Abstract Title Page

Title:

Reading Outcomes of Success for All: Early Results from the MDRC Investing in Innovation Evaluation

Authors and Affiliations:

Robert E. Slavin
Nancy A. Madden
Johns Hopkins University
-and-
Success for All Foundation

Janet Quint
MDRC

Background/Context:

Success for All (SFA) is one of the best known and thoroughly evaluated school reform models. First implemented in 1987 and focused on ensuring that every child learns to read well in the elementary grades, it combines three basic elements:

- A set of strategies for securing teacher buy-in, providing school personnel with initial training and ongoing professional development, and fostering shared leadership in schools;
- An instructional model that emphasizes phonics for beginning readers and comprehension for students at all levels, and that is characterized by a highly specified curriculum, an emphasis on cooperative learning, frequent assessments, and tutoring for students who need extra help;
- Whole-school improvement components that address issues of behavior, attendance, parental involvement, and other aspects of the educational experience that can affect student learning.

Previous evaluations, both experimental and nonexperimental, showed that students in SFA classrooms outperformed students receiving other kinds of reading instruction on standardized reading tests.

The most salient of these evaluations was a three-year longitudinal cluster randomized experiment in which 35 Title I schools were randomly assigned to use Success for All either in grades K-2 or 3-5, with the 3-5 group serving as a control group for the K-2 schools. Children in the K-2 schools scored significantly higher than their counterparts in the 3-5 schools on three scales from the Woodcock Reading Mastery Test. Impacts grew over time as the children progressed from kindergarten to second grade. (See Borman, Slavin, Cheung, Chamberlain, Madden, & Chambers, 2007.) In other large-scale studies, results for students in SFA schools have outstripped those for students in matched comparison schools. (See, for example, Rowan, Correnti, Miller, & Camburn, 2009.)**.** The strength of the program's evidentiary base was critical to the selection in 2010 of the Success for All Foundation (the nonprofit organization that provides materials, training, and support to schools implementing the intervention) as one of only four recipients of five-year scale-up grants awarded under the U.S. Department of Education's first Investing in Innovation (i3) competition. The i3 grant called for SFAF to expand its operations substantially over the five-year period and for an independent evaluation of the implementation and impacts of that expansion to be conducted.

Purpose/Objective:
Further evaluation of the initiative is especially important for two reasons. First, the program model has continued to evolve over time, with a greater emphasis placed on the use of engaging technology in the classroom and on the deployment of school district personnel trained by SFAF to provide professional development services and technical assistance to schools along with SFAF coaches. Second, many school reading programs have also modified their practices since earlier SFA evaluations were conducted. In particular, they have strengthened their teaching of phonics and, like SFA, they have incorporated increasingly intensive instructional supports for students who are not making adequate progress in the classroom (an approach commonly

referred to as Response to Intervention). All these developments leave open the question of whether SFA continues to lead the early reading field.

MDRC, a nonprofit, nonpartisan education and social policy research organization, is conducting the evaluation of the i3 scale-up. The data reported here come from the first of a series of three reports from that evaluation. That report examines the implementation and effects on student learning of SFA during the 2011-2012 school year, the first year that the program was put in place. The evaluation employs an experimental design, in which 37 schools in five school districts participating in the scale-up effort were assigned at random to a program group and a control group. The analysis compares the experiences of adults and the performance of students in the 19 program-group schools that were randomly selected to receive SFA with those of their counterparts in the 18 control-group schools in the same five districts that did not receive the intervention.

This report uses quantitative and qualitative data from a wide variety of sources. Through teacher and student surveys, implementation summaries completed by SFAF staff, teacher instructional logs, interviews and focus groups conducted in the course of site visits with school personnel, school district data bases, and individual assessments of students' reading skills, it addresses the following main questions:

1. Are the SFA and control schools participating in the i3 evaluation similar to each other and to the other schools receiving SFA under the i3 grant that are not part of the evaluation?
2. What was involved in putting the program in place, and how did school personnel respond?
3. To what extent were SFA's features implemented during the program's first year, and what factors were associated with more complete implementation?
4. How distinct were the treatment and control schools in various aspects of school functioning?
5. Did SFA produce impacts on students' early reading skills?

Setting:

The impact analysis reported here centers on a group of students who entered kindergarten in the 37 schools in the fall of 2011 and whose reading skills were assessed the following spring.

Participants:

The main analysis sample consisted of 2,956 kindergartners who took both pre- and posttests. An attrition rate of 10% was equal in experimental and control groups.

Intervention/Program:

Table 1 lays out the key features associated with each of the three main program elements: implementation strategies, the instructional model, and the non-instructional components. The key features include both structures (for example, a 90-minute reading block, a group of staff members whose mission is to improve relationships with students' families) and processes (for example, use of data for fine-tuning instruction and for regrouping students, ongoing professional development).

