

Challenges in the Implementation of Success for All in Four High-Need Urban Schools

Janette Klingner

University of Colorado at Boulder

Elizabeth Cramer

Florida International University

Beth Harry

University of Miami

Abstract

We examined the challenges faced by 4 high-need urban schools when trying to implement Success for All (SFA). We wanted to understand SFA instruction as well as how SFA fit into the larger school context. Over a span of almost 2 years, we observed 45 SFA lessons (21 complete and 24 partial) across the 4 schools, taught by 30 teachers. We analyzed our data using a constant-comparison procedure and found that several challenges affected the fidelity with which the model was implemented. One of the greatest challenges was students stagnating and not passing to higher levels of the program, thus recycling through material they had already covered. We also found grouping and scheduling difficulties, and students of different reading levels and widely varying grade levels placed in the same class. We discuss the implications of these and other challenges.

The achievement gap between African American and Hispanic students and their white peers increased in the 1990s, despite numerous school reform efforts (Lee, 2002; Pogrow, 2002). Title I was reauthorized in 1994 with an emphasis on using school-wide funds to improve educational outcomes for students in high-need schools. In a companion study to the national evaluation of Title I (Stringfield, Millsap, & Herman, 1997), the two school-wide programs found to be most promising for educational outcomes were James Comer's School Development Program and Success for All (SFA). These results were replicated by a later study showing that each of these programs was associated with improved reading comprehension, but that neither program had a statistically significant advantage over the other (Clarke, 2001). SFA was one of three school reform programs found to be associated strongly with

improvement in student achievement by the American Institutes for Research (Herman et al., 2000).

Success for All in Urban Schools

Success for All is a school-based achievement-oriented program for students considered at risk in grades pre-K through 5. The program was developed in 1987, and during the 2000–2001 school year served approximately 1 million children in 1,800 schools (Slavin & Madden, 2001b). The program is built on the assumption that every child can learn to read. Its primary goal is to ensure that virtually every child will have reading skills at grade level or above by the end of third grade (Wells, Blendinger, & Greene, 2000).

The key components of the program include a research-based school-wide curriculum, assessments every 8 weeks, ability grouping, one-to-one tutoring for primary-grade students who struggle with reading, a family support team, and a program facilitator (Slavin & Madden, 2001b). Participating students are grouped according to their instructional reading level and by age, and the curriculum is designed to vary depending on the age of the students. Because students are grouped in this way, they do not necessarily have their homeroom teacher as their SFA instructor. During SFA time, students move to their reading classes for a 90-minute block of time, and no outside interruptions are allowed. Every certified teacher in the school participates, so SFA classes are generally smaller than regular classes. Lessons are scripted and include explicit phonics instruction as well as a focus on meaning.

An extensive review of research on the efficacy of SFA showed it to be highly effective in the majority of school districts in which it was implemented, yielding increases in reading achievement in comparison with other reading programs (Slavin, Madden, Dolan, & Wasik, 1996). Evaluations focused on assessment of student outcomes on standardized achievement mea-

asures (e.g., Madden, Slavin, Christie, & Karpouzis, 2001). Since Slavin et al.'s (1996) review, several studies have demonstrated significant educational improvements for students in culturally diverse urban schools implementing SFA programs (e.g., Fashola & Slavin, 1998; Slavin & Madden, 1999, 2001a). The achievement gap between white students and African American and Latino students in Texas was significantly diminished for schools using SFA school-wide models, leading Slavin and Madden (2001a) to conclude that this program was particularly effective for African American and Latino students. In a review of the research relating SFA to outcomes for special education students, Slavin and Madden (2001b, p. 33) concluded that SFA "both reduces the need for special education services (by raising the reading achievement of very low achievers) and reduces special education referrals and placements." These findings were corroborated by Borman and Hewes (2001).

However, claims of SFA's effectiveness have not gone without dispute. Studies of the effects of a 3-year implementation of SFA in Miami-Dade County, Florida, showed that for the lowest-performing schools, achievement was not improved (Levitt, 2000), and English-language learners in SFA schools made smaller gains in English proficiency than students in comparison schools (Urdegar, 2000). In an analysis of these studies and others in cities nationwide, Pogrow (2002) claimed that previous studies that showed the benefits of SFA were based on faulty or misleading methods and were conducted either by the program's creators or employees of the Success for All Foundation. Others (Walberg & Greenberg, 1999) claimed that independent evaluations had found no improvements or improvements only at kindergarten and first grade where more cost-effective programs had also been shown to be effective. These critics of SFA have been engaged in an ongoing debate with supporters of the

program (Joyce, 1999; Pogrow, 1998, 1999; Slavin, 1999).

