect size was not large enough to be considered substantively important according to WWC criteria.
- Smith et al. (1993) found a statistically significant positive effect of *SFA*® on the Durrell Oral Reading subtest. The WWC computations found that the effect was not statistically significant, but large enough to be considered substantively important.
- Ross et al. (1998) reported a positive but not statistically significant effect on the Durrell Oral Reading subtest. The effect size was not large enough to be considered substantively important according to WWC criteria.

In the general reading domain, three studies reported substantively important positive effects and three studies showed indeterminate effects. No study had a strong design.

Rating of effectiveness

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

The WWC found *Success for All*® to have positive effects in the alphabetics domain, potentially positive effects on general reading achievement, and mixed effects on comprehension

Improvement index

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

The average improvement index for alphabetics is +13 percentile points across the seven studies, with a range of 0 to +32 percentile points across findings. The average improvement index for comprehension is +8 percentile points across the six studies, with a range of 0 to +17 percentile points across findings. The average improvement index for general reading is +10 percentile points across the six studies, with a range of +2 to +19 percentile points across findings.

Summary

The WWC reviewed 112 studies on *Success for All*®. One of these studies meets WWC evidence standards; six studies meet WWC evidence standards with reservations; the remaining 105 studies do not meet either WWC evidence standards or eligibility screens. Based on the seven studies, the WWC found positive effects in the alphabetics domain, potentially positive effects in the general reading achievement domain, and mixed effects in the comprehension domain. The conclusions presented in this report may change as new research emerges.

References

Meets WWC evidence standards

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2006). *Final reading outcomes of the national randomized field trial of* Success for All. Retrieved from *Success for All* website: http://www.successforall.net/_images/pdfs/Third_Year_Results_06.doc.

Additional sources:

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A., & Madden, N. A. (2004). Success for All: *Preliminary first-year results from the national randomized field trial*. Baltimore, MD: Success for All Foundation.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2005). *Success for All*: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis, 27*(1), 1–22.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2005, Winter). The national randomized field trial of *Success for All*: Second-year outcomes. *American Educational Research Journal, 42*(4), 673–696.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of *Success for All*. *American Educational Research Journal, 44*(3), 701.

Slavin, R. E., Madden, N. A., Cheung, A., Chamberlain, A., Chambers, B., & Borman, G. (2005). *A randomized evaluation of* Success for All: *Second-year outcomes*. Baltimore, MD: Success for All Foundation.

Meets WWC evidence standards with reservations

Dianda, M., & Flaherty, J. (1995, April). *Effects of* Success for All *on the reading achievement of first graders in California bilingual programs*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Additional sources:

Dianda, M. R., & Flaherty, J. F. (1995). Report on work station uses: *Effects of* Success for All *on the reading achievement*

of first graders in California bilingual programs. Los Alamitos, CA: Southwest Regional Laboratory. (ERIC Document Reproduction Service No. ED394327)

Livingston, M., & Flaherty, J. (1997). *Effects of* Success for All *on reading achievement in California schools*. San Francisco, CA: WestEd.

Slavin, R. E., & Madden, N. A. (1999). Effects of bilingual and English as a second language adaptations of *Success for All* on the reading achievement of students acquiring English. *Journal of Education for Students Placed At Risk, 4*(4), 393–416. (Study: California)

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Success for All:* A summary of research. *Journal of Education for Students Placed At Risk, 1*(1), 41–76. (Study: California)

Madden, N. A., Slavin, R. E., Karweit, N., Dolan, L., & Wasik, B. A. (1993). *Success for All:* Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30*(1), 123–148.

Additional sources:

Borman, G. D., & Hewes, G. M. (2002). The long-term effects and cost effectiveness of *Success for All*. *Educational Evaluation and Policy Analysis, 24*(4), 243–266.

Madden, N. A., Slavin, R. E., Karweit, N., Dolan, L., & Wasik, B. A. (1991). Success for All: *Multi-year effects of a schoolwide elementary restructuring program*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students. (Study: Baltimore, MD)

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1993). Success for All *in the Baltimore City Public Schools: Year 6 report*. Baltimore, MD: Johns Hopkins University, Center for Research in Effective Schooling for Disadvantaged Students.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Success for All*: A summary of research. *Journal of Education for Students Placed At Risk, 1*(1), 41–76. (Study: Baltimore, MD)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994). "Whenever and wherever we choose": The replication of *Success for All*. *Phi Delta Kappan, 75*(8), 639–647. (Study: Baltimore, MD)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994, April). Success for All: *Longitudinal effects of systemic school-by-school reform in seven districts*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Baltimore, MD)

Slavin, R. E., Madden, N. A., Karweit, N. L., Dolan, L., & Wasik, B. A. (1990). Success for All: *Second year report*. Baltimore, MD: Baltimore Public Education Institute and Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Slavin, R. E., Madden, N. A., Karweit, N., Dolan, L., & Wasik, B. A. (1993). Success for All *in the Baltimore City Public Schools: Year 5 report*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Ross, S. M., Alberg, M., & McNelis, M. (1997). *Evaluation of elementary school school-wide programs: Clover Park School District, year 1: 1996–97*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Additional source:

Ross, S. M., Alberg, M., McNelis, M., & Rakow, J. (1998). *Evaluation of elementary school school-wide programs: Clover Park School District year 2: 1997–98*. Memphis: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., & Casey, J. (1998). *Longitudinal study of student literacy achievement in different Title I school-wide programs in Fort Wayne community schools—Year 2: First grade results*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Additional sources:

Casey, J., Smith, L. J., & Ross, S. M. (1994). *1993–1994 Fort Wayne, Indiana, SFA results*. Memphis, TN: Center for Research in Educational Policy.

Ross, S. M., Smith, L. J., & Casey, J. (1995). *Final report: 1994–95* Success for All *program in Fort Wayne, Indiana*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., Smith, L. J., Casey, J., & Johnson, B. (1993). *Final report: 1992–93* Success for All *program in Ft. Wayne, Indiana*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., Smith, L. J., Casey, J., Johnson, B., & Bond, C. (1994, April). *Using* Success for All *to restructure elementary schools: A tale of four cities*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Ft. Wayne, IN)

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Success for All*: A summary of research. *Journal of Education for Students Placed At Risk, 1*(1), 41–76. (Study: Ft. Wayne, IN)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994, April). Success for All: *Longitudinal effects of systemic school-by-school reform in seven districts*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Ft. Wayne, IN)

Smith, L. J., Ross, S. M., & Casey, J. (1996). Multi-site comparison of the effects of *Success for All* on reading achievement. *Journal of Literacy Research, 28*(3), 329–353. (Study: Ft. Wayne, IN)

Smith, L. J., Ross, S. M., Faulks, A., Casey, J., Shapiro, M., & Johnson, B. (1993). *Final report: 1991–92* Success for All *program in Ft. Wayne, Indiana*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., McNelis, M., Lewis, T., & Loomis, S. (1998). *Evaluation of* Success for All *programs: Little Rock School District year 1: 1997–1998*. Memphis, TN: The University of Memphis, Center for Research in Education Policy.

Additional sources:

Wang, L. W., & Ross, S. M. (1999). *Evaluation of* Success for All *program: Little Rock School District year 2: 1998–1999*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Wang, L. W., & Ross, S. M. (1998). *Evaluation of* Success for All *programs: Little Rock School District year 1: 1997–1998*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Smith, L. J., Ross, S. M., Faulks, A., Casey, J., Shapiro, M., & Johnson, B. (1993). *1991–1992 Ft. Wayne, Indiana* SFA *results*. Memphis, TN: University of Memphis, Center for Research in Education Policy.

Additional source:

Smith, L. J., Ross, S. M., & Casey, J. (1996). Multi-site comparison of the effects of *Success for All* on reading achievement. *Journal of Literacy Research, 28*(3), 329–353.

Studies that fall outside the Beginning Reading protocol or do not meet WWC evidence standards

Ahearn, E. M. (1994). *Involvement of students with disabilities in the New American Schools development corporation projects*. Alexandria, VA: National Association of State Directors of Special Education. (ERIC Document Reproduction Service No. ED371513) This study is ineligible for review because it does not include a student outcome.

Atkinson, C. L. H. (1998). An analysis of the impact of *Success for All* on reading, attendance, and academic self-efficacy with at-risk elementary school students. *Dissertation Abstracts International, 59*(10), 3699A. (UMI No. 9905180) The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Barnes, C., Camburn, E., Kim, J. S., & Rowan, B. (2005, April). *School leadership and instructional improvement in CSR schools*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA. This study is ineligible for review because it does not include a student outcome.

References *(continued)*

Berends, M., Chun, J., Schuyler, G., Stockly, S., & Briggs, R. J. (2002). *Challenges of conflicting school reforms: New American Schools in a high-poverty district*. Santa Monica, CA: RAND Education. (ERIC Document Reproduction Service No. ED464984) This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Berends, M., Kirby, S. N., Naftel, S., & McKelvey, C. (2000). *Implementation and performance in New American Schools: Three years into scale-up*. Santa Monica, CA: RAND Education. (ERIC Document Reproduction Service No. ED451204) This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Bifulco, R. (2001). Do whole-school reform models boost student performance: Evidence from New York City. *Dissertation Abstracts International, 62*(06), 1991A. (UMI No. 3019134) The study does not meet evidence standards because the overall attrition rate exceeds WWC standards for this area.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*, 125–230. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Casey, J., Smith, L. J., & Ross, S. M. (1994). *Final report: 1993–94* Success for All *program in Memphis, Tennessee: Formative evaluation of new* SFA *schools*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Chamberlain, A., Daniels, C., Madden, N. A., & Slavin, R. E. (n.d.). A randomized evaluation of the *Success for All* middle school reading program. *Middle Grades Research Journal, 2*(1), 1–21. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Chambers, B., Abrami, P. C., & Morrison, S. (2001). Can *Success for All* succeed in Canada? In R. E. Slavin & N. A. Madden (Eds.), Success for All: *Research and reform in elementary education* (pp. 93–109). Mahwah, NJ: Lawrence Erlbaum Associates. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Chambers, B., Cheung, A. C. K., Madden, N. A., Slavin, R. E., & Gifford, R. (2006). Achievement effects of embedded multimedia in a *Success for All* reading program. *Journal of Educational Psychology, 98*(1), 232–237. This study is ineligible for review because it does not use a comparison group.

Additional source:

Chambers, B., Cheung, A., Gifford, R., Madden, N., & Slavin, R. E. (2004). *Achievement effects of embedded multimedia in a* Success for All *reading program*. Baltimore, MD: Success for All Foundation.

Chambers, B., Slavin, R. E., Madden, N. A., Cheung, A., & Gifford, R. (2005). *Enhancing* Success for All *for Hispanic students: Effects on beginning reading achievement*. Baltimore, MD: Success for All Foundation. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Additional source:

Chambers, B., Slavin, R. E., Madden, N. A., Cheung, A., & Gifford, R. (2004). *Effects of* Success for All *with embedded video on the beginning reading achievement of Hispanic children*. Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Clarke, P. A. (2001). *Analysis of the* Success for All *and* School Development *programs and their effects on reading comprehension*. Union, NJ: Kean University. (ERIC Document Reproduction Service No. ED456426) The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Comprehensive School Reform Quality Center. (2006). *CSRQ center report on elementary school CSR models*. Washington, DC:

Comprehensive School Reform Quality Center, American Institutes for Research. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. *American Educational Research Journal, 44*(2), 298. The study is ineligible for review because it does not include a student outcome.

Datnow, A., Borman, G. D., Stringfield, S., Overman, L. T., & Castellano, M. (2003). Comprehensive school reform in culturally and linguistically diverse contexts: Implementation and outcomes from a four-year study. *Educational Evaluation and Policy Analysis, 25*(2), 143–170. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Dicembre, E. (2002). How they turned the ship around. *Journal of Staff Development, 23*(2), 32–35. This study is ineligible for review because it does not use a comparison group.

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.

Gines, B. E. (2005). *Comparing two comprehensive reform models: Their effect on student reading achievement*. Retrieved from https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/4120/research.pdf?sequence=3. This study is ineligible for review because it does not use a comparison group.

Greenlaw, M. J. (2004). A case study examining the relationships among teachers' perceptions of the *Success for All* reading program, teachers' sense of efficacy, students' attitudes toward reading and students' reading achievement. *Dissertation Abstracts International, 65*(07), 2541A. (UMI No. 3139431) This study is ineligible for review because it does not use a comparison group.

Grehan, A. W. (2001). The effects of the *Success for All* program on improving reading readiness skills for at-risk students in kindergarten. *Dissertation Abstracts International, 62*(10), 3292A. (UMI No. 3029892). The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Hankerson, K. M. (2004). A cross-case study of the practices of the *Success for All* (*SFA*) facilitator. *Dissertation Abstracts International, 65*(03), 836A. (UMI No. 3126318) This study is ineligible for review because it does not use a comparison group.

Harris, A., Hopkins, D., & Wordsworth, J. (2001). The implementation and impact of *Success for All* in English schools. In R. E. Slavin & N. A. Madden (Eds.), Success for All: *Research and reform in elementary education* (pp. 81–92). Mahwah, NJ: Lawrence Erlbaum Associates. This study is ineligible for review because it does not examine an intervention conducted in English.

Hertz-Lazarowitz, R. (2001). *Success for All:* A community model for advancing Arabs and Jews in Israel. In R. E. Slavin & N. A. Madden (Eds.), Success for All: *Research and reform in elementary education* (pp. 149–177). Mahwah, NJ: Lawrence Erlbaum Associates. This study is ineligible for review because it does not examine an intervention conducted in English.

Hess, P. M. (2004). A study of teachers' selection and implementation of meta-cognitive reading strategies for fourth/fifth grade reading comprehension from a *Success for All* reading program perspective: Moving beyond the fundamentals. *Dissertation Abstracts International, 65*(07), 2542A. (UMI No. 3140930) This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Hicks, J. E. (2006). A systems analysis of the *Success for All* reading program: Student achievement and teacher perceptions. (Texas A&M University—Commerce; 1287 Advisor: Adviser Wayne Linek). *Dissertation Abstracts International, 67*(12A), 190–4433. This study is ineligible for review because it does not use a comparison group.

References *(continued)*

Hurley, E. A., Chamberlain, A., Slavin, R. E., & Madden, N. A. (2001). Effects of *Success for All* on TAAS reading scores—a Texas statewide evaluation. *Phi Delta Kappan, 82*(10), 750–756. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

James, D. W., Jurich, S., & Estes, S. (2001). *Raising minority academic achievement: A compendium of education programs and practices*. Retrieved from American Youth Policy Forum website: http://www.aypf.org/publications/rmaa/pdfs/Book.pdf. This study is ineligible for review because it does not use a comparison group.

James, L. R. D. (2003). The effect of the *Success for All* reading approach on fourth- and fifth-grade students' standardized reading assessment scores. *Dissertation Abstracts International 63*(11), 3896A. (UMI No. 3072259) This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Jenkins, J. R., & O'Connor, R. E. (2005). Cooperative learning for students with learning disabilities: Evidence from experiments, observations, and interviews. In H. L. Swanson, K. R. Harris, & S. Graham (Eds), *Handbook of learning disabilities*. New York: Guilford Publications. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Jones, E. M., Gottfredson, G. D., & Gottfredson, D. C. (1997). Success for some: An evaluation of a *Success for All* program. *Evaluation Review, 21*(6), 643–670. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Additional sources:

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Success for All*: A summary of research. *Journal of Education for Students Placed At Risk, 1*(1), 41–76. (Study: Charleston, SC)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994, April). Success for All: *Longitudinal effects of systemic school-by-school reform in seven districts*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Charleston, SC)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994). "Whenever and wherever we choose": The replication of *Success for All*. *Phi Delta Kappan, 75*(8), 639–647. (Study: Charleston, SC)

Kapushion, B. M. (2003). A qualitative study of "*Success for All*—Roots and Wings" on four Jefferson County schools. *Dissertation Abstracts International, 64*(01), 58A. (UMI No. 3078189) This study is ineligible for review because it does not use a comparison group.

Klingner, J., Cramer, E., & Harry, B. (2006). Challenges in the implementation of *Success for All* in four high-need urban schools. *The Elementary School Journal, 106*(4), 333–349. The study is ineligible for review because it does not does not include a student outcome.

Koh, M. S., & Robertson, J. S. (2003). School reform models and special education. *Education and Urban Society, 35*(4), 421–442. This study is ineligible for review because it does not use a comparison group.

Lewis, J. L., & Bartz, M. (1999). *New American Schools designs: An analysis of program results in district schools—Cincinnati Public Schools*. Cincinnati, OH: Cincinnati Public Schools, Research and Evaluation Office. This study is ineligible for review because it does not use a comparison group.

Lucius, L. B. (2000). A comparison of three kindergarten curricula on language and literacy performance. *Dissertation Abstracts International, 62*(01), 65A. (UMI No. 3003007) This study is ineligible for review because it does not include a student outcome.

Madden, N. A. (2006). Reducing the gap: *Success for All* and the achievement of African American students. *Journal of Negro Education, 75*(3), 389–400. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Madden, N. A., Slavin, R. E., & Simons, K. (1999). *MathWings: Effects on student mathematics performance* (Report No. 39).

Baltimore, MD: Center for Research on the Education of Students Placed at Risk. This study is ineligible for review because it does not use a comparison group.

Manset, G., St. John, E. P., Simmons, A., Michael, R., Bardzell, J., Hodges, D., et al. (1999). *Indiana's early literacy intervention grant program impact study for 1997–98*. Retrieved from ERIC: http://www.eric.ed.gov/PDFS/ED439392.pdf. This study is ineligible for review because it does not include a student outcome.

Mason, B. (2005). *Achievement effects of five comprehensive school reform designs implemented in Los Angeles Unified School District*. Doctoral dissertation, Pardee RAND Graduate School. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Massue, F. M. (1999). *Effects of engaging in* Success for All *on children's causal attributions*. Montreal, Quebec, Canada: Concordia University, Department of Education. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

McCollum, S., McNeese, M. N., Styron, R., & Lee, D. E. (2007). A school district comparison of reading achievement based on three reading programs. *Journal of At-Risk Issues, 13*(1; 1), 1–6. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

McCollum-Rogers, S. A. (2004). Comparing *Direct Instruction* and *Success for All* with a basal reading program in relation to student achievement. *Dissertation Abstracts International, 65*(10), 3642A. (UMI No. 3149920) The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

McCoy, E. B. (2005). *The performance of student participants in externally developed Title I schoolwide reading interventions in the Little Rock school district as measured by national, state and local standardized tests*. Unpublished dissertation, University of Arkansas at Little Rock, College of Education. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Motte, A. M. (2006). Success for All: *Program effects on three elementary schools in Lafayette Parish, Louisiana*. Unpublished dissertation (70244960). This study is ineligible for review because it does not use a comparison group.

Munoz, M. A., Dossett, D., & Judy-Gallans, K. (2003). *Educating students placed at risk: Evaluating the impact of* Success for All *in urban settings*. Retrieved from ERIC: http://www.eric.ed.gov/PDFS/ED480178.pdf. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Nunnery, J. A., Slavin, R. E., Madden, N. A., Ross, S. M., Smith, L. J., Hunter, P., et al. (1997, March). *Effects of full and partial implementations of* Success for All *on student reading achievement in English and Spanish*. Paper presented at the meeting of the American Educational Research Association, Chicago, IL. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Additional sources:

Nunnery, J. A. (1995). An assessment of *Success for All* program component effects on the reading achievement of at-risk first grade students. *Dissertation Abstracts International, 57*(01), 155A. (UMI No. 9615378)

Nunnery, J. A., Slavin, R. E., Ross, S. M., Smith, L. J., Hunter, P., & Stubbs, J. (1996, April). *An assessment of* Success for All *program component configuration effects on the reading achievement of at-risk first grade students*. Paper presented at the meeting of the American Educational Research Association, New York.

Peurach, D. J. (2005). *Designing and managing comprehensive school reform: The case of* Success for All. Unpublished doctoral dissertation, University of Michigan, Ann Arbor. The study

is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Pogrow, S. (2002). *Success for All* is a failure. *Phi Delta Kappan, 83*(6), 463–468. This study is ineligible for review because it does not use a comparison group.

Additional source:

Pogrow, S. (2003). [Pogrow final debate]. Unpublished raw data.

Reis, S. M., McCoach, D. B., Coyne, M., Schreiber, F., Eckert, R. D., & Gubbins, E. J. (2007). Using planned enrichment strategies with direct instruction to improve reading fluency, comprehension, and attitude toward reading: An evidence-based study. *The Elementary School Journal, 108*(1), 3–25. The study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Reynolds, M., & Wheldall, K. (2007). Reading recovery 20 years down the track: Looking forward, looking back. *International Journal of Disability, Development and Education, 54*(2), 199–223. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Ross, S. M., & Casey, J. (1998). Success for All *evaluation: 1997–1998 Tigard-Tualatin School District*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Ross, S. M., & Smith, L. J. (1994). Effects of the *Success for All* model on kindergarten through second-grade reading achievement, teachers' adjustment, and classroom-school climate at an inner-city school. *The Elementary School Journal, 95*(2), 121–138. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Additional sources:

Ross, S. M., & Smith, L. J. (1992). *Final report: 1991–92* Success for All *program in Memphis, Tennessee*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., Smith, L. J., Casey, J., Johnson, B., & Bond, C. (1994, April). *Using* Success for All *to restructure elementary schools: A tale of four cities*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Memphis, TN)

Ross, S. M., Smith, L. J., Crawford, A., Eck, L., Lohr, L., & Faulks, A. (1991). *Final report:* Success for All *1990–91 Memphis program*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Success for All*: A summary of research. *Journal of Education for Students Placed At Risk, 1*(1), 41–76. (Study: Memphis, TN)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994, April). Success for All: *Longitudinal effects of systemic school-by-school reform in seven districts*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Memphis, TN)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994). "Whenever and wherever we choose": The replication of *Success for All. Phi Delta Kappan, 75*(8), 639–647. (Study: Memphis, TN)

Smith, L. J., Ross, S. M., & Casey, J. (1994). *Final report: 1993–1994* Success for All *program in Memphis, Tennessee*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., Fleischman, S. W., & Hornbeck, M. (2003). *Progress and options regarding the implementation of* Direct Instruction *and* Success for All *in Toledo public schools*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

References *(continued)*

Ross, S. M., Nunnery, J. A., Goldfeder, E., McDonald, A., Rachor, R., Hornbeck, M., et al. (2004). Using school reform models to improve reading achievement: A longitudinal study of direct instruction and *Success for All* in an urban district. *Journal of Education for Students Placed At Risk, 9*(4), 357–388. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Additional source:

Ross, S. M., Nunnery, J. A., & Smith, L. J. (1996). *Evaluation of Title I reading programs: Amphitheater public schools—Year 1: 1995–1996*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., Nunnery, J. A., & Smith, L. J. (1996). *Evaluation of Title I reading programs: Amphitheater public schools—Year 1: 1995–1996*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—the intervention was combined with another intervention.

Additional source:

Slavin, R. E., & Madden, N. A. (1999). Effects of bilingual and English as a second language adaptations of *Success for All* on the reading achievement of students acquiring English. *Journal of Education for Students Placed At Risk, 4*(4), 393–416. (Study: Arizona)

Ross, S. M., Sanders, W. L., & Wright, S. P. (2000). *Fourth-year achievement results on the Tennessee Value-Added Assessment System for restructuring schools in Memphis*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Ross, S. M., Sanders, W. L., Wright, S. P., Stringfield, S., Wang, L. W., & Alberg, M. (2001). *Two- and three-year achievement results from the Memphis Restructuring Initiative*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Ross, S. M., Smith, L. J., & Casey, J. (1992). *Final report: 1991–92* Success for All *program in Caldwell, Idaho*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Additional sources:

Ross, S. M., Smith, L. J., Casey, J., Johnson, B., & Bond, C. (1994, April). *Using* Success for All *to restructure elementary schools: A tale of four cities*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Caldwell, ID)

Ross, S. M., Smith, L. J., Casey, J., & Slavin, R. E. (1995). Increasing the academic success of disadvantaged children: An examination of alternative early intervention programs. *American Educational Research Journal, 32*(4), 773–800.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Success for All*: A summary of research. *Journal of Education for Students Placed At Risk, 1*(1), 41–76. (Study: Caldwell, ID)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994, April). Success for All: *Longitudinal effects of systemic school-by-school reform in seven districts*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Caldwell, ID)

Smith, L. J., Ross, S. M., & Casey, J. (1996). Multi-site comparison of the effects of *Success for All* on reading achievement. *Journal of Literacy Research, 28*(3), 329–353. (Study: Caldwell, ID)

Ross, S. M., Smith, L. J., & Casey, J. (1997). *Final report: 1996–97* Success for All *program in Clark County, Georgia*.

Memphis, TN: University of Memphis, Center for Research in Educational Policy. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Ross, S. M., Smith, L. J., & Casey, J. (1997). Preventing early school failure: Impacts of *Success for All* on standardized test outcomes, minority group performance, and school effectiveness. *Journal of Education for Students Placed At Risk, 2*(1), 29–53. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Additional source:

Ross, S. M., Smith, L. J., & Casey, J. (1999). Bridging the gap: The effects of the *Success for All* program on elementary school reading achievement as a function of student ethnicity and ability level. *School Effectiveness and School Improvement, 10*(2), 129–150.

Ross, S. M., Smith, L. J., Casey, J., & Slavin, R. E. (1995). Increasing the academic success of disadvantaged children: An examination of alternative early intervention programs. *American Educational Research Journal, 32*(4), 773–800. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Ross, S. M., Smith, L. J., & Nunnery, J. A. (1998, April). *The relationship of program implementation quality and student achievement*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Additional source:

Ross, S. M., Smith, L. J., Lewis, T., & Nunnery, J. A. (1996). *1995–96 evaluation of Roots and Wings in Memphis City Schools*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., Smith, L. J., & Nunnery, J. A. (1998, April). *Title I as a catalyst for school improvement: Impact of alternative schoolwide models on the reading achievement of students at risk*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Ross, S. M., Tabachnick, S., & Sterbinsky, A. (2002). Using comprehensive school reform models to raise student achievement: Factors associated with success in Memphis. Retrieved from New American Schools website: http://web.archive.org/web/20031218233713/http://www.naschools.org/uploadedfiles/Ross+Using+Comprehensive+School+Reform+Models.pdf. This study is ineligible for review because it does not include a student outcome.

Ross, S. M., Wang, W., Sanders, W. L., & Wright, S. P. (1999). Teacher mobility and effectiveness in restructuring and non-restructuring schools in an inner-city district. Retrieved from SAS Institute website: http://www.sas.com/govedu/edu/teacher_mobility.pdf. This study is ineligible for review because it does not include a student outcome.

Ross, S. M., Wang, L. W., Sanders, W. L., Wright, S. P., & Stringfield, S. (1999). *Two- and three-year achievement results on the Tennessee Value-Added Assessment System for restructuring schools in Memphis*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Ross, T. (2005). *An investigation of the* Success for All *reading program in three Mississippi elementary schools*. Unpublished dissertation, Mississippi State University, Starkville. This study is ineligible for review because it does not use a comparison group.

Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers.

American Educational Research Journal, 44(2), 252–297. The study is ineligible for review because it does not examine the effectiveness of an intervention.

Sanders, W. L., Wright, S. P., Ross, S. M., & Wang, L. W. (2000). Value-added achievement results for three cohorts of Roots and Wings schools in Memphis: 1995–1999 outcomes. Retrieved from Success for All Foundation website: http://successforall.com/_images/pdfs/Ross_Roots_Wings_99.pdf. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Schneider, F. H. (1999). Impact of the *Success for All* program in the teaching of reading for third grade students in selected elementary schools in the Pasadena Independent School District. *Dissertation Abstracts International, 60*(06), 1965A. (UMI No. 9934489) The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Seligo Boehle, D. (2003). How schema appropriate alternate assessments affect the oral reading accuracy and oral reading fluency of selected first-grade students using the school-wide reform model *Success for All*. *Masters Abstracts International, 42*(02), 383. (UMI No. 1416013) This study is ineligible for review because it does not use a comparison group.

Shippen, M. E., Houchins, D. E., Calhoon, M. B., Furlow, C. F., & Sartor, D. L. (2006). The effects of comprehensive school reform models in reading for urban middle school students with disabilities. *Remedial and Special Education, 27*(6), 322–328. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Simpson, S. H. (1997). A principal's perspective of the implementation of *Reading Recovery* in six metropolitan Nashville elementary schools. *Dissertation Abstracts International, 58*(08), 2948A. (UMI No. 9806596) The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

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. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Slavin, R. E. (2005). *Evidence-based reform: Advancing the education of students at risk*. Washington, DC: Center for American Progress and the Institute for America's Future. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Slavin, R. E., Leighton, M., & Yampolsky, R. (1990). Success for All: *Effects on the achievement of limited English proficient children (Report No. 5)*. Baltimore, MD: The Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students. (ERIC Document Reproduction Service No. ED331585) The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Additional sources:

Slavin, R. E., & Yampolsky, R. (1991). Effects of *Success for All* on students with limited English proficiency: A three-year evaluation. Retrieved from Success for All Foundation website: http://www.successforall.net/_images/pdfs/englanlearn.htm.

Slavin, R. E., & Yampolsky, R. (1991). Success for All: *Effects on language minority students*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Slavin, R. E., & Madden, N. A. (1991). Success for All *at Buckingham Elementary: Second year evaluation*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Additional source:

Madden, N. A., Slavin, R. E., Karweit, N., Dolan, L., & Wasik, B. A. (1991). Success for All: *Multi-year effects of a schoolwide elementary restructuring program*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students. (Study: Worchester County, MD)

Slavin, R. E., & Madden, N. A. (1994, April). *Lee Conmigo: Effects of* Success for All *in bilingual first grades*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. This study is ineligible for review because it does not examine an intervention conducted in English.

Slavin, R. E., & Madden, N. A. (1998). Success for All/*Éxito Para Todos—Effects on the reading achievement of students acquiring English* (Report No. 19). Baltimore, MD: Center for Research on the Education of Students Placed at Risk, Johns Hopkins University. (ERIC Document Reproduction Service No. ED423327) This study is ineligible for review because it does not examine an intervention conducted in English.

Slavin, R. E., & Madden, N. A. (1999). Effects of bilingual and English as a second language adaptations of *Success for All* on the reading achievement of students acquiring English. *Journal of Education for Students Placed At Risk, 4*(4), 393–416. (Study: Philadelphia) The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Additional sources:

Madden, N. A., Slavin, R. E., Karweit, N., Dolan, L., & Wasik, B. A. (1991). Success for All: *Multi-year effects of a schoolwide elementary restructuring program*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students. (Study: Philadelphia, PA)

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1993). Success for All: *Evaluations of national replications* (Report No. 43). Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1994). *Implementing* Success for All *in the Philadelphia Public Schools: Final report to the Pew Foundation*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1996). *Success for All*: A summary of research. *Journal of Education for Students Placed At Risk, 1*(1), 41–76. (Study: Philadelphia, PA)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994). "Whenever and wherever we choose": The replication of *Success for All. Phi Delta Kappan, 75*(8), 639–647. (Study: Philadelphia, PA)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994, April). Success for All: *Longitudinal effects of systemic school-by-school reform in seven districts*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Philadelphia, PA)

Slavin, R. E., & Madden, N. A. (2006). Success for All/*Roots & Wings: 2006 summary of research on achievement outcomes*. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Slavin, R. E., & Madden, N. A. (2006). Reducing the gap: *Success for All* and the achievement of African American students. *Journal of Negro Education, 75*(3), 389–400. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Slavin, R. E., Madden, N. A., Cheung, A., & Liang, C. (2002). Success for All *in California: Gains on SAT-9 Reading and the Academic Performance Index*. Baltimore, MD: Success for All Foundation. This study is ineligible for review because it does

not disaggregate findings for the age or grade range specified in the protocol.

Slavin, R. E., Madden, N. A., & Datnow, A. (2007). Research in, research out: The role of research in the development and scale-up of *Success for All*. In S. Fuhrman, D. Cohen, & F. Mosher (Eds.), *The state of education policy research*. Mahwah, NJ: Lawrence Erlbaum Associates. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Slavin, R. E., Madden, N. A., Karweit, N., Livermon, B. J., & Dolan, L. (1990). *Success for All*: First-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal, 27*(2), 255–278. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Smith, L. J., Ross, S. M., & Casey, J. (1996). Multi-site comparison of the effects of *Success for All* on reading achievement. *Journal of Literacy Research, 28*(3), 329–353. (Study: Memphis, TN) The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Smith, L. J., Ross, S. M., & Casey, J. (1996). Multi-site comparison of the effects of *Success for All* on reading achievement. *Journal of Literacy Research, 28*(3), 329–353. (Study: Montgomery, AL) The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Additional sources:

Ross, S. M., Smith, L. J., Casey, J., Johnson, B., & Bond, C. (1994, April). *Using* Success for All *to restructure elementary schools: A tale of four cities*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Montgomery, AL)

Slavin, R. E., Madden, N. A., Dolan, L., Wasik, B. A., Ross, S. M., & Smith, L. J. (1994, April). Success for All: *Longitudinal effects of systemic school-by-school reform in seven districts*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. (Study: Montgomery, AL)

Smith, L. J., Ross, S. M., Johnson, B., & Casey, J. (1993). *Final report: 1992–1993 Memphis, Tennessee* SFA *results*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Smith-Davis, S. L. (2008). Does *Success for All* impact reading achievement of students with learning disabilities. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 68*(8-A), 3328. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

St. John, E. P., Manset, G., Chung, C., Simmons, A. B., & Musoba, G. D. (2000). *Research-based reading interventions: The impact of Indiana's Early Literacy Grant Program*. Bloomington: Indiana University, Indiana Education Policy Center, Smith Center for Research in Education. (ERIC Document Reproduction Service No. ED447466) This study is ineligible for review because it does not use a comparison group.

St. John, E. P., Manset, G., Chung, C., & Worthington, K. (2001). *Assessing the rationales for educational reforms: A test of the professional development, comprehensive reform, and direct instruction hypotheses*. Bloomington: Indiana University, Indiana Education Policy Center, Smith Center for Research in Education. (ERIC Document Reproduction Service No. ED458641) This study is ineligible for review because it does not include a student outcome.

Sterbinsky, A., Ross, S. M., & Redfield, D. (2002). *The effects of implementing comprehensive school reform models in 12 elementary schools: Year 3 study results*. Charleston, SC: Appalachia Educational Laboratory. This study is ineligible for

review because it does not disaggregate findings for the age or grade range specified in the protocol.

Additional source:

Sterbinsky, A., Ross, S., & Redfield, D. (2003, April). *Comprehensive school reform: A multi-site replicated experiment*. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Sterbinsky, A., Ross, S., & Redfield, D. (2006). Effects of comprehensive school reform on student achievement and school change: A longitudinal multi-site study. *School Effectiveness & School Improvement, 17*(3), 367–397. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Stevens, R. J., & Slavin, R. E. (1995). The Cooperative Elementary School: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal, 32*(2), 321–351. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Stringfield, S., Millsap, M. A., Herman, R., Yoder, N., Brigham, N., Nesselrodt, P., et al. (1997). Urban and suburban/rural special strategies for educating disadvantaged children: Findings and policy implications of a longitudinal study. Retrieved from Johns Hopkins University, Center for Social Organization of Schools website: http://www.csos.jhu.edu/Otherlinks/SpecialStrategies/index.htm. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Success for All Foundation. (2007). Independent reviews of *Success for All*. Retrieved from Success for All Foundation website: http://www.successforall.net/_images/pdfs/410054000_IndependReviews.pdf. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Thomas, R., & Lopes, J. (2007). Systematic reviews of learning by children with learning disabilities. *Psicologia: Teoria, Investigacao e Pratica, 12*(1), 3–12. The study is ineligible for review because it is not a primary analysis of the effectiveness of an intervention.

Tivnan, T., & Hemphill, L. (2005). Comparing four literacy reform models in high-poverty schools: Patterns of first-grade achievement. *The Elementary School Journal, 105*(5), 419–441. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Urdegar, S. M. (2000). *Evaluation of the* Success for All *program 1998–99*. Miami, FL: Miami-Dade County Public Schools, Office of Evaluation and Research. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Additional source:

Urdegar, S. M. (1998). *Evaluation of the* Success for All *program 1997–98*. Miami, FL: Miami-Dade Public Schools, Office of Educational Evaluation.

Veals, C. J. (2002). The impact of the *Success for All* reading program on the reading performance of third grade students in two southwest Mississippi schools. *Dissertation Abstracts International, 63*(04), 1291A. (UMI No. 3049586) This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Wang, L. W., & Ross, S. M. (1999). *Results for* Success for All *program: Alhambra School District*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. The study does not meet WWC evidence standards because the measures of effect cannot be attributed solely to the intervention—there was only one unit of analysis in one or both conditions.

Wang, L. W., & Ross, S. M. (2003). *Comparisons between elementary school programs on reading performance: Albuquerque Public Schools*. Memphis, TN: University of Memphis, Center for Research in Educational Policy. This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Weaver, L. M. (2005). An investigation of the *Success for All* reading program in three Mississippi schools. *Dissertation*

References *(continued)*

Abstracts International, 66(04A), 166–1310. This study is ineligible for review because it does not use a comparison group.

Wells, L. R. (2000). An investigation of the *Success for All* reading program at two Mississippi elementary schools. *Dissertation Abstracts International, 61*(04), 1342A. (UMI No. 9970370) This study is ineligible for review because it does not disaggregate findings for the age or grade range specified in the protocol.

Appendix

Appendix A1.1 Study characteristics: Borman, Slavin, Cheung, Chamberlain, Madden, & Chambers, 2006 (randomized controlled trial)

Characteristic	Description
Study citation	Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2006). *Final reading outcomes of the national randomized field trial of* Success for All®. Retrieved from Success for All website: http://www.successforall.net/_images/pdfs/Third_Year_Results_06.doc.
Participants	The study piloted the *SFA*® program in fall 2001, when three schools were randomly assigned to the *SFA*® and three schools to the comparison condition. In fall 2002, 35 new schools were recruited with 18 schools randomly assigned to implement *SFA*® in grades K–2, and 17 schools randomly assigned to serve as comparisons.[1] The study presented findings after the intervention students completed one, two, and three years of the program. For the effectiveness ratings, the WWC focused on findings from the longitudinal sample, that is, schools and students who completed three years of the program.[2] After three years, 18 *SFA*® schools with 707 students and 17 comparison schools with 718 students remained in the longitudinal sample.
Setting	The analysis sample included 35 elementary schools across 14 states located in rural and small towns in the South and urban areas of the Midwest.
Intervention	Intervention students received the *SFA*® school reform program, including the *SFA*® reading curriculum, tutoring for students' quarterly assessments, family support teams for students' parents, a facilitator who worked with school personnel, and training for all intervention teachers. Intervention schools implemented *SFA*® in grades K–2 and used their previously planned curriculum in grades 3–5. Some schools took a year to fully implement the program.
Comparison	Comparison schools continued using their regular, previously planned curriculum for grades K–2 (*SFA*® was implemented in grades 3–5). Authors conducted observations at all schools and indicated that there was no evidence that when *SFA*® was implemented in grades 3–5, students in grades K–2 were also exposed to *SFA*®. All sample students were pretested with the Peabody Picture Vocabulary Test (PPVT) prior to *SFA*® implementation, and schoolwide PPVT scores show equivalence between the program and comparison schools. Researchers also used information from the Common Core of Data (a database maintained by the National Center for Education Statistics) at several points over the course of the study to demonstrate the equivalence between the program and comparison schools on race/ethnicity, gender, English as a second language, special education, and free and reduced-price lunch. All equivalency tests were assessed at the school level, and no statistically significant differences were found.
Primary outcomes and measurement	Three subtests of the Woodcock Reading Mastery Test were administered during the period reflected in the intervention rating: Word Identification, Word Attack, and Passage Comprehension.[3] (See Appendices A2.1–A2.3 for more detailed descriptions of outcome measures.)
Staff/teacher training	*SFA*® teachers received three days of training during the summer and approximately eight days of on-site follow-up during the first implementation year. Success for All Foundation trainers visited classrooms, met with groups of teachers, looked at data on children's progress, and provided feedback to school staff on implementation quality and outcomes.

1. The 17 additional comparison schools implemented *SFA*® in grades 3–5, but students in grades K–2—the focus of this study and the WWC review—did not receive the intervention.
2. The study provided analysis for two samples, the "longitudinal sample," which included students who participated in the program for all three years, and the "in-mover sample," which included the longitudinal sample plus students who transferred into the school. The WWC analysis focuses on the longitudinal sample. The WWC prioritized outcomes that reflected students' exposure to the intervention for the longest period of time available. Findings reflecting students' outcomes after shorter periods of implementation can be found in Appendices A4.1–A4.6.
3. One additional subtest of the Woodcock Reading Mastery Test (Letter Identification) was administered during an earlier time period and is presented as an additional finding in Appendix A4.1

Appendix A1.2 Study characteristics: Dianda & Flaherty, 1995 (quasi-experimental design)

Characteristic	Description
Study citation	Dianda, M., & Flaherty, J. (1995, April). *Effects of* Success for All® *on the reading achievement of first graders in California bilingual programs*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
Participants	This study involved seven elementary schools in California in which the majority of students were English language learners. Six schools remained by the third year of program implementation. Students were grouped into four language categories and received instruction in English, Spanish, or "Sheltered English."[1] Only the English-speaking sub-sample was reviewed.[2] The report includes three cohorts of students who began participating in the study as kindergarteners in 1992 (99 intervention and 120 comparison students), 1993 (105 intervention and 62 comparison students), or 1994 (94 intervention and 59 comparison students), for a total of 539 participants. For the effectiveness rating, the WWC used data that reflected students' exposure to the intervention for the longest period of time, which varied for the different cohorts and domains.[3] Exact attrition rates are not known for this study; however, the post-attrition intervention and comparison samples were equivalent for the English-speaking subgroup. In the overall sample, the percentage of students eligible for free lunch varied from 70% to 98% in intervention schools, and from 47% to 80% in comparison schools. The percentages of minority students were between 50% and 70% for each study condition.
Setting	The analysis sample included seven elementary schools in California.
Intervention	Intervention students received the typical *SFA*® curriculum, including the *SFA*® reading curriculum, tutoring for students, quarterly assessments, family support teams for students' parents, a facilitator who worked with school personnel, and training for all intervention teachers.
Comparison	Comparison schools continued using their regular, previously planned curriculum. Each comparison school was matched with an *SFA*® school in the same district with students who had similar demographics and pretest scores on the Peabody Picture Vocabulary Test measure.
Primary outcomes and measurement	Three subtests of the Woodcock Language Proficiency Battery were administered: Letter-Word Identification, Word Attack, and Passage Comprehension. The authors presented findings from each Woodcock subtest separately and also pooled findings from the Woodcock Letter-Word Identification subtests (see Appendices A2.1–A2.3 for more detailed descriptions of outcome measures).
Staff/teacher training	*SFA*® teachers received three days of training during the summer and approximately eight days of on-site follow-up during the first implementation year. Success for All Foundation trainers visited classrooms, met with groups of teachers, looked at data on children's progress, and provided feedback to school staff on implementation quality and outcomes. Specially trained certified teachers or qualified aides worked one-to-one with the students.

1. English language learners participate in *SFA*® in English alongside their English-dominant classmates during a common period in the morning. During the rest of the day, they receive sheltered content instruction or ESL instruction, depending on their level of English proficiency.
2. The WWC Beginning Reading topic focuses only on students learning to read in English (see Beginning Reading Protocol).
3. Findings include outcomes after two years of exposure for the alphabetics and comprehension domains, and after two (1994 cohort), three (1993 cohort), and four (1992 cohort) years of exposure for the general reading domain. Findings reflecting students' outcomes after shorter periods of implementation can be found in Appendix A4.3.

Appendix A1.3 Study characteristics: Madden, Slavin, Karweit, Dolan, & Wasik, 1993 (quasi-experimental design)

Characteristic	Description
Study citation	Madden, N. A., Slavin, R. E., Karweit, N., Dolan, L., & Wasik, B. A. (1993). *Success for All®*: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30*(1), 123–148.
Participants	The study investigated the effects of two versions of the *SFA®* program: full implementation and dropout prevention. Although these versions varied in their implementation of the whole-school reform model, the reading curricula are essentially the same at all schools, with each school receiving the same training, coaching support, and materials. Ratings presented in this report are not disaggregated by the variations in implementation of whole-school reforms. Within each comparison school, one-third of the students were randomly selected for testing purposes. The study focused on cohorts of students who started *SFA®* in prekindergarten, kindergarten, and first grade and received the intervention for multiple years. To determine the effectiveness ratings, the WWC focused on the latest term results available. The third-year analytic sample included 671 students within five *SFA®* schools and 671 students within five comparison schools spread across three grade levels.[1] African-American students constituted 97% to 100% of students in five intervention schools, with 83% to 98% of students qualified for free lunch. In comparison (Chapter 1) schools, at least 75% of students qualified for free lunch.
Setting	The analysis sample included 10 elementary schools in Baltimore, Maryland.
Intervention	Intervention students in the full implementation version received the typical *SFA®* program, including the *SFA®* reading curriculum, tutoring for students in grades 1–3, quarterly assessments, family support teams for students' parents, a facilitator who worked with school personnel, and training for all intervention teachers. Intervention students in the dropout prevention version had a reduced number of tutors and family support staff. Chapter 1 monies supported the dropout prevention program.
Comparison	The comparison condition included schools that implemented a traditional reading program built around *Macmillan Connections* basal series. Each comparison school was matched with an intervention school based on the percentage of students getting free or reduced-price lunch and historical achievement level. Students were then individually matched on a standardized test given by the school district. Pretest scores on WRMT Letter-Word Identification, Word Attack, and Durrell Oral Reading subtests served as covariates in analyses.
Primary outcomes and measurement	Two subtests of the Woodcock Language Proficiency Battery were administered: *Letter-Word Identification* and *Word Attack*. Additional measures included Durrell Analysis of Reading Difficulty Silent Reading and Oral Reading subtests and the California Achievement Test (CAT) Total Reading (see Appendices A2.1–A2.3 for more detailed descriptions of outcome measures).
Staff/teacher training	The teachers and tutors were regular certified teachers. They received detailed teacher's manuals supplemented by two to three days of in-service at the beginning of the school year. For teachers of grades 1–3 and for reading tutors, these training sessions focused on the implementation of the reading program. Preschool and kindergarten teachers and aids were trained in the use of the thematic units, and other aspects of the preschool and kindergarten models. School facilitators also organized many information sessions to allow teachers to share problems and solutions, suggest changes, and discuss individual children.

1. Additional findings for a subsample of low-achieving students can be found in Appendices A4.4–A4.6.

Appendix A1.4 Study characteristics: Ross, Alberg, & McNelis, 1997 (quasi-experimental design)

Characteristic	Description
Study citation	Ross, S. M., Alberg, M., & McNelis, M. (1997). *Evaluation of elementary school school-wide programs: Clover Park School District. Year 1: 1996–97.* Memphis, TN: The University of Memphis, Center for Research in Education Policy.
Participants	The study compared whole-school improvement programs, *Success for All®*, *Accelerated Schools*, and locally developed programs, in 19 schools. Schools were divided into four groups based on the similarity of several school characteristics, including enrollment, percentage of minority students, percentage of students eligible for free/reduced-price lunch, and initial academic performance. The WWC focused on only one group, "cluster 2A," the third highest with respect to socioeconomic status, which included three *SFA®* schools and three *Accelerated Schools*, with a total number of 252 first-grade students (148 students who attended *SFA®* schools, 104 students who attended *Accelerated Schools*).[1] The study included data that reflected students' outcomes after one year of program implementation. In the overall sample, the percentage of minority students in three intervention schools was between 47% and 63%. In the three comparison schools, the range was between 42% and 54%. The percentage of students eligible for free/reduced-price lunch varied from 63% to 66% in intervention schools, and from 66% to 71% in comparison schools.
Setting	The analysis sample included six elementary schools in Clover Park, Washington.
Intervention	Intervention students received the typical *SFA®* program, including the *SFA®* reading curriculum, tutoring for students in grades 1–3, quarterly assessments, family support teams for students' parents, a facilitator who worked with school personnel, and training for all intervention teachers.
Comparison	*Accelerated Schools* is a comprehensive school reform program that is designed to close the achievement gap between at-risk and not at-risk children. The program redesigns and integrates curricular, instructional, and organizational practices so that they provide enrichment for at-risk students.
Primary outcomes and measurement	Three subtests of the Woodcock Reading Mastery Test were administered: Word Identification, Word Attack, and Passage Comprehension. The Durrell Analysis of Reading Difficulty Oral Reading subtest was also used (see Appendices A2.1–A2.3 for more detailed descriptions of outcome measures).
Staff/teacher training	No information on training for the specific teachers in this study was provided.

1. An additional group included one *SFA®* school and three comparison schools (one school used *Accelerated Schools* design, and the other two used locally developed programs), but this comparison did not meet WWC evidence screens because the effect of *SFA®* could not be separated from the effect of the school.

Appendix A1.5 Study characteristics: Ross & Casey, 1998 (quasi-experimental design)

Characteristic	Description
Study citation	Ross, S. M., & Casey, J. (1998). *Longitudinal study of student literacy achievement in different Title I school-wide programs in Fort Wayne Community Schools year 2: First grade results*. Memphis, TN: The University of Memphis, Center for Research in Education Policy.
Participants	This study examines the effects of *SFA*® in two Title I schools by comparing them with five other Title I schools that were implementing locally developed schoolwide programs.[1] The study did not report on the initial sample size, but 288 students in kindergarten (83 students in the *SFA*® schools, 205 students at comparison schools) were included in the final analysis sample, and the post-attrition intervention and comparison samples were equivalent on the achievement pretest measure (PPVT). The study included data that reflected students' outcomes after two years of program implementation.[2] School populations ranged between 31% and 50% minority enrollment; between 62% and 81% of students received free or reduced-price lunch.
Setting	The analysis sample included seven Title I elementary schools in Fort Wayne, Indiana.
Intervention	Intervention students received the typical *SFA*® curriculum, including the Reading Roots reading curriculum in grade 1 and the Reading Wings reading curriculum in grade 2; one-to-one tutoring for the lowest-achieving students by certified teacher tutors, quarterly assessments, family support teams for students' parents, a facilitator who worked with school personnel, and training for all intervention teachers.
Comparison	The five comparison schools implemented locally developed schoolwide programs. The schools were comparable with *SFA*® schools on pretest PPVT measures, socio-economic status, and ethnicity. Four out of the five local school programs incorporated components of other branded programs, including *Reading Recovery*, *Accelerated Reader*, *Four-Block*, and *STAR*. These curricula place considerable emphasis on reading, use of basal readers, and multifaceted reading activities.
Primary outcomes and measurement	Three subtests of the Woodcock Reading Mastery Test were administered: Word Identification, Word Attack, and Passage Comprehension. The study presented a combined measure of Word Identification and Word Attack. The Durrell Analysis of Reading Difficulty Oral Reading subtest was also used (see Appendices A2.1–A2.3 for more detailed descriptions of outcome measures).
Staff/teacher training	No information on training for the specific teachers was provided in this study.

1. The article reported on an additional intervention school that supplemented *SFA*® with another branded intervention (*Reading Recovery*), but results from this portion of the study do not meet WWC evidence standards because the effect of *SFA*® cannot be separated from the effect of *Reading Recovery*.
2. Additional findings for a subsample of low-achieving students (that is, lowest 25% with respect to reading achievement) are reported in Appendices A4.1–A4.6.

Appendix A1.6 Study characteristics: Ross, McNelis, Lewis, & Loomis, 1998 (quasi-experimental design)

Characteristic	Description
Study citation	Ross, S. M., McNelis, M., Lewis, T., & Loomis, S. (1998). *Evaluation of* Success for All® *programs: Little Rock school district year 1: 1997–1998*. Memphis, TN: The University of Memphis, Center for Research in Education Policy.
Participants	This study involved 97 first-grade students with both pretest and posttest data in four schools. Two schools implemented the *Success for All*® program (40 students), and two schools were selected as their matched comparison schools (47 students). The *SFA*® schools and the comparison schools were similar in poverty level, achievement level, and enrollment. The study reported data on students' outcomes after one year of program implementation.
Setting	The study took place in four elementary schools in Little Rock, Arkansas.
Intervention	Intervention students received the typical *SFA*® program, including the *SFA*® reading curriculum, tutoring for students in grades 1–3, quarterly assessments, family support teams for students' parents, a facilitator who worked with school personnel, and training for all intervention teachers.
Comparison	No information was provided on the nature of the comparison curriculum. The two comparison schools were matched to the *SFA*® schools based on poverty level, achievement level, and enrollment. Pretest PPVT scores were used as a covariate to adjust for differences in students' abilities.
Primary outcomes and measurement	Three subtests of the Woodcock Reading Mastery Test were administered: Word Identification, Word Attack, and Passage Comprehension. The Durrell Analysis of Reading Difficulty Oral Reading subtest was also used (see Appendices A2.1–A2.3 for more detailed descriptions of outcome measures).
Staff/teacher training	No information on training for the teachers in this study was provided.

Appendix A1.7 Study characteristics: Smith, Ross, Faulks, Casey, Shapiro, & Johnson, 1993 (quasi-experimental design)

Characteristic	Description
Study citation	Smith, L. J., Ross, S. M., Faulks, A., Casey, J., Shapiro, M., & Johnson, B. (1993). *1991–1992 Ft. Wayne, Indiana* SFA® *results*. Memphis, TN: The University of Memphis, Center for Research in Education Policy.
Participants	This study involved approximately 286 students in kindergarten and first grade in four elementary schools in Fort Wayne, Indiana. Two schools implemented the *SFA*® program. Two comparison schools were matched to the intervention schools based on poverty level, historical achievement level, and ethnicity; then pairs of students were matched on PPVT pretest scores. There were 74 kindergarteners and 69 first-grade students in the intervention group, and 74 kindergarteners and 69 first-grade students in the comparison group. Exact student attrition rates are not known for this study; however, the post-attrition intervention and comparison samples were equivalent on achievement pretest. School-level data—poverty level, achievement, and enrollment—were similar across all schools. The study included data on students' outcomes after one year of program implementation.[1]
Setting	The study took place in four elementary schools in Fort Wayne, Indiana.
Intervention	Intervention students received the typical *SFA*® program, including the *SFA*® reading curriculum, tutoring for students, quarterly assessments, family support teams for students' parents, a facilitator who worked with school personnel, and training for all intervention teachers.
Comparison	Comparison schools continued using their regular, previously planned curriculum. No other information was provided on the comparison curriculum.
Primary outcomes and measurement	Four subtests of the Woodcock Reading Mastery Test were used: Letter Identification, Word Identification, Word Attack, and Passage Comprehension. Additional measures included the Peabody Picture Vocabulary Test and Durrell Analysis of Reading Difficulty Oral Reading subtest. The Merrill Language Screening Test and the Test of Language Development were also administered but have not been included in this review because they were outside the scope of the Beginning Reading review (see Appendices A2.1–A2.3 for more detailed descriptions of outcome measures).
Staff/teacher training	Teachers in their first year of teaching *SFA*® classes received three days of summer training and two to four additional in-service days during the school year. A school facilitator monitored and provided feedback throughout the year. Twice a year, trainers provided by the developer visited and observed teachers. After the first year, training was reinforced by regular in-services, an annual *SFA*® conference, and implementation checks for the facilitators and trainers.

1. Additional findings for a low-achieving subset of students (lowest 25% with respect to reading achievement) are presented in Appendices A4.1–A4.6.

Appendix A2.1 Outcome measures in the alphabetics domain by construct

Outcome measure	Description
Letter knowledge	
Woodcock Reading Mastery Test (WRMT): Letter Identification subtest	The standardized test measures the number of letters that students are able to identify correctly (Smith et al., 1993).
Phonics	
WRMT: Word Identification subtest	The Word Identification subtest is a test of decoding skills. The standardized test requires the child to read aloud isolated real words that range in frequency and difficulty (as cited in Borman et al., 2006; Ross, Alberg, & McNelis, 1997; Ross & Casey, 1998; Ross et al., 1998; Smith et al., 1993).
Woodcock Language Proficiency Battery (WLPB): Letter-Word Identification subtest	The Letter/Word Identification subtest is a standardized test that requires the child to read aloud isolated letters and real words that range in frequency and difficulty (as cited in Dianda & Flaherty, 1995; Madden et al., 1993).
WRMT and WLPB: Word Attack subtest	The standardized test measures phonemic decoding skills by asking students to read pseudowords. Students are aware that the words are not real (as cited in Borman et al., 2006; Dianda & Flaherty, 1995; Madden et al., 1993; Ross, Alberg, & McNelis, 1997; Ross & Casey, 1998; Ross et al., 1998; Smith et al., 1993).

Appendix A2.2 Outcome measures in the comprehension domain by construct

Outcome measure	Description
Reading comprehension	
WRMT and WLPB: Passage Comprehension subtest	In this standardized test, comprehension is measured by having students fill in missing words in a short paragraph (as cited in Borman et al., 2006; Dianda & Flaherty, 1995; Ross, Alberg, & McNelis, 1997; Ross & Casey, 1998; Ross et al., 1998; Smith et al., 1993).
Durrell Analysis of Reading Difficulty (DARD): Silent Reading Test	An individually administered, standardized diagnostic test that measures reading rate while students read passages silently and answer comprehension questions (as cited in Madden et al., 1993).
Vocabulary development	
Peabody Picture Vocabulary Test (PPVT)	A standardized, receptive vocabulary test that asks students to choose which one of four pictures corresponds to a test word spoken aloud (as cited in Smith et al., 1993).

Appendix A2.3 Outcome measures in the general reading domain by construct

Outcome measure	Description
California Achievement Test (CAT) Total Reading	A group-administered, standardized assessment battery composed of numerous reading and language-oriented subtests (as cited in Madden et al., 1993).
DARD Oral Reading Test	An individually administered, standardized diagnostic test that measures reading accuracy, reading rate, and oral reading comprehension (as cited in Madden et al., 1993; Ross, Albert, & McNelis, 1997; Ross & Casey, 1998; Ross et al., 1998; Smith et al., 1993).

Appendix A3.1 Summary of findings for all domains[1]

Outcome measure	Domain: Alphabetics – Letter Identification	Alphabetics – Phonics	Comprehension – Reading Comprehension	Comprehension – Vocabulary Development	General Reading Achievement
Meets evidence standards					
Borman et al., 2006	nr	+	+	nr	nr
Meets evidence standards with reservations					
Dianda & Flaherty, 1995	nr	(+)	(+)	nr	(+)
Madden et al., 1993	nr	+	nr	nr	(+)
Ross, Alberg, & McNelis, 1997	nr	ind	ind	nr	ind
Ross & Casey, 1998	nr	ind	ind	nr	ind
Ross et al., 1998	nr	(+)	ind	nr	ind
Smith et al., 1993	(+)	(+)	(+)	ind	(+)
Rating of Effectiveness	**Positive**		**Mixed Effects**		**Potentially Positive**

nr = no reported outcomes under this construct

+ = study finding was positive and statistically significant

(+) = study finding was positive and substantively important, but not statistically significant

ind = study finding was indeterminate, that is, neither substantively important nor statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices in each domain. More detailed information on findings for all measures within the domains and the constructs that factor into the domains can be found in Appendices A3.2–A3.4.

Appendix A3.2 Summary of findings for alphabetics domain[1]

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Borman et al., 2006 (randomized controlled trial)[8]—Three years of intervention									
WRMT: Word ID subtest[9]	Phonics	Kindergarten	35/1,425	462.96 (23.56)	457.41 (25.72)	5.55	0.22	Statistically significant	+9
WRMT: Word Attack subtest[9]	Phonics	Kindergarten	35/1,425	493.43 (16.45)	487.73 (17.64)	5.70	0.33	Statistically significant	+13
Madden et al., 1993 (quasi-experimental design)[8,10]—Three years of intervention									
WLPB: Letter-Word ID subtest	Phonics	Prekindergarten (Cohort 1)	10/492	18.53 (5.34)	15.91 (6.59)	2.62	0.44	ns	+17
WLPB: Word Attack subtest	Phonics	Prekindergarten (Cohort 1)	10/492	5.46 (4.11)	2.25 (3.55)	3.21	0.83	Statistically significant	+30
WLPB: Letter-Word ID subtest	Phonics	Kindergarten (Cohort 2)	10/440	25.09 (6.65)	21.54 (6.72)	3.55	0.53	ns	+20
WLPB: Word Attack subtest	Phonics	Kindergarten (Cohort 2)	10/440	8.63 (6.27)	5.21 (4.76)	3.42	0.61	ns	+23
WLPB: Letter-Word ID subtest	Phonics	Grade 1 (Cohort 3)	10/410	28.69 (6.72)	25.56 (6.19)	3.12	0.48	ns	+19
WLPB: Word Attack subtest	Phonics	Grade 1 (Cohort 3)	10/410	10.77 (6.94)	7.02 (5.49)	3.74	0.60	ns	+23
Dianda & Flaherty, 1995 (quasi-experimental design)[8]—Two years of intervention									
WLBP: Letter-Word ID subtest	Phonics	English-speaking kindergarten (1992 cohort)	7/219	nr	nr	na	0.34[11]	ns	+13
WLBP: Word Attack subtest	Phonics	English-speaking kindergarten (1992 cohort)	7/219	nr	nr	na	0.26[11]	ns	+10

(continued)

Appendix A3.2 Summary of findings for alphabetics domain[1] *(continued)*

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Ross & Casey, 1998 (quasi-experimental design)[8]—Two years of intervention									
WRMT: Word ID subtest	Phonics	Kindergarten	7/288	32.14 (14.63)	31.30 (14.20)	0.84	0.06	ns	+2
WRMT: Word Attack subtest	Phonics	Kindergarten	7/288	12.25 (7.36)	10.40 (8.20)	1.85	0.23	ns	+9
Ross, Alberg, & McNelis, 1997 (quasi-experimental design)[8]—One year of intervention									
WRMT: Word ID subtest	Phonics	Grade 1	6/252	nr	nr	na	–0.01[12]	ns	0
WRMT: Word Attack subtest	Phonics	Grade 1	6/252	18.35	15.86	2.49 (8.89)[13]	0.28[12]	ns	+11
Ross et al., 1998 (quasi-experimental design)[8]—One year of intervention									
WRMT: Word ID subtest	Phonics	Grade 1	4/97	38.27	36.21	2.06 (12.31)[14]	0.17	ns	+7
WRMT: Word Attack subtest	Phonics	Grade 1	4/97	15.17	11.19	3.98 (8.89)[14]	0.44	ns	+17
Smith et al., 1993 (quasi-experimental design)[8]—One year of intervention									
WRMT: Word ID subtest	Phonics	Kindergarten (Cohort 1)	4/148	10.26 (9.82)	3.15 (4.95)	7.11	0.91	ns	+32
WRMT: Letter ID subtest	Letter knowledge	Kindergarten (Cohort 1)	4/148	32.43 (4.28)	29.36 (7.81)	3.07	0.48	ns	+19
WRMT: Letter ID subtest	Letter knowledge	Grade 1 (Cohort 2)	4/138	nr	nr	na	0.08[11]	ns	+3
WRMT: Word ID subtest	Phonics	Grade 1 (Cohort 2)	4/138	35.04 (10.63)	28.00 (14.70)	7.04	0.55	ns	+21
WRMT: Word Attack subtest	Phonics	Grade 1 (Cohort 2)	4/138	12.60 (7.43)	7.90 (7.91)	4.70	0.61	ns	+23

(continued)

Appendix A3.2 Summary of findings for alphabetics domain[1] *(continued)*

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Averages for alphabetics[15]									
Borman et al., 2006—Three years of intervention							0.28	Statistically significant	+11
Madden et al., 1993—Three years of intervention							0.58	ns	+22
Dianda & Flaherty, 1995—Two years of intervention							0.30	ns	+12
Ross & Casey, 1998—Two years of intervention							0.14	ns	+6
Ross, Alberg, & McNelis, 1997—One year of intervention							0.13	ns	+5
Ross et al., 1998—One year of intervention							0.31	ns	+12
Smith et al., 1993—One year of intervention							0.56	ns	+21
Domain average for alphabetics across all studies							**0.33**	**na**	**+13**
Averages by years of *SFA®* implementation									
Average of results from studies with three years of intervention (two studies)							0.43	na	+17
Average of results from studies with two years of intervention (two studies)							0.22	na	+9
Average of results from studies with one year of intervention (three studies)							0.33	na	+13

na = not applicable nr = not reported ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Earlier findings from longitudinal studies are not included in these ratings but are reported in Appendix A4.1. Subgroup findings from the studies are not included in these ratings but are reported in Appendix A4.4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time the pretest is administered. For example, kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Borman et al. (2006), a correction for multiple comparisons was needed so the significance levels may differ from those reported in the original study. There was no need to adjust for clustering because the findings were based on HLM analyses. In the case of the six other studies, corrections for both clustering and multiple comparisons were needed so the significance levels may differ from those reported in the original studies.
9. Standard deviations and adjusted means have been received through communication with the author (G. Borman, personal communication, 2006).

(continued)

Appendix A3.2 Summary of findings for alphabetics domain[1] *(continued)*

10. WWC combined means and standard deviations for the *SFA*® schools and for the control schools. Adjusted posttest means (with pretests standard scores as covariates) were used for effect size calculations. Kindergarten and grade 1 cohorts from Abbottston elementary school received four years of intervention.
11. Authors reported effect sizes that used the comparison group standard deviation in the denominator (Glass's delta). Effect size was computed by subtracting the comparison group mean from the intervention group mean and dividing the result by the comparison group standard deviation.
12. Authors reported effect sizes adjusted for PPVT pretest scores.
13. The WWC derived the pooled standard deviation from the reported means and effect size.
14. Authors reported the pooled standard deviation.
15. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A3.3 Summary of findings for comprehension domain[1]

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Borman et al., 2006 (randomized controlled trial)[8]—Three years of intervention									
WRMT: Passage Comprehension subtest[9]	Reading comprehension	Kindergarten	35/1,425	481.41 (14.20)	478.33 (15.33)	3.08	0.21	Statistically significant	+8
Dianda & Flaherty, 1995 (quasi-experimental design)[8]—Two years of intervention									
WLPB: Passage Comprehension subtest	Reading comprehension	English-speaking kindergarten (1992 cohort)	7/219	nr	nr	na	0.44	ns	+17
Ross & Casey, 1998 (quasi-experimental design)[8]—Two years of intervention									
WRMT: Passage Comprehension subtest	Reading comprehension	Kindergarten	7/288	16.09 (8.46)	15.40 (8.70)	0.69	0.08	ns	+3
Ross, Alberg, & McNelis, 1997 (quasi-experimental design)[8]—One year of intervention									
WRMT: Passage Comprehension subtest	Reading comprehension	Grade 1	6/252	nr	nr	na	0.01[11]	ns	0
Ross et al., 1998 (quasi-experimental design)[8]—One year of intervention									
WRMT: Passage Comprehension subtest	Reading comprehension	Grade 1	4/97	19.19	17.73	1.46 (8.19)[12]	0.18	ns	+7
Smith et al., 1993 (quasi-experimental design)[8]—One year of intervention									
Peabody Picture Vocabulary Test Vocabulary	Vocabulary development	Kindergarten (Cohort 1)	4/148	nr	nr	na	0.17[10]	ns	+7
WRMT: Passage Comprehension subtest	Reading comprehension	Grade 1 (Cohort 2)	4/136	16.37 (8.07)	13.91 (9.31)	2.46	0.28	ns	+11
Averages for comprehension[13]									
Borman et al., 2006—Three years of intervention							0.21	Statistically significant	+8
Dianda & Flaherty, 1995—Two years of intervention							0.44	ns	+17

(continued)

Appendix A3.3 Summary of findings for comprehension domain[1] *(continued)*

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Ross & Casey, 1998—Two years of intervention							0.08	ns	+3
Ross, Alberg, & McNelis, 1997—One year of intervention							0.01	ns	0
Ross et al., 1998—One year of intervention							0.18	ns	+7
Smith et al., 1993—One year of intervention							0.23	ns	+9
Domain average for comprehension across all studies							**0.19**	**na**	**+8**
Averages by years of *SFA®* implementation									
Results from study with three years of intervention (one study)							0.21	Statistically significant	+8
Average of results from studies with two years of intervention (two studies)							0.26	na	+10
Average of results from studies with one year of intervention (three studies)							0.14	na	+6

na = not applicable nr = not reported ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Earlier findings from longitudinal studies are not included in these ratings but are reported in Appendix A4.2. Subgroup findings from the studies are not included in these ratings but are reported in Appendix A4.5.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time pretest is administered. For example, the kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Borman et al. (2006), there was no need to adjust for clustering because the findings were based on HLM analyses. In the cases of Dianda and Flaherty (1995), Ross and Casey (1998), Ross, Alberg, and McNelis (1997), and Ross et al. (1998), a correction for clustering was needed so the significance levels may differ from those reported in the original studies. In the case of Smith et al. (1993), correction for both clustering and multiple comparisons were needed so the significance levels may differ from those reported in the original study.
9. Standard deviations and adjusted means have been received through communication with the author.
10. Authors reported effect sizes that used the comparison group standard deviation in the denominator (Glass's delta). Effect size was computed by subtracting the comparison group mean from the intervention group mean and dividing the result by the comparison group standard deviation.
11. Authors reported effect sizes adjusted for PPVT pretest scores.
12. Authors reported the pooled standard deviation.
13. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A3.4 Summary of findings for general reading achievement domain[1]

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Dianda & Flaherty, 1995 (quasi-experimental design)[8, 9]									
Four years of intervention									
3 WLPB subtests and Durrell Reading subtest combined	General reading	English-speaking kindergarten (1992 cohort)	6/136	nr	nr	na	0.23[10]	ns	+9
Three years of intervention									
3 WLPB subtests and Durrell Reading subtest combined	General reading	English-speaking kindergarten (1993 cohort)	6/167	nr	nr	na	0.34[10]	ns	+13
Two years of intervention									
3 WLPB subtests and Durrell Reading subtest combined	General reading	English-speaking kindergarten (1994 cohort)	6/153	nr	nr	na	0.27[10]	ns	+11
Madden et al., 1993 (quasi-experimental design)[9,11]—Three years of intervention									
Durrell Oral Reading subtest	General reading	Prekindergarten (Cohort 1)	10/492	5.59 (4.78)	4.26 (5.16)	1.33	0.27	ns	+11
Durrell Oral Reading subtest	General reading	Kindergarten (Cohort 2)	10/440	11.99 (7.28)	8.84 (6.05)	3.15	0.47	ns	+18
Durrell Oral Reading subtest	General reading	Grade 1 (Cohort 3)	10/410	16.66 (7.00)	13.25 (7.13)	3.41	0.48	ns	+19
Ross & Casey, 1998 (quasi-experimental design)[9]—Two years of intervention									
Durrell Oral Reading subtest	General reading	Kindergarten	7/288	5.35 (4.63)	4.70 (4.30)	0.65	0.15	ns	+6
Ross, Alberg, & McNelis, 1997 (quasi-experimental design)[9]—One year of intervention									
Durrell Oral Reading subtest	General reading	Grade 1	6/252	nr	nr	na	0.04[12]	ns	+2

(continued)

Appendix A3.4 Summary of findings for general reading achievement domain[1] *(continued)*

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Ross et al., 1998 (quasi-experimental design)[9]—One year of intervention									
Durrell Oral Reading subtest	General reading	Grade 1	4/97	7.01	6.46	0.55 (3.52)[13]	0.16	ns	+6
Smith et al., 1993 (quasi-experimental design)[9]—One year of intervention									
Durrell Oral Reading subtest	General reading	Grade 1	4/138	6.74 (4.25)	4.68 (3.83)	2.06	0.51	ns	+19
Averages for general reading achievement[14]									
Dianda & Flaherty, 1995[10]—Two to four years of intervention							0.28	ns	+11
Madden et al., 1993—Three years of intervention							0.41	ns	+16
Ross & Casey, 1998—Two years of intervention							0.15	ns	+6
Ross, Alberg, & McNelis, 1997—One year of intervention							0.04	ns	+2
Ross et al., 1998—One year of intervention							0.16	ns	+6
Smith et al., 1993—One year of intervention							0.51	ns	+19
Domain average for general reading achievement across all studies							**0.26**	**na**	**+10**
Averages by years of *SFA®* implementation									
Results from study with four years of intervention (one study)							0.23	ns	+9
Average of results from studies with three years of intervention (two studies)							0.37	na	+14
Average of results from studies with two years of intervention (two studies)							0.21	ns	+8
Average of results from studies with one year of intervention (three studies)							0.24	na	+9

na = not applicable nr = not reported ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Earlier findings from longitudinal studies are not included in these ratings but are reported in Appendix A4.3. Subgroup findings from the studies are not included in these ratings but are reported in Appendix A4.6.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time the pretest is administered. For example, kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.

(continued)

Appendix A3.4 Summary of findings for general reading achievement domain[1] *(continued)*

6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. Data are taken from Livingston and Flaherty (1997).
9. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the cases of Dianda and Flaherty (1995), Madden et al. (1993), and Smith et al. (1993), a correction for clustering and multiple comparisons was needed so the significance levels may differ from those reported in the original studies. In the cases of Ross and Casey (1998), Ross, Alberg, and McNelis (1997), and Ross et al. (1998), a correction for clustering was needed so the significance levels may differ from those reported in the original studies.
10. Authors reported effect sizes that used the comparison group standard deviation in the denominator (Glass's delta). Effect size was computed by subtracting the comparison group mean from the intervention group mean and dividing the result by the comparison group standard deviation.
11. WWC combined means and standard deviations for the *SFA*® schools and for the control schools. Adjusted posttest means (with pretests standard scores as covariates) were used for effect size calculations. Kindergarten and grade 1 cohorts from Abbottston elementary school received four years of intervention.
12. Authors reported effect sizes adjusted for PPVT pretest scores.
13. Authors reported the pooled standard deviation.
14. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4.1 Summary of earlier findings from longitudinal studies for alphabetics domain[1]

				Authors' findings from the study					
				Mean outcome (standard deviation)[2]		WWC calculations			
Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	*Success for All®* group	Comparison group	Mean difference[4] (*SFA®*–comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Borman et al., 2006 (randomized controlled trial)[8]—Two years of intervention									
WRMT: Letter ID subtest	Letter knowledge	Kindergarten and Grade 1	38/3,353	451.42 (14.08)	449.46 (11.19)	1.96	0.15	ns	+6
WRMT: Word ID subtest	Phonics	Kindergarten and Grade 1	38/3,353	449.52 (28.31)	444.82 (29.18)	4.70	0.16	ns	+6
WRMT: Word Attack subtest	Phonics	Kindergarten and Grade 1	38/3,353	487.92 (18.20)	483.29 (19.82)	4.63	0.24	Statistically significant	+10

ns = not statistically significant

1. This appendix presents earlier longitudinal findings for measures that fall in the alphabetics domain. Data that reflected students' exposure to the intervention for the longest period of time were used for intervention rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time the pretest is administered. For example, kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not applied to findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Borman et al. (2006), there was no need to adjust for clustering because the data were based on HLM analyses.

Appendix A4.2 Summary of earlier findings from longitudinal studies for comprehension domain[1]

				Authors' findings from the study					
				Mean outcome (standard deviation)[2]		WWC calculations			
Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	*Success for All®* group	Comparison group	Mean difference[4] (*SFA®*– comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Borman et al., 2006 (randomized controlled trial)[8]—Two years of intervention									
WRMT: Passage Comprehension subtest	Reading comprehension	Kindergarten and Grade 1	38/3,353	472.00 (18.29)	469.87 (19.53)	2.13	0.11	ns	+4

ns = not statistically significant

1. This appendix presents earlier longitudinal findings for measures that fall in comprehension domain. Data that reflected students' exposure to the intervention for the longest period of time were used for intervention rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time the pretest is administered. For example, kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Borman et al. (2006), there was no need to adjust for clustering because the findings were based on HLM analyses.

Appendix A4.3 Summary of earlier findings from longitudinal studies for general reading achievement domain[1]

				Authors' findings from the study					
				Mean outcome (standard deviation)[2]		WWC calculations			
Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	*Success for All®* group	Comparison group	Mean difference[4] (*SFA®*– comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Dianda & Flaherty, 1995 (quasi-experimental design)[8,9]									
Three years of intervention									
3 WLPB subtests and Durrell Reading subtest combined	General reading	English-speaking kindergarten (1992 cohort)	6/136	nr	nr	na	0.44[10]	ns	+17
Two years of intervention									
3 WLPB subtests and Durrell Reading subtest combined	General reading	English-speaking kindergarten (1993 cohort)	6/167	nr	nr	na	0.87[10]	Statistically significant	+31

na = not applicable nr = not reported ns = not statistically significant

1. This appendix presents earlier longitudinal findings for measures that fall in general reading domain. Data that reflected students' exposure to the intervention for the longest period of time were used for intervention rating purposes and are presented in Appendix A3.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time the pretest is administered. For example, kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. Data are taken from Livingston and Flaherty (1997).
9. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Dianda and Flaherty (1995), a correction for clustering was needed so the significance levels may differ from those reported in the original study.
10. Authors reported effect sizes that used comparison group standard deviation in the denominator (Glass's delta). Effect size was computed by subtracting the comparison group mean from the intervention group mean and dividing the result by the comparison group standard deviation.

Appendix A4.4 Summary of subgroup findings for alphabetics domain[1]

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Madden et al., 1993 (quasi-experimental design)[8,9]—Three years of intervention									
WLPB: Letter-Word ID subtest	Phonics	Prekindergarten/ lowest 25% (Cohort 1)	10/126	16.65 (5.34)	12.56 (6.66)	4.09	0.67	Statistically significant	+25
WLPB: Word Attack subtest	Phonics	Prekindergarten/ lowest 25% (Cohort 1)	10/126	4.92 (4.38)	1.52 (3.39)	3.40	0.86	Statistically significant	+31
WLPB: Letter-Word ID subtest	Phonics	Kindergarten/ lowest 25% (Cohort 2)	10/112	19.19 (4.80)	15.50 (5.54)	3.69	0.71	Statistically significant	+26
WLPB: Word Attack subtest	Phonics	Kindergarten/ lowest 25% (Cohort 2)	10/112	4.73 (3.68)	1.48 (2.17)	3.25	1.07	Statistically significant	+36
WLPB: Letter-Word ID subtest	Phonics	Grade 1/ lowest 25% (Cohort 3)	10/104	25.06 (6.85)	21.31 (4.75)	3.75	0.63	ns	+24
WLPB: Word Attack subtest	Phonics	Grade 1/ lowest 25% (Cohort 3)	10/104	7.85 (6.52)	4.02 (4.02)	3.83	0.70	Statistically significant	+26
Ross & Casey, 1998 (quasi-experimental design)[9]—Two years of intervention									
WRMT: Word ID subtest	Phonics	Kindergarten/ lowest 25%	7/79	27.10 (14.25)	25.10 (13.40)	2.00	0.15	ns	+6
WRMT: Word Attack subtest	Phonics	Kindergarten/ lowest 25%	7/79	10.11 (6.13)	7.80 (8.10)	2.31	0.30	ns	+12

(continued)

Appendix A4.4 Summary of subgroup findings for alphabetics domain[1] *(continued)*

Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation)[2]: *Success for All®* group	Authors' findings from the study: Mean outcome (standard deviation)[2]: Comparison group	WWC calculations: Mean difference[4] (*SFA®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Smith et al., 1993 (quasi-experimental design)[9]—One year of intervention									
WRMT: Letter ID subtest	Letter knowledge	Kindergarten/ lowest 25% (Cohort 1)	4/38	nr	nr	na	0.38[10]	ns	+15
WRMT: Word ID subtest	Phonics	Kindergarten/ lowest 25% (Cohort 1)	4/38	nr	nr	na	2.56[10]	Statistically significant	+49
WRMT: Letter ID subtest	Letter knowledge	Grade 1/ lowest 25% (Cohort 2)	4/38	nr	nr	na	–0.07[10]	ns	–3
WRMT: Word ID subtest	Phonics	Grade 1/ lowest 25% (Cohort 2)	4/38	28.16 (10.02)	18.53 (12.78)	9.63	0.82	ns	+29
WRMT: Word Attack subtest	Phonics	Grade 1/ lowest 25% (Cohort 2)	4/38	9.05 (5.37)	4.68 (5.76)	4.37	0.77	ns	+28

na = not applicable nr = not reported ns = not statistically significant

1. This appendix presents subgroup findings (students in the lowest 25% of their grades) for measures that fall in the alphabetics domain. Total group scores were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time the pretest is administered. For example, the kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. WWC combined means and standard deviations for the *SFA®* schools and for the control schools. Adjusted posttest means (with pretests standard scores as covariates) were used for effect size calculations. Kindergarten and grade 1 cohorts from Abbottston elementary school received four years of intervention.
9. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the cases of Ross and Casey (1998), Madden et al. (1993), and Smith et al. (1993), a correction for clustering was needed, so the significance levels may differ from those reported in the original studies.
10. Authors reported effect sizes that used comparison group standard deviation in the denominator (Glass's delta).

Appendix A4.5 Summary of subgroup findings for comprehension domain[1]

				Authors' findings from the study					
				Mean outcome (standard deviation)[2]		WWC calculations			
Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	*Success for All®* group	Comparison group	Mean difference[4] (*SFA®*– comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Ross & Casey, 1998 (quasi-experimental design)[8]—Two years of intervention									
WRMT: Passage Comprehension subtest	Reading comprehension	Kindergarten/ lowest 25%	7/79	12.29 (7.79)	11.20 (8.20)	1.09	0.13	ns	+5
Smith et al., 1993 (quasi-experimental design)[8]—One year of intervention									
Peabody Picture Vocabulary Test	Vocabulary development	Kindergarten/ lowest 25% (Cohort 1)	4/38	nr	nr	na	0.26[9]	ns	+10
WRMT: Passage Comprehension subtest	Reading comprehension	Grade 1/ lowest 25% (Cohort 2)	4/38	9.84 (6.18)	8.11 (7.13)	1.73	0.25	ns	+10

na = not applicable nr = not reported ns = not statistically significant

1. This appendix presents subgroup findings (students in the lowest 25% of their grades) for measures that fall in the comprehension domain. Total group scores were used for rating purposes and are presented in Appendix A3.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time the pretest is administered. For example, the kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the cases of Ross and Casey (1998), Madden et al. (1993), and Smith et al. (1993), a correction for clustering was needed, so the significance levels may differ from those reported in the original studies.
9. Authors reported effect sizes that used the comparison group standard deviation in the denominator (Glass's delta). Effect size was computed by subtracting the comparison group mean from the intervention group mean and dividing the result by the comparison group standard deviation.

Appendix A4.6 Summary of subgroup findings for general reading achievement domain[1]

				Authors' findings from the study					
				Mean outcome (standard deviation)[2]		WWC calculations			
Outcome measure	Construct	Study sample[3]	Sample size (schools/ students)	*Success for All®* group	Comparison group	Mean difference[4] (*SFA®*–comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Madden et al., 1993 (quasi-experimental design)[8,9]—Three years of intervention									
Durrell Oral Reading subtest	General reading	Prekindergarten/ lowest 25% (Cohort 1)	10/126	4.35 (4.30)	1.81 (3.66)	2.54	0.63	ns	+24
Durrell Oral Reading subtest	General reading	Kindergarten/ lowest 25% (Cohort 2)	10/112	6.04 (4.62)	3.32 (3.37)	2.72	0.67	Statistically significant	+25
Durrell Oral Reading subtest	General reading	Grade 1/ lowest 25% (Cohort 3)	10/104	12.92 (6.39)	8.08 (4.87)	4.85	0.85	Statistically significant	+30
Ross & Casey, 1998 (quasi-experimental design)[9]—Two years of intervention									
Durrell Oral Reading subtest	General reading	Kindergarten/ lowest 25%	7/79	4.14 (3.84)	3.00 (3.60)	1.14	0.31	ns	+12

ns = not statistically significant

1. This appendix presents subgroup findings (students in the lowest 25% of their grades) for measures that fall in the general reading domain. Total group scores were used for rating purposes and are presented in Appendix A3.4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The cohort is defined by the time the pretest is administered. For example, the kindergarten cohort describes students who completed pretest measures in kindergarten.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. WWC combined means and standard deviations for the *SFA®* schools and for the control schools. Adjusted posttest means (with pretests standard scores as covariates) were used for effect size calculations. Kindergarten and grade 1 cohorts from Abbottston elementary school received four years of intervention.
9. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the cases of Madden et al. (1993) and Ross and Casey (1998), a correction for clustering was needed so the significance levels may differ from those reported in the original studies.

Appendix A5.1 *Success for All*® rating for the alphabetics domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of alphabetics, the WWC rated *Success for All*® as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Success for All*® was assigned the highest applicable rating.

Rating received
Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.
• Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.
Met. Two studies showed a statistically significant positive effect, and one of the studies had a strong design.
AND
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 *Success for All*® rating for the comprehension domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of comprehension, the WWC rated *Success for All*® as having mixed effects. It did not meet the criteria for positive effects because only one study showed statistically significant positive effects. In addition, it did not meet the criteria for potentially positive effects because more studies showed indeterminate effects than substantively important or statistically significant positive effects. The remaining ratings (no discernible effects, potentially negative effects, and negative effects) were not considered because *Success for All*® was assigned the highest applicable rating.

Rating received

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing 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.

  Not met. No studies showed a statistically significant or substantively important negative effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Met. One study showed a statistically significant positive effect, one study showed a substantively important positive effect, and four studies showed indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

  Not met. Only one study had a statistically significant positive effect in this domain.

AND

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

  Met. No studies showed statistically significant or substantively important negative effects in this domain.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. One study had a statistically significant positive effect, and one study had a substantively important positive effect in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. No studies showed statistically significant or substantively important negative effects in this domain, and more studies showed indeterminate effects (four) than statistically significant (one) or substantively important positive effects (one) in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.3 *Success for All*® rating for the general reading achievement domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of general reading achievement, the WWC rated *Success for All*® as having potentially positive effects. It did not meet the criteria for positive effects because only one study showed a statistically significant positive effect. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Success for All*® was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. Three studies showed substantively important positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies showed statistically significant or substantively important negative effects. Three studies showed indeterminate effects and three studies showed substantively important positive effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

  Not met. No studies showed a statistically significant positive effect.

AND

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

  Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	7	73	3,909	Medium to large
Fluency	0	0	0	na
Comprehension	6	65	2,565	Medium to large
General reading achievement	6	37	2,573	Medium to large

na = not applicable/not studied

1. A rating of "medium to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."