Figure 1 depicts the theory of change guiding the program and the evaluation. (In the interest of simplicity and clarity, the diagram has been stripped down to its essentials more or less as they unfold in chronological order.) Strategies for implementing the program are introduced to participating schools, which then put into place SFA's instructional model and its whole-school improvement components. The operationalization of these program elements leads to intermediate changes in teaching and learning and in the school environment more generally. The SFA instructional model results in improved classroom instruction for all students, in greater individualization of instruction, and to teachers' greater confidence in their ability to help all their students achieve success. The whole-school improvement components benefit individual students (for example, by securing eyeglasses for students who need them, or by enlisting parents in support of their children's learning) as well as the broader school environment (for example, by creating a more orderly environment and by engaging teachers in more collaborative efforts). Finally, these intermediate changes produce changes in student outcomes: greater engagement and improved attendance and behavior, higher levels of achievement, and steady academic progress through the elementary grades.

Research Design and Analysis:

The study uses an experimental design with random assignment of a roughly equal number of schools either to a program group, which puts in place the SFA program, or a control group, which implements the reading programs in regular use by their schools. The difference in outcomes between the program schools and the control schools can be interpreted as the average effect of the SFA program relative to "business as usual" across all participating districts.

Recruitment for the evaluation occurred as part of the general outreach to schools, districts, and states for the i3 scale-up grant. Schools willing to participate in the evaluation received the program gratis. Each school had to meet certain eligibility criteria: 1) it had to serve students from kindergarten through 5th grade; 2) at least 40 percent of the students at the school had to be eligible for the free and reduced price lunch program; 3) the school had to be willing to participate in a random assignment experiment; 4) it had to identify a school staff member to serve as the SFA facilitator; and 5) at least 75 percent of its teachers had to vote to adopt the SFA program.

Collection and Analysis of Outcome Data:

Measures. Students were pre- and posttested on the Woodcock-Johnson Word Attack (WA) and Letter-Word (LW) scales by independent testers unaware of students' treatment placements.

Analysis. Data were analyzed using HLM, with students' pretest scores serving as covariates for their respective posttests. Students were nested within schools in a two-level hierarchical design. Districts were treated as fixed effects.

Findings/Results:

The key impact findings are summarized in Table 2:

- By the end of the first implementation year, the SFA program produced a positive and statistically significant impact for kindergarten students on one of the two reading

outcomes measured. The program impact on the Woodcock- Johnson Word Attack (WA) score is 0.55 raw score points, or 0.18 standard deviations in effect size.

- A similar impact on the WA test score was found for the spring analysis sample, which includes all kindergarten students with at least one valid spring test score, including those who moved into a study school over the course of the year.
- The program impact on WA score seems to be robust across a range of demographic and socioeconomic subgroups. Positive and statistically significant impacts were found for male students, students in poverty (as defined by each district), and students who were not English Language Learners.
- The program impact on WA does not vary by students' baseline reading level.
- The SFA program as implemented in the first year did not produce any meaningful impact (positive or negative) on Woodcock-Johnson Letter-Word Identification (LWI) test scores.

Conclusions:

Success for All (SFA) is a complex and far-reaching intervention that includes both instructional and whole-school reforms. It requires that schools put in place new structures and processes that are time-consuming and labor intensive. It requires even experienced elementary school reading teachers to adopt new practices that may appear unfamiliar to them and whose effectiveness may seem questionable (at least until they have seen the outcomes of these practices). Nonetheless, by the end of the first year, almost all the program schools had put in place three-quarters or more of the elements of SFA that the Success for All Foundation (SFAF) considers to have the highest priority, along with many lower-priority elements. And many teachers were beginning to feel more comfortable with the program and looking forward to a smoother second year of operations.

Instructional logs and principal and teacher surveys point to key ways in which reading instruction in SFA schools and control schools differed. Reading lessons in SFA schools were more likely than those in control schools to emphasize comprehension. Teachers in SFA schools were also more likely to emphasize aspects of instruction that are hallmarks of the SFA program: grouping and regrouping of students for reading from first grade on by ability level and across grades, and, from kindergarten on, regular use of cooperative learning techniques and close adherence to a highly scripted curriculum.

The first-year impact analysis of the study centers on students who entered kindergarten in the program and control schools in fall 2011; it assesses their reading skills in the spring of 2012. These findings are encouraging: They show that kindergartners in Success for All schools scored significantly higher than their control counterparts on one of two standardized measures of early literacy. The impact is robust: It holds up across several subgroups and remains significant at the 0.10 level when a procedure that corrects for multiple hypothesis testing is applied. The effect size (0.19) is on a par with that achieved by a number of other prominent school reform initiatives. If the findings hold up in later years, they will suggest that even as it continues to scale up to hundreds of additional schools, Success for All is maintaining its effectiveness. This would add to the policy importance of the model, in showing that a model capable of working at large scale could, for example, broadly improve outcomes in Title I schools.