Research on school-wide interventions has indicated that keys to interpreting the success of such models are understanding the extent to which they have been implemented with fidelity and examining the constraints under which they operate (Herman et al., 2000). Even the creators of SFA offer the caveat that their program is effective only when "fully implemented" (Slavin & Madden, 2001b, p.34). In a study of two Mississippi elementary schools, Wells et al. (2000) found that teachers' adherence to SFA program protocols was partial and that effectiveness varied greatly due to inconsistencies in program implementation.

The Current Study

In this study we did not attempt to answer questions about SFA's overall effectiveness, though our research was conducted in a school district in which SFA did not prove to be successful and was ultimately abandoned (Levitt, 2000; Urdegar, 2000). Our purpose was to shed light on the problems encountered when trying to implement SFA in high-need urban schools with culturally and linguistically diverse populations. This research was part of a larger project that investigated the disproportionate representation of culturally and linguistically diverse students in high-incidence disability categories in special education (Harry & Klingner, 2006). As part of this project, we observed instruction in general and special education classrooms in 12 schools. Because four of the highest-need urban schools in our sample were using SFA as their reading program, information on SFA curriculum and instruction formed a substantial subset of our data and raised interesting questions about the effectiveness of the SFA model with struggling readers. We wanted to understand the nature of SFA instruction as well as how SFA fit into the larger school context.

Our work in the larger project was based on the National Academy of Sciences'

(Heller, Holtzman, & Messick, 1982) approach to understanding disproportionality, which stated that disproportionate representation is problematic if any aspect of early instruction or the referral, evaluation, and placement process works against students' success. We were operating from the belief that the events that occur in the general education classroom will ultimately affect student performance and teachers' decision making about referrals.

Method

Participants

The student population at all four schools was predominantly black (e.g., African American, Haitian, Haitian American), although one school also had a growing Spanish-speaking population. Almost every child was receiving free or reduced-price lunch, as shown in Table 1. Teacher ethnicity was mixed, with no more than 45% or less than 20% of the teachers at any school of white, black, or Hispanic origin. At three of the schools, 41% to 46% of the teachers had completed advanced degrees, and at the fourth school this percentage was 26 (see Table 2). We observed in the classrooms of 30 teachers, five of whom were special education teachers.

Data Sources

Data for the larger study were collected through a process that progressed in a funnel-like fashion over three phases, moving from a description of district-wide placement rates and referral/placement policies, to an examination of the implementation of the referral/placement policies in 12 schools that represented a variety of ethnic and socioeconomic configurations, to individual case studies of 12 students. During Phase 1 of the study we interviewed district administrators and school personnel and examined documents. Data for this study of SFA were collected during Phase 2 of the project. During this phase we observed in every kindergarten through third-grade general education classroom in each of the

TABLE 1. Student Demographics, in Percentages, by School

School	Ethnicity			Free or Reduced-Price Lunch	ELL	EMH	LD
	W	B	H				
1	0	98	2	97	1	8.0	13.4
2	0+	91	8	98	39	.2	3.0
3	0+	87	11	99	26	.1	5.7
4	1	69	30	99	16	1.9	4.4

NOTE.—W = white; B = black (used to refer to African American students as well as those from Haiti, Jamaica, and other Caribbean countries); H = Hispanic (the term generally preferred in this region for individuals from Cuba and other Caribbean countries, Mexico, Central America, and South America); ELL = English language learners not yet considered fluent in English (referred to as Limited English Proficient by the district); EMH = educable mental handicaps (term used in this district); LD = learning disabilities.

TABLE 2. Teacher Demographics, by School

School	Ethnicity of Teachers (%)			Teachers' Educational Level (%)			New to School (%)	Mean Years Teaching	No. Observed Teaching SFA
	White	Black	Hispanic	Master's	Specialist	Doctorate			
1	44	36	20	36	10	0	17.1	11	8
2	36	34	25	35	6	0	8.5	9	6
3	36	36	26	35	4	2	11.8	13	8
4	30	25	45	21	5	0	17.0	10	8

12 schools (with a few exceptions) and interviewed administrators, teachers, and other support personnel in each school. We also selected two teachers and their students from each school for more extensive observations (between eight and 12 observations in each classroom). We selected teachers to reflect a range of grade levels, ethnicity, experience, and referral patterns (high, medium, or low). We selected students to represent a range in possible disability categories and, to the extent that we could, ethnicity, age, and gender. We believe that the use of multiple sources and data-collection strategies provided us with considerable triangulation of data. We entered all interview transcripts and summaries as well as all observation field notes into an ATLAS.ti qualitative analysis database (a software program that facilitates the examination and testing of associations among analytic categories and assists in the development of a conceptual model).

Observations. We took extensive field notes during observations, using proce-

dures suggested by LeCompte and Preissle (1993). During the earliest days of the research project, our field notes were fairly general. As the research progressed and emergent issues were identified, our field notes became more focused. In our observations of SFA, we started with a broad view of the instruction itself, recording the teacher's actions and to some extent the behaviors of students who stood out as having problems. Over time, our focus narrowed as we honed in on specific target students and their interactions with teachers and other students. Altogether across the 12 schools we conducted 627 classroom observations and observed 42 child study team meetings, five psychological evaluations, 15 special education placement and/or individual education plan (IEP) meetings, 14 other school-based meetings, and 15 home and community settings. At the four SFA schools we conducted 162 classroom observations. Among these were 21 complete or almost complete SFA lessons and 24 partial lessons, for a total of 45 SFA observations.

These were conducted in 30 teachers' classrooms.

Interviews. We primarily used a semi-structured interview format. Interviews were semistructured in that they followed a set of guiding questions, though because our emphasis was on eliciting spontaneous and personalized perspectives, our interview style was open-ended, flexible, and responsive to the leads of the interviewees. Thus, we frequently used follow-up probes that were unique to a particular interview. Sample guiding questions included: (1) Please tell me about the referral process at your school and the child study team. Who is involved? What steps are followed? What challenges do you face? (2) Please tell me about your reading program. How well do you think it is working for struggling readers? What challenges do you face?

Across the 12 schools we conducted 272 individual interviews with students, parents, and school-based and district personnel, including principals, assistant principals, counselors, social workers, parent liaisons, staffing specialists, psychologists, bilingual assessors, SFA facilitators, SFA tutors, classroom teachers, paraprofessionals, and a security guard. We also conducted 84 informal conversations (for which we kept field notes). At the four SFA schools the total number of interviews and conversations was 156. Some of teachers' most revealing comments about SFA were made during, or at the end of, observations of SFA instruction.

Documents. We collected documents such as students' work samples (including SFA work), individualized educational plans, psychological and other evaluation reports, school district guidelines and policies, and extant data on special education placement in the school district.

Data Analysis

Our data analysis for the larger project proceeded through several phases. We applied grounded theory and ethnographic techniques to develop theory inductively

using the constant-comparison procedure (Glaser & Strauss, 1967; Strauss & Corbin, 1990). The recursive nature of the constant-comparison method allowed each phase of data collection and analysis to feed into the next. In the first phase of data analysis, during and immediately following our first phase of data collection, we read through interview transcripts and "chunked" and categorized sections that seemed to fit similar patterns. From these interview data we identified a set of themes that represented school personnel's explanations of the genesis of disproportionate representation. We further analyzed each of these themes into a set of codes and categories that we applied to all subsequent data (see Harry, Sturges, & Klingner, 2005).

As we entered the second year of our study, we continued conducting interviews and began observing in classrooms. We added codes to capture what we were observing, mapping these onto the initial set of themes that had emerged from our early interviews: teachers' classroom management skills, instructional skills, relationships with students, and perceptions of students. We rated each of these as positive, neutral, or negative. This coding and rating process was first conducted by our entire team over several meetings while we developed a common perspective, and then in pairs. For SFA data, we added additional codes: SFA grouping, SFA levels, SFA scheduling, SFA content, and SFA activities (see Klingner, Sturges, & Harry, 2003).

Once we had categorized and coded data using this set of descriptors, we applied finer levels of coding. We did this by comparing a given data chunk to another to see if the same code would apply, thus developing consistency in usage of the codes. After we had developed a refined set of codes, each of us applied these codes to the existing dataset as well as to new incoming data, and we met regularly to compare our coding decisions (see the Appendix for a list of these observation codes). There was no attempt to develop a numerical reliability

check. Rather, our goal was consensus, with each difference being debated and clarified until we agreed on appropriate use of the codes.

Next we clustered codes into related, overarching categories and moved to the interpretation of themes that applied across categories and pointed to tentative explanations of the data. For SFA, these themes included problems with grouping and students with a variety of grade or reading levels in the same class, students recycling through the same materials over and over, not enough assistance for the lowest students, scheduling procedures, students and teachers getting bored, adaptations to SFA instructional and assessment procedures, and influences on the referral process for special education. In addition, we noticed that the quality of instruction seemed to reflect teachers' own skills more than the SFA curriculum and instructional strategies. To focus on these issues, we reflected about data as we typed our field notes and then discussed our ideas during weekly project meetings. These discussions played an important role in that they assisted us in thinking about what we were finding and helped guide subsequent data collection. For example, after one of us observed a third-grade class with mixed instead of homogeneous reading groups, as SFA requires, she brought this up at a meeting, describing the observation and asking if others had witnessed anything similar. Through discussion we agreed that we should explore the phenomenon of mixed-level classes further. Thus, the processes of data collection and analysis were recursive and continued throughout the project.

Results

Implementation of SFA varied greatly across the four schools and 30 classrooms. Although we saw numerous instances of excellent classroom practice, we did not find any examples of effective school-wide implementation. We noted several patterns, which we describe next.

Good Teachers Are Good SFA Teachers

Those teachers who seemed to be effective in their use of the program were the same teachers we rated as effective in other areas of teaching (13 of the 30 teachers). Teachers judged to be less effective were, in turn, less effective in their use of the SFA program (11 of the teachers; the remaining six we considered adequate). Likewise, teachers who were rated to have better classroom management skills were able to implement the program with fewer management complications, whereas teachers who were judged to have trouble with classroom management had management problems that interfered with SFA instruction. This can be illustrated through examples of our observations of SFA in classrooms.

In a kindergarten classroom in an urban school with a 98% free and reduced-price lunch rate and 27 of 28 students of African American descent, an observation of SFA from the first month of school showed that students were engaged, on task, and motivated to learn. Taken from the observer's field notes:

She tells the students that she has some "super hard" words for them today. Smiling broadly, she says, "I don't know, if you go home and read these words, they're going to take you out of kindergarten because they'll say you are too smart for kindergarten." The kids laugh gleefully. The first word is "optometrist." The kids really are proud of themselves when they are able to "read" the word. She asks what it means, and a student responds, "A doctor who looks at your eyes." (She had introduced the word by saying this, asking who had been to a doctor who looked at their eyes.)

This teacher was also rated as having effective classroom management skills, and this carried over to her SFA instruction:

The teacher says, "I need eyes back on me. I need your hands back in your lap." They do "Grammar's Glasses" (a brief

rhyme designed to focus students' attention) and finish ready and eager to listen to the story. She tells them, "Sometimes we read stories, and sometimes we listen to stories." She reads the story to them with enthusiasm and expression—all students listen attentively, their eyes on the teacher or the book, some with their mouths slightly open as if transfixed. Then the teacher asks questions: "Why do you think they might not want to play with him?" She tells them, "I want to see different hands." She regularly reinforces students raising their hands rather than calling out and doesn't call on students who don't raise their hands quietly. To one student who has already responded, she says, "It's okay, my love, I want to hear from someone different." She asks students, "What do you think about that?" "What might happen now?" "Why did you think it was good?"

In a third-grade class in a different school, a teacher struggled with both instruction and management during her SFA lessons, as demonstrated by the following field notes (names are pseudonyms):

Randy is becoming noticeably more agitated. He says, "I not finished, mommy, mommy. I not going to catch up." He continues writing. . . . The next sentence on the board the students are supposed to be copying says: "He run after the bird and catch it." [Observer's comment: Yes, "run" rather than "ran" and "catch" rather than "caught," written by the teacher.] The teacher is saying, "The book says 'chased,' but we aren't going to put that because it's too hard for you." . . . Randy is up out of his seat again. The teacher asks him sternly, "What are you doing?" He responds matter-of-factly, "Looking for my pencil." Walter now says, "Look at Mark on the floor!" . . . Mark is now doing caterpillar- or worm-like dancing movements. Randy asks him, in a tone of admiration, "How you do that?" Mark does it again. The teacher says, "Don't pay attention to Mark because that's what he wants." Now Randy is on the floor, too. . . . The teacher calls students to come up to an area on the rug and sit on the floor. Some students comply, but many are fooling around.

Adaptations to SFA Instructional Procedures

Beyond the fact that effective teachers' instruction and management skills did not depend on the SFA program, we also noted that such teachers did not follow the SFA manual precisely. Two excellent teachers in a school serving a large English language learner (ELL) population predominantly comprising Haitian Creole speakers asserted that, although they found the SFA materials quite good, the scripted lessons and strategies did not address the language and acculturation needs of their first and third graders.

The first-grade teacher explained that she only used the key aspects of the manual that she found important and that the timing expected for most activities was inappropriate for her lessons. In particular, she spent considerable time on pronunciation and vocabulary with the children, with the result that a supposedly 2–3-minute reading activity took much longer. For example, our observations noted her pausing in the reading of the sentence, "Mrs. Stanton stands by the pond," to model and have the children practice the "st" sounds. She did the same with words such as "slip" and "spin"; in the latter case the phrase "spins past" was difficult for children both to understand and pronounce.

The third-grade teacher in this school made similar comments but focused more on the need to make decisions about the appropriateness of SFA material to children's abilities and interests. With regard to a requirement that the children read and edit each other's work, she exclaimed: "I said, 'Forget it.' It's just a time waster. You're not learning. . . . If you don't know and I don't know then there's nothing that I can do to learn from you and you can't learn from me. So I just cut it out. They (the SFA supervisors) didn't push it this year, so I'm assuming they're cutting it out too."

Another strategy that this teacher modified was called "jump-in reading," which is supposed to enable the children to read

spontaneously (i.e., take turns reading aloud individually without being called on by the teacher to do so). Her students, she said, had neither the reading skills nor the confidence to do this so she modified the activity by choosing the readers, but in lively and spontaneous fashion that was observed to be highly successful with her third graders. In the excerpt that follows, the student we were monitoring, Manuel, was a native Spanish speaker whose oral expression and listening comprehension were much more advanced than his reading and writing skills. The field notes read:

Then the teacher says, "It's time to have some jump-in reading!" As she says this, she wiggles her legs and puts on a funny facial expression that makes the children giggle. She begins going around the room touching kids on their shoulders to indicate the next reader. The kids are all attentive. Manuel turns to me and says, "This is called jump-in reading!" The teacher pauses to ask some questions about the story . . . [which lead to lively responses from the children]. . . . Then she goes on to more jump-in reading, hurrying around the room picking new readers. She exclaims, "I'm looking for people who haven't read!"

This teacher also modified the time she spent on stories:

If we start a story and it's really boring, I'll go through it quickly and just go on . . . they're supposed to be interested in reading. It's supposed to be fun. If you give them too many that are boring, they're turned off. But don't tell SFA that. . . . For example, they [the kids] were dying cause they couldn't get to the *Titanic*, which was in level 3-1, so when we got there it took us 3 weeks to get through that story because we just did a lot of extra things to go with it. That's what makes it interesting to the children.

Teachers felt justified in making these adaptations, explaining, "That's what I went to school for." For example, in an inner-city school in the sample from our

larger study (other than one of the four SFA schools reported on here), a teacher told us that the school had used SFA for 1 year prior to the start of our study, and then the faculty had voted it out. She explained, "I didn't get a master's degree to read strictly out of a manual. I can follow a format. You give me a program, your guidelines, and that's all I need. That's what I went to school for. If not, all you need to do is get someone with common sense, who likes children, and give them that (SFA). And it's boring. The material is boring. There are some interesting components, but mostly you're doing the same thing all the time. . . . The teachers hated it; the children hated it."

Lowest Students Not Adequately Supported

Although SFA has built-in supports for struggling students, we observed that for the lowest students, these supports were not sufficient and not easy for schools to implement. We drew this conclusion because several students were not able to progress to higher levels of the program. Though SFA includes one-on-one tutoring for students not making adequate progress, this assistance was only provided to first graders, leaving older struggling learners without tutoring or other extra support. The tutoring helped those first graders whose progress accelerated when they received it, though for some it ended too soon.

We observed numerous students who appeared to be languishing. For example, a third grader named Rex was not making progress. At his staffing, the teacher explained that he was "very eager to learn but can't learn to read and can't retain information from one day to the next. We have tried different ways of teaching him. This is his fourth year in SFA, and he is still in Level 6 (an early kindergarten level)." Also, in the third-grade observation described previously, the lower students were not receiving reading instruction at the appropriate level.

In our sample of 12 schools for the larger

study that formed the backdrop for this study of SFA, we noted a common trend whereby the lowest-achieving students tended to be taught by the least qualified or least effective teachers. In one school this was also true for SFA groups, where the lowest achievers were taught by a paraprofessional rather than a certified teacher. The teacher in a first-grade, self-contained ESOL class explained that SFA grouping was “supposed to be by ESOL level, but it’s really like the low readers go with the paraprofessional.”

Recycling through the Same Materials

Students who did not make adequate progress, such as Rex, recycled through the same materials repeatedly, in general and special education classes. After being placed in special education, Rex continued to be taught with SFA using the same books he had already “read” many times and with which he had not progressed. We noted that many older students had already memorized the books they were supposed to be reading and had lost interest in them. In one observation the teacher told the class they were going to read “the best story” they had read in a long time, “Land of the Midnight Sun.” She asked, “How many of you know this story?” All of the older students raised their hands. These students then ran around the classroom and climbed on desks while the teacher tried to teach the rest of the class. The teacher remarked later that students were bored and frustrated. Boredom was a concern for many teachers as well as students. We heard numerous comments such as the following from a first-grade teacher: “My students already know all of the stories. I keep teaching the same thing over and over, and it’s so boring.”

Mixed Instructional Levels

In one SFA school in particular, some classes seemed to be “dumping grounds” for students considered to be troublemakers, although they were not all at the same level and therefore should not have been in

the same class. Field notes from a conversation with a third-grade teacher in this school demonstrate this:

She explained that in SFA, it’s only supposed to be one level, but she has three, because the administration groups students “not by level, it’s by the type of kid.” She says that she has the students who are behavior problems, and that the assistant principal told her to only read one level with them. But she tells me that she wouldn’t feel right, couldn’t go home at the end of the day knowing that she’d only read one when she has three (or more). So that’s why they read three books at the beginning of my observation, one for each level. Some students are still at Levels 1–5, and others are 16 or 17. That’s why some kids weren’t able to read the last two books. Randy is at the highest level—he can read, but he’s in her class because of his problematic behavior.

We observed this in special education as well as general education classrooms. In a second SFA school, the special education teacher told us that the previous year she had had five groups (levels) of four to seven students each at once. She said that the class size increased too much, so she got permission to reverse the schedule and teach language arts and math in the morning and spend the rest of the day on SFA. This was the only way she could get in enough SFA instruction for all her groups.

Concerns with Testing Procedures

The issue of appropriate placement of students at SFA reading levels also reflected concerns with testing procedures. Teachers often believed that the determined level was incorrect and substituted their own methods of assessment. For example, one teacher said that she tested her English language learners herself and found that her testing showed different results from those the SFA tester obtained. After her testing, she moved children to the “correct” groups. She explained that in giving the tests she “explored the meaning [of the story on the

SFA test] more" because she could not assume that students had understood the story enough to answer the official test question. So she "expanded on" the questions, as in the following example: "The story is about how Arthur lost his dog. The SFA question is, 'What happened after Arthur realized the dog was lost?' But I'll question them about prior information before asking that. Like, 'What was he doing with the dog?' and so on, step by step, to be sure they know the context." By so doing this teacher placed students in levels that were higher than those indicated by the SFA tests.

A third-grade teacher of Haitian Creole speakers moved in the opposite direction. She said that she made her own decisions with her group of third graders who were "struggling at a 2-2 level. . . . Whoever did the testing for placement helped them with the test enough to get them into that level, because they just weren't ready. So we went back to 2-1 and started lower. . . . The program doesn't like them to go backward, but . . . I'm like—you're building a house on quicksand here!"

One principal was so dissatisfied with the accuracy of the SFA testing that she replaced it with another assessment strategy to determine reading levels and form more homogeneous groups. She said that she believed the problem with the SFA testing was that "they are tested at the beginning and never re-tested. . . . They may be moving along reading levels but not really reading at those levels." Overall, she found that the SFA testing tended to place children at higher levels than they should be.

Differences in how students were assessed and placed at levels meant that, when students transferred from one SFA school to another, they did not necessarily continue at the same level. When a student at one SFA school transferred and was tested at his new school, he was found to be reading at a much lower level than he had been at the previous school. His new teacher said that "at [the other school] they just move kids along whether they are

ready or not, but here we don't move kids until they pass to the next level." From our notes on a CST meeting (child study team or prereferral meeting): "Meanwhile, the SFA facilitator explained that when Gordon had first come to Stowe, the paperwork from (the other school) indicated that he was reading in Book 42 in SFA. But they tested him, and he placed in Book 21, much lower, at a mid-first grade level. It became clear to me quickly that the school thought the mother had been upset about their putting Gordon back in SFA, but in fact the opposite was true. The mother was upset that the previous school has pushed him along when he couldn't read the books."

Mixed Grade Levels

Not only did we see mixed instructional levels in the same class, but we also observed students from mixed primary and intermediate grade levels. For instance, one class with first through fifth graders was a particular challenge for the teacher because the first graders were motivated and on task, whereas the older students were turned off and unruly. As the teacher lamented, "It's so bad for the first graders to see those examples." SFA stipulates that students from such a wide range of grades not be placed in the same classroom, but this teacher told us that sometimes this occurred because the principal (or other SFA administrator) did not know where to put the older students who were reading at very low levels. She explained, "They will tell you that this shouldn't happen, that when they are that low, there should have been a meeting with the parent (at a family support team meeting), and a CST, but the CSTs take a year, a year and a half, or sometimes even 2 years (e.g., because of a backlog of students or inaction on the part of an administrator)." The following year this same teacher commented that teachers who do not make referrals to CSTs are one of the reasons her school has had problems with SFA and with students repeating levels

“over and over” and with students of varying ages placed in the same class.

A special education teacher at another school explained that she was supposed to have two kindergartners mixed with second and third graders for SFA instruction but that she preferred to work individually with kindergartners. Therefore, she arranged with the kindergarten teachers to keep the two students for SFA reading in the mornings. In the afternoons she took them out of their classroom and helped them with reading. “This way they get reading twice. They would not get the one on one they need if they’re in with the older children.” She said this was unofficial, she was not supposed to do this, but she thought this was the only way. She gestured to her lips with her finger, to indicate that the researcher should keep this confidential.

Problems with Scheduling

Teachers considered it problematic that they did not have their own students for SFA and were assigned different students every 9 weeks, so that they were not able “to really get to know their students.” This was also a problem during the entire school day because their homeroom students also switched to other teachers later in the day for other subjects (e.g., physical education, Spanish or Haitian Creole, art, music, computers). One teacher complained:

“I don’t have time to teach. I have them for less than 30 minutes before they go to SFA, and then afterward they go to PE. Just when they have settled down and are working well, it’s time for them to leave again. Then they come back for 10 minutes, then they go to lunch. Then they come back and go to Spanish. Then they come back at 2:00. We’re the ones that make them hyper. We have them running around all day. . . . I went to the principal and she told me, ‘Make the most of it.’”

Because of this fragmentation, a teacher would often not know the reading abilities of homeroom students who were not in the teacher’s SFA group. Another teacher

summed up the feelings of many by saying: “With (the state’s high-stakes mandatory testing), SFA, CCC [Computer Curriculum Corporation, a required program that necessitated students working at a computer for 30–40 minutes every day], teachers don’t have time. . . . Teachers need more flexibility to do things with the students and for themselves. Teachers feel very powerless at times. We are not empowered. They’ve taken away our flexibility. I can’t meet a child’s needs if I’m not allowed to be flexible.”

Another teacher said that this pattern of first to third graders moving from class to class and program to program made school an “uncertain” place, especially for her immigrant ESOL students who were, as she put it, “just off the boat.” The fragmentation of the school day made it extra difficult for these children to begin the process of attachment and acculturation they desperately needed.

Special Education Referrals

SFA did not seem to directly affect the special education referral process, other than at one school where teachers used SFA placement test results to determine which students to refer for a special education evaluation. A first-grade teacher explained, “It is the ones who have made no progress in reading (on the SFA level placement tests) we are referring.” She went on to say that the principal had directed teachers to do this. Yet this school did not have more overall referrals than the other schools and had the second lowest percentage of students identified as learning disabled (4.2%).

In two schools the lack of timely special education placement for students who seemed to need it affected SFA implementation because it meant that students remained in general education classrooms, repeating the same SFA levels rather than receiving instruction from a special education teacher. Two teachers commented that one reason SFA did not work as well as it was supposed to was that some teachers

did not refer their lowest readers to their school's child study team. A third-grade teacher told us, "One of the reasons they've had problems with SFA is kids repeating levels over and over. These are the kids their teachers should have referred, but some don't."

SFA is supposed to include a family support team made up of the SFA facilitator, an administrator, teacher, social worker, and/or a parent liaison. The team is intended to provide assistance for students who struggle and to facilitate parental involvement, but in two schools there was no evidence of the team's existence and in the other two schools staff used the team only sporadically. In general, school personnel seemed unsure of the purpose of the team, their roles, or how this team coordinated with the child study team. The SFA facilitator at one school, however, seemed clear about what was expected from the two teams. She explained the process to us:

If they are having problems, we pull them during the afternoon for one-to-one tutoring. This step takes place before a student is referred to special education. If the student is still having a lot of problems, they refer the child and ask me how they are doing, if they have been getting tutoring. As part of SFA we have a family support team. This meeting is in addition to the CST [child study team], but sometimes they overlap. I'm in charge of the family support team. . . . We plan, organize, talk with the parents. We try to avoid a CST and talk about how to provide extra help. We try to come up with solutions, and then we follow up. Then they use that information in CST meetings [she herself rarely attended CST meetings]. . . . We try to hold family support team meetings every week, sometimes every 2 weeks [in reality, it was not this often]. . . . but I have to do so many things, check attendance, parent involvement, school-based interventions.

When asked how many students who have been referred to the family support team then get referred to a CST, she replied, "About 70%. After we try and we see no change, then we push for them to be re-

ferred, not only for reading, but in general." Interestingly, none of the teachers, administrators, or other support staff in this school mentioned the family support team to us during our several conversations and interviews with them.

Discussion

In our observations of SFA in four urban schools, we noted that challenges affected the fidelity with which schools and teachers implemented the model. This issue of fidelity has been at the center of debates about SFA and other school-wide instructional models (Herman et al., 2000; Pogrow, 2002). Though our data clearly indicate uneven implementation among teachers, it is difficult to assess either the reasons for or the direct outcomes of this pattern.

Similar to others (Wells et al., 2000), we found that teachers in the schools we studied varied in adherence to SFA protocol and in instructional effectiveness. Also, as other researchers have found (Walberg & Greenberg, 1999), we noted that older learners who struggled with reading often entered a cycle of boredom and repetition of materials. This led to grouping and scheduling challenges, with students of different reading abilities and widely varying grade levels sometimes placed in the same class. One of the greatest implementation challenges involved students stagnating and not passing to higher levels of the program, thus recycling through material they had already covered. Or, in some cases, students were moved to the next level even though they had not passed the placement test for that level.

The finding that highly effective teachers adapted the program to address the needs of struggling readers raises the questions of whether scripted materials and strategies can meet a wide range of learning needs and whether skilled teachers should have their instruction scripted in this way. We believe that experienced and talented teachers can make adaptations that overcome program limitations and enhance

rather than undermine the program's effectiveness. In their work on factors affecting the sustainability of research-based practices, Klingner and colleagues consistently found that teachers were more likely to continue using a practice when they felt they had the flexibility to modify the practice to fit the needs of their students (Klingner, Ahwee, Pilonieta, & Menendez, 2003; Klingner, Arguëlles, Hughes, & Vaughn, 2001; Klingner, Vaughn, Hughes, & Arguëlles, 1999).

However, data from the study reported here suggest that neither SFA's scripted nature nor its materials can solve the problems of teachers with weak instructional or management skills. Yet it is hard to separate teacher effects from program effects. We do not mean to imply that teacher effectiveness and program fidelity are inversely related, but simply wish to show that effective teachers are more skilled at making essential adaptations that can enhance program effects. We and others have found that, for teachers to implement multicomponent reading programs effectively, they must understand not only what the components of the program are and how to use them but also why and when (conditional as well as declarative and procedural knowledge). Conditional knowledge is necessary for teachers to tailor instruction to the needs of their students and seems to be the hardest for teachers to acquire (Klingner, Vaughn, Arguëlles, Hughes, & Ahwee, 2004; National Reading Panel, 2000; Pressley & El-Dinary, 1997).

We argue that teachers who have a solid grasp of a practice, confidence in their knowledge of how to teach, and sophisticated understandings of their students' needs as well as contextual factors are better able to adjust instruction. These adjustments may be planned or happen spontaneously during teaching by teachers skilled at "thinking on their feet." Would the teachers in our study have changed procedures if they had been presented with classrooms of look-alike students all at the same level, with similar needs? We do not think so, but

that they adjusted when their situations varied from the SFA ideal is a testament to their skill. We believe that SFA would be more effective if teachers were officially granted more flexibility and empowered to make decisions when their classroom context varied from what the developers of SFA envisioned. Yet SFA is presented as "all or nothing" by the developers of the program. Eventually this district discontinued SFA: "It is recommended that . . . Success for All be discontinued . . . unless there is full support of staff and the necessary financial and staff resources leading to the possibility that the curriculum can be implemented **exactly** as required by the Success for All Foundation" (Levitt, 2000, p. 10). None of the four schools in our sample used SFA the following year.

Besides problems with instruction, difficulties with accurate assessment of students' reading levels were endemic. We do not have adequate data to reveal whether this problem was related to the implementation of the assessment or to its design. A corollary question is whether the SFA assessment, curriculum, and strategies adequately took into account the needs of young English language learners. The teacher who said that when assessing students' reading she probed their understanding of the story "step by step," until she felt they were ready for the question on the SFA test, may have been doing them a disservice by inflating their competence and thus "building on quicksand." Nevertheless, this teacher's need to modify the SFA testing procedure suggests that it may not have accurately assessed her students' vocabulary and comprehension levels.

We do not know whether some students did not fare well because their schools and teachers did not implement the program more effectively, or, conversely, whether schools and teachers could not more effectively administer the program because they faced so many external challenges. It is easy for those not directly involved to say (as we have heard), "Well, it was not working be-

cause they were not doing it right—if they had just implemented it the way they were supposed to, then they would have achieved better results.” Our impression, however, is that these schools were trying their best to solve the dilemmas they faced. They were confronted with numerous pressures from many levels that complicated both the implementation of the program and its potential for effective outcomes. Specifically, we noted a heightened emphasis on accountability through high-stakes testing, excessive poverty and associated factors that put students at risk for school failure, high mobility rates among teachers and students, and numerous mandated programs that schools were expected to implement simultaneously. It seems unfair to blame a school’s administration for poor implementation of an instructional model if implementation is not feasible because of other pressures. The unique characteristics of the student population and the school context exerted a powerful influence on how these schools implemented SFA. Thus, we argue that the issue is not only one of fidelity of implementation but also of feasibility in a particular context.

Could the leadership in these schools have responