THE SUCCESS FOR ALL MODEL OF SCHOOL REFORM

Early Findings from the Investing in Innovation (i3) Scale-Up

Executive Summary

Janet C. Quint
Rekha Balu
Micah DeLaurentis
Shelley Rappaport
Thomas J. Smith
Pei Zhu

October 2013

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

Executive Summary

The Success for All Model of School Reform

Early Findings from the Investing in Innovation (i3) Scale-Up

Janet C. Quint
Rekha Balu
Micah DeLaurentis
Shelley Rappaport
Thomas J. Smith
Pei Zhu

with

Emma Alterman
Herbert Collado
Emily Pramik

October 2013

Funding for this report came from the U.S. Department of Education under its Investing in Innovation (i3) initiative. The i3 grant called for the Success for All Foundation to scale up its elementary school whole-school reform initiative and for MDRC to conduct an independent evaluation of the implementation and impacts of that expansion.

Dissemination of MDRC publications is supported by the following funders that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, The George Gund Foundation, Sandler Foundation, and The Starr Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

The findings and conclusions in this report do not necessarily represent the official positions or policies of the funders.

For information about MDRC and copies of our publications, see our Web site: www.mdrc.org.

Overview

First implemented in 1987, the Success for All (SFA) school reform model combines three basic elements:

- Reading instruction that is characterized by an emphasis on phonics for beginning readers and comprehension for students at all levels, a highly structured curriculum, an emphasis on cooperative learning, across-grade ability grouping and periodic regrouping, frequent assessments, and tutoring for students who need extra help
- Whole-school improvement components that address noninstructional issues
- Strategies to secure teacher buy-in, provide school personnel with initial and ongoing training, and foster shared school leadership

Success for All was selected to receive a five-year scale-up grant under the U.S. Department of Education's first Investing in Innovation (i3) competition. This report, the first of three, examines the program's implementation and impacts in 2011-2012, the first year of operation, at 37 kindergarten through grades 5 and 6 (K-5 and K-6) schools in five school districts that agreed to be part of the scale-up evaluation: 19 "program group" schools were randomly selected to operate SFA, and 18 "control group" schools did not receive the intervention. Program and control group schools were very similar at the start of the study. The analysis compares the experiences of school staff as well as the reading performance of a cohort of kindergarten students who remained in SFA schools throughout the year (and therefore received the maximum "dosage" of the program) with those of their counterparts in the control group schools.

Key Findings

- While teachers in the SFA schools initially expressed concerns about implementing this new, complex, and demanding initiative, by the end of the first year, many teachers were beginning to feel more comfortable with the program.
- Almost all the program group schools had reached a satisfactory level of early implementation as determined by the Success for All Foundation, the nonprofit organization that provides materials, training, and support to schools operating the reform. Yet there was also ample room for schools to implement additional program elements and to refine the elements that they had put in place.
- Reading instruction in the two sets of schools was found to differ in key ways.
- Kindergartners in the SFA schools scored significantly higher than their control group counterparts on one of two standardized measures of early reading. The impact on this measure seems to be robust across a range of demographic and socioeconomic subgroups, as well as across students with different levels of literacy skills at baseline.

Subsequent reports will examine the reading skills of these students as they progress through first and second grades and will also measure the reading skills of students in the upper elementary grades.

Preface

Improving reading instruction and student reading achievement has been a major focus of federal, state, and local education programs in recent decades because reading is a foundational skill for all academic success. Though progress has been made in understanding the characteristics of effective reading instruction, student reading achievement has improved only gradually over the past 20 years, and black and Hispanic students continue to lag behind their white counterparts.

A continuing issue has been how to secure wider take-up by schools and school districts of instructional approaches that have been shown to be effective. One such approach is Success for All (SFA). Starting in the 1980s, SFA's developers, Robert Slavin and Nancy Madden, developed a reading program — including curriculum materials and teacher professional development — that emphasizes both phonics and comprehension and that makes extensive use of cooperative learning techniques. In the following years, they and their colleagues built a substantial research record demonstrating SFA's positive effects on students' reading achievement.

The U.S Department of Education's Investing in Innovation (i3) program has created an opportunity to expand programs that have previously been shown to be effective and to test their continued effectiveness at scale and in new settings. SFA's solid evidentiary base positioned it to win one of the first scale-up grants awarded under the i3 competition. This report is the first of three that examine the implementation and impacts of the i3 scale-up of SFA.

This fresh look at SFA will address important new questions about this much-studied initiative. As schools' approaches to reading instruction continue to evolve, does the SFA model remain substantially different from other reading programs used in the participating districts, or have the strategies used by those schools begun to look more like SFA? How do teachers and principals respond to SFA in a world of high-stakes testing and school accountability? Can schools implement SFA with the needed fidelity? And, finally, does SFA continue to produce better reading outcomes for students than other programs? This report provides early, encouraging answers based on the first year of program implementation. Later reports on this project will continue the analysis through two more years of program implementation and will follow children into later grades.

Gordon L. Berlin
President

Acknowledgments

This report reflects the efforts of a great many people. First and foremost, we want to acknowledge the principals, research liaisons, and teachers at the schools that participated in this study of the Success for All program. Their assistance and cooperation were vital for providing the rich and detailed information on which this report is based. Staff members in the central offices of the five school districts where the study took place provided us with critical student records data.

At the Success for All Foundation, Nancy Madden, Dan Anderson, Sharon Fox, and Jill Hanson were unfailingly responsive to our information requests. Nancy Madden and Robert Slavin provided helpful critiques of earlier drafts.

Joseph Jenkins, Professor Emeritus at the University of Washington, helped us assess the advantages and drawbacks of several assessment measures. Brian Rowan and Robin Jacob at the University of Michigan provided useful insights on the study's instructional logs and their analysis.

Pamela Wells and her capable team at Decision Information Resources, Inc., produced the student assessment data that are at the heart of the impact study. MDRC staff members Jo Anna Hunter and Matthew Au worked with DIR to ensure that these data were collected on schedule and with minimal disruption to school operations.

Also within MDRC, Seth Muzzy helped prepare the principal and teacher surveys and the instructional logs that constitute important data sources for the report. Shirley James, Zuleka Abakoyas, Donna George, and Carmen Troche keyed surveys and logs. Marjorie Singer, Mariella Quimoyoz-Cruz, Ann Fenton, and Amor Aquino helped resolve a number of issues that arose.

Fred Doolittle ensured that the team received both material and moral support at every juncture. He, along with Margaret Bald, Jean Grossman, Robert Ivry, and Leigh Parise, carefully reviewed earlier drafts of the report and made comments that improved the final product. Mario Flecha was his usual unflappable self in handling a variety of administrative and other tasks. Elizabeth Nelson assisted with the coding of the qualitative data.

Robert Weber edited the report with his customary mix of sensitivity and punctiliousness, and Stephanie Cowell prepared the report for publication.

The Authors

Executive Summary

First implemented in 1987, Success for All (SFA) is one of the best-known and most thoroughly evaluated school reform models. It combines three basic elements:

- Reading instruction that emphasizes phonics for beginning readers and comprehension for students at all levels, and that is characterized by a highly structured curriculum, an emphasis on cooperative learning, across-grade ability grouping and periodic regrouping, frequent assessments, and tutoring for students who need extra help
- Whole-school improvement components that address noninstructional issues that can affect student learning, such as behavior, attendance, and parental involvement
- 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

Previous evaluations of Success for All showed positive effects on students' reading performance. The strength of the program's evidentiary base was critical to the selection in 2010 of the Success for All Foundation (SFAF) as one of four recipients of five-year scale-up grants awarded under the U.S. Department of Education's first Investing in Innovation (i3) competition. SFAF is the nonprofit organization that provides materials, training, and support to schools implementing the intervention. The i3 grant called for SFAF to expand its operations to 1,100 additional schools over the five-year period and for MDRC to conduct an independent evaluation of the implementation and impacts of that expansion.

Further study of the initiative is especially important for two reasons. First, over the decade since SFA was rigorously studied, the program model has continued to evolve, with greater emphasis being placed on the use of engaging technology in the classroom. Second, many school reading programs have also modified their practices, strengthening their teaching of phonics and incorporating additional instructional supports for students who are not making adequate progress in the classroom. These developments leave open the question of whether SFA continues to lead the early reading field.

This report, the first of three, examines the program's implementation and impacts in 2011-2012, the first year of operation. Thirty-seven kindergarten through grades 5 and 6 (K-5 and K-6) schools in five school districts agreed to be part of the scale-up evaluation; 19 "program group" schools were randomly selected to operate SFA, and 18 "control group" schools

did not receive the intervention. The analysis compares the experiences of school staff as well as the reading performance of a cohort of kindergarten students in the SFA schools with those of their counterparts in the control group schools. Subsequent reports will examine the reading skills of these students as they progress through first and second grades and will also measure the reading skills of students in the upper elementary grades.

Data sources for the report include principal surveys, teacher surveys, and teacher-completed logs describing reading instruction, which were administered at all schools; "School Achievement Snapshot" forms completed by SFAF coaches to report on the extent of program implementation; assessments administered to kindergartners at the beginning and end of the school year; and administrative records obtained from the districts. During spring 2012 site visits, researchers also conducted interviews with principals at SFA schools and control group schools as well as interviews with SFA facilitators and focus groups with teachers at SFA schools.

In brief, this report finds that while teachers initially expressed concerns about implementing this new, complex, and demanding initiative, by the end of the first year, almost all the program schools had reached what SFAF considers a satisfactory level of early implementation, and many teachers were beginning to feel more comfortable with the program. Reading instruction in SFA schools was found to differ in key ways from instruction in the control group schools. Finally, kindergartners in the SFA schools scored significantly higher than their control group counterparts on one of two standardized measures of early reading.

Characteristics of Participating Schools and Students

- **The 19 program group schools were similar to the 18 control group schools on all school-level characteristics at baseline, although the 37 evaluation schools were not fully representative of all schools participating in SFA's i3 scale-up.**

As expected, random assignment produced two groups of schools that, at the outset of the demonstration, had similar characteristics. The evaluation schools tended to be larger than other schools participating in the SFA scale-up and to serve more Hispanic students — not surprising, given the location of the majority of the evaluation schools in districts within 200 miles of the U.S. border with Mexico.

- **While on most student-level characteristics students in the program and control group schools were statistically indistinguishable, students in SFA schools were significantly more likely than those in control group schools to be English language learners, and control group students had**

slightly higher scores, on average, on one of two measures of early reading skills.

Randomization in the research design ensures that any baseline differences between the program group and the control group are themselves random — that is, due to chance. Nonetheless, these chance differences can be statistically significant when the samples are very large. The impact analysis controls statistically for these baseline differences.

- **Mobility was fairly high in the study schools. About 10 percent of kindergarten students who were enrolled in fall 2011 (the baseline point for the study) left the study schools over the course of the school year, and about 11 percent of students who were enrolled in these schools in spring 2012 had transferred during the year from a school that was not in the study.**

Mobility patterns were similar in program and control group schools. The main analysis sample for which impacts are measured consists of students who did not move either in or out of the study schools. It comprises 2,567 students who were enrolled in fall 2011 in regular (not special education) classes and who could be tested in English, remained in these schools in the spring, and had valid scores on at least one of the two standardized reading measures used in the evaluation. This group of students had the best chance of receiving the full amount of the SFA program during the year.

Implementing the Initiative

- **The adoption, summer workshops, and professional development processes prescribed by SFAF were generally followed, although some teachers voiced concerns about each of these sets of activities.**

At all the study schools, at least 75 percent of the teachers voted to adopt SFA, although, in retrospect, some teachers reported that they were given limited information beforehand and that the decision was rushed. In a similar vein, teachers reported that workshops held before the school year began and professional development from SFA coaches did not fully prepare them for the day-to-day experience of teaching in an SFA classroom.

- **Teachers and facilitators at the SFA schools frequently reported that insufficient staff made it difficult for the schools to do everything that they were expected to do. Teachers also voiced concerns about implementing some aspects of the program model.**

Although 15 of the 19 schools had SFA facilitators who were supposed to be available full time, many of these individuals had to divide their efforts between the program and other

non-SFA responsibilities. Many schools lacked the staff needed to put in place the daily tutoring for struggling students that is called for by the program model. A mismatch between the number of students identified for instruction at a certain level and the number of teachers prepared to teach at that level sometimes complicated the regrouping process, making for too many students at some levels and too few at others (so that students at somewhat different levels had to be grouped together). And staffing challenges also meant that about half the schools did not put in place the committees charged with implementing the whole-school aspects of SFA.

Features of the program model also posed implementation difficulties. Some teachers complained that the highly structured curriculum stifled their creativity. They also feared that classes were moving too quickly for struggling students, and some distrusted SFAF's reassurances that students who did not grasp the material the first time around would have opportunities to relearn it at a later point. Finally, school staff complained that SFA's data system was complicated and demanding.

- **Such issues notwithstanding, by the end of the first year, all but one of the study schools were deemed to have met SFAF's standards for adequate first-year implementation, although there was also considerable room for improving the breadth and depth of that implementation.**

On average, the 19 study schools were judged to have put in place 85 percent of the items on the school Snapshot that describe program features that SFAF considers to have the highest priority for first-year implementation. Similarly, they were judged to have put in place 79 percent of all the features whose implementation was measured during the first year.

However, the standard for what constitutes faithful implementation of the SFA model changes as the program rolls out. Only two-thirds of all the Snapshot items were rated during the first year, mostly because SFAF does not expect schools to implement the remaining items until the second year of program operations. The Snapshot ratings indicate that all schools could improve their implementation of SFA, not only by putting in place additional program features but also by improving the depth and quality of features already in place.

- **The SFA schools varied a good deal in their implementation ratings, with higher ratings generally being found at the schools that had more experienced teachers.**

In collaboration with SFAF, the researchers used the Snapshot ratings to create a scoring system — a more refined measure of implementation that accounts for the extent of implementation by weighting key practices and taking into account the proportion of classrooms demonstrating a given practice. On average, schools earned just over half the maximum possible score for the items rated in 2011-2012 (55.8 of a maximum possible score of 105). The

lowest-scoring school achieved just 40 percent of the maximum possible score, while the highest-scoring school achieved 74 percent of the maximum score. The higher-scoring schools not only pursued more practices but also implemented those practices in more classrooms. The schools in the top quartile of the scale measuring fidelity of implementation had teachers with two years' more experience, on average, than schools in the bottom three quartiles.

- **Teachers and principals agreed that SFA benefited their schools.**

Despite the issues that they faced in implementing the program, in response to a survey item, 71.4 percent of the teachers expressed agreement with the statement "Overall, your school has benefited from the SFA program." Principals were unanimous in agreeing with this statement. Moreover, by the end of the first year, many teachers reported in focus groups that they felt more comfortable with the program.

Instructional and Other Characteristics of SFA Schools and Control Group Schools

- **The reading programs used in the control group schools appear to cover similar content as SFA, and all programs provide similar kinds of materials.**

The majority of the control group schools taught reading/language arts using commonly used basal programs available from leading educational publishers. Like SFA, these cover phonics, phonemic awareness, vocabulary, fluency, and reading comprehension. Like SFA, too, these materials include a teacher's manual, reading selections for students, assessments, and strategies for assisting struggling readers.

- **Teachers in the SFA schools received more professional development in reading and rated it more highly than did teachers in control group schools.**

Teachers in the SFA schools were more likely to report having received professional development in reading instruction, and on a greater number of reading-related topics, than their counterparts in the control group schools. These patterns held whether SFA teachers were compared with teachers in all control group schools or only with teachers at those control group schools that also adopted new reading programs. In general, SFA teachers also rated the professional development that they received as more helpful than did control group teachers.

- **Several factors appear to have differentiated reading instruction in the SFA and control group schools.**

Cooperative learning and cross-grade ability grouping and periodic regrouping are key features of the SFA instructional model, and teacher survey responses make it clear that these practices were much more common in program group schools than in control group schools. Furthermore, instructional logs completed by the teachers indicate that early-grade reading instruction in SFA schools was much more likely to center on comprehension and vocabulary; teachers in control group schools, in contrast, were more likely to emphasize spelling. Finally, teachers in SFA schools were much less likely than those in control group schools to say that they changed parts of the reading program that they disliked or with which they disagreed.

- **There are no significant differences between program and control group schools along other dimensions of reading instruction that SFA deems important.**

In both program and control group schools, the average length of the reading period was approximately an hour and forty minutes. Teachers in both sets of schools gave their principals virtually identical ratings on a scale measuring instructional leadership in reading (giving the principals ratings that were somewhat higher than a neutral midpoint). The use of data to check on students' progress was equally common in the two groups of schools. Finally, while a higher proportion of program group school principals than control group school principals reported that staff members provided tutoring to students needing extra assistance, the difference was not statistically significant. (As previously noted, staffing issues made it difficult to implement the tutoring component at a number of SFA schools.)

- **There are no significant differences between SFA and control group schools in the extent to which the schools had staff members charged with improving attendance and behavior, securing parental and community support, or undertaking other whole-school improvement activities not directly related to instruction.**

While principal surveys indicate that SFA schools were generally more likely than control group schools to have an individual or a group of people who were responsible for carrying out various activities associated with non-instructional whole-school reforms, the differences were not statistically significant. Again, this may be because the committees charged with implementing the whole-school improvement aspects of SFA were not fully operational at a number of schools.

Early Program Impacts

- **By the end of the first implementation year, SFA produced a positive and statistically significant impact on one of the two reading outcomes**

measured for the main sample of kindergarten students who remained in their study schools for the whole school year and who had maximum possible exposure to the program.

Both the Woodcock-Johnson Letter-Word Identification test and the Woodcock-Johnson Word Attack test measure phoneme awareness and decoding. The Letter-Word Identification test asks the student to read real words of increasing complexity, while the Word Attack test has the student apply phonic/decoding skills to nonsense words. The program impact on the Woodcock-Johnson Word Attack score is 0.55 raw score point, or 0.18 standard deviation in effect size. (Program and control group students had similar scores on the second measure, the Woodcock-Johnson Letter-Word Identification test.)

- **The program impact on the Word Attack score seems to be robust across a range of demographic and socioeconomic subgroups as well as across students with different levels of literacy skills at baseline.**

Positive and statistically significant impacts were found for male students, female students, black students, and Hispanic students, students in poverty (as defined by each district), non-English language learners, and students not in special education. (The program's impact on English language learners is positive but not statistically significant; the sample size for this subgroup is quite small.) The program's impact also does not differ for students with more or fewer literacy skills measured at baseline.

- **The positive findings on the Word Attack measure are consistent with findings in a previous study of SFA and are on a par with the impacts of other prominent school reform measures.**

Borman et al.'s study of SFA found a significant positive effect on the Word Attack measure for kindergarten students after one year of program implementation (and no impact on the Letter-Word Identification measure). Furthermore, the effect size registered in the present study is similar in magnitude to those of other major reforms.[1]

The results of this study are encouraging. They are also preliminary, for a number of reasons. First, students were tested after only one year of exposure to the SFA program. Second,

[1]Borman et al. used a meta-analysis to show that the effect size of 29 of the most widely deployed comprehensive school reforms ranged between 0.09 and 0.15 standard deviation (Borman, Hewes, Overman, and Brown, "Comprehensive School Reform and Achievement: A Meta-Analysis," *Review of Educational Research* 73: 125-230 [2003]). Similarly, the Tennessee Student-Teacher Ratio (STAR) experiment found that reducing early-grade classes in elementary schools from their standard size of 22 to 26 students to only 13 to 17 students significantly increased average student reading performance by 0.11 to 0.22 standard deviation in effect size (Nye, Hedges, and Konstantopoulos, "The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment," *Educational Evaluation and Policy Analysis* 21: 127-142 [1999]).

the measures used for kindergartners are less precise than those for older students, and they test phonetic skills; what ultimately matters for reading is comprehension, and this will not be assessed until students are slightly older. Third, teachers are likely to be able to implement SFA in their classrooms more easily and more smoothly in subsequent years than in this first year. Finally, it is anticipated that a number of program elements not now in place in the SFA schools — especially tutoring — will be added over time.

About MDRC

MDRC is a nonprofit, nonpartisan social and education policy research organization dedicated to learning what works to improve the well-being of low-income people. Through its research and the active communication of its findings, MDRC seeks to enhance the effectiveness of social and education policies and programs.

Founded in 1974 and located in New York City and Oakland, California, MDRC is best known for mounting rigorous, large-scale, real-world tests of new and existing policies and programs. Its projects are a mix of demonstrations (field tests of promising new program approaches) and evaluations of ongoing government and community initiatives. MDRC's staff bring an unusual combination of research and organizational experience to their work, providing expertise on the latest in qualitative and quantitative methods and on program design, development, implementation, and management. MDRC seeks to learn not just whether a program is effective but also how and why the program's effects occur. In addition, it tries to place each project's findings in the broader context of related research — in order to build knowledge about what works across the social and education policy fields. MDRC's findings, lessons, and best practices are proactively shared with a broad audience in the policy and practitioner community as well as with the general public and the media.

Over the years, MDRC has brought its unique approach to an ever-growing range of policy areas and target populations. Once known primarily for evaluations of state welfare-to-work programs, today MDRC is also studying public school reforms, employment programs for ex-offenders and people with disabilities, and programs to help low-income students succeed in college. MDRC's projects are organized into five areas:

- Promoting Family Well-Being and Children's Development
- Improving Public Education
- Raising Academic Achievement and Persistence in College
- Supporting Low-Wage Workers and Communities
- Overcoming Barriers to Employment

Working in almost every state, all of the nation's largest cities, and Canada and the United Kingdom, MDRC conducts its projects in partnership with national, state, and local governments, public school systems, community organizations, and numerous private philanthropies.