Appendices

Appendix A. References

Borman, G.D., Slavin, R.E., Cheung, A., Chamberlain, A., 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), 701-731.

Rowan, B., Correnti, R., & Miller, R. (2002). *What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools.* Philadelphia, Pennsylvania: Consortium for Policy Research in Education, University of Pennsylvania..

Appendix B. Tables and Figures

Table 1: Key Elements of the Success for All Program

The instructional model

- A K-6 reading program with three levels:
    - KinderCorner (kindergarten)
    - Reading Roots (usually first grade – beginning readers)
    - Reading Wings (usually second grade and up)
- An emphasis on phonics in the lower levels and on vocabulary and comprehension at all levels
- A 90-minute reading period
- "Scripted" lesson plans that lay out timed activities and language for teachers to use in presenting them
- Instruction that is rapidly paced, uses technology, and employs cooperative learning in pairs and small groups
- Cross-grade ability grouping for reading, with students frequently leaving their regular classroom to receive reading instruction from another teacher ("walking to read")
- Frequent use of data to monitor student learning
- Quarterly assessments to measure students' progress toward grade-level standards and to regroup students into the highest levels at which they can be successful ("aggressive placement")
- A team of staff members charged with fostering instructional improvement efforts
- Computerized small-group tutoring and individual tutoring for students who need extra assistance

Whole-school improvement features

- A "Leading for Success" continuous improvement model whose key elements include distributed leadership, quarterly review of student achievement data, and the harnessing of school resources to meet specified achievement goals
- A Leadership team (including the principal, SFA facilitator, and Schoolwide Solutions coordinator, among others) that provides vision, direction, and monitoring
- Leading for Success teams that include:
    - Instructional component teams of teachers for each level (KinderCorner, Roots, Wings)
    - "Solutions" teams of teachers and other staff members charged with:
        - Improving student attendance
        - Developing appropriate interventions (academic, behavioral, health-related, social, and attendance-related) for particular students with learning difficulties
        - Putting in place "Getting Along Together," a schoolwide program for social skills development and conflict resolution, as well as other behavioral interventions
        - Increasing family involvement
        - Engaging community businesses and institutions to support the school

Implementation strategies

- An adoption process that includes a presentation on the program followed by a teacher referendum
- Designation of a school staff member as the program facilitator
- Initial training of school leaders, program facilitators, and teachers
- Delivery of SFA curricular and other materials
- Ongoing professional development supplied by "coaches" (SFAF employees or district employees trained by SFAF) and by school-based program facilitators

Table 2
Key Impact Findings

Early Impact of SFA on Kindergarten Student Reading Achievement for the Main Analysis Sample (Implementation Year 2011-2012)

Outcome	Program Group	Control Group	Estimated Impact	Estimated Impact Effect Size	P-Value
Stable student subgroups					
Woodcock-Johnson Letter-Word Identification	20.02	20.02	-0.01	0.00	0.991
Woodcock-Johnson Word Attack	5.87	5.32	0.55	0.18	0.032 **
Number of schools: 37	19	18			

SOURCES: Woodcock-Johnson Letter-Word Identification test (spring 2012), Woodcock-Johnson Word Attack test (spring 2012), student records data collected from the five districts in the study sample.

NOTES: The "main analysis sample" consists of students from 37 schools (19 program group schools and 18 control group schools), and includes any student who had at least one valid spring test score and who was enrolled in a study school during the fall of the same year.

The student sample size for the Woodcock-Johnson Letter-Word Identification test is 2,564 students (1,331 in the program group and 1,233 in the control group).

The student sample size for the Woodcock-Johnson Word Attack test is 2,564 students (1,334 in the program group and 1,230 in the control group).

The impact analyses for student reading achievement were conducted using raw scores. The estimated impacts are based on a two-level model with students nested within schools, controlling for random assignment block and school- and student-level covariates. The program group and control group columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the program group as the basis for the adjustment.

Effect sizes were computed using the full control group's standard deviations for the respective measures. The control group standard deviations are as follows:
WJLWI: 6.98
WJWA: 3.07

A two-tailed t-test was applied to differences between program and control groups. Statistical significance levels are indicated as follows: *** = 1 percent; ** = 5 percent; * = 10 percent.

Rounding may cause slight discrepancies in calculating sums and differences.

Figure 1: Success for All Theory of Change

Implementation strategies	Program operations	Intermediate pathways	Student outcomes
• Program adoption • Initial training • Materials delivery • Ongoing professional development	Instructional model Whole-school improvement components	**Improved classroom instruction** **Greater individualization of instruction** **Teachers' greater confidence in ability to reach difficult students and greater belief that such students can succeed** **More attention to health/vision issues** **Greater parental engagement to support students** **More orderly environment conducive to learning** **Greater cooperation among teachers** **Greater focus on success**	• Improved attendance • Improved behavior • Improved engagement • Improved achievement in reading • Lower Special Education assignment • Lower retention in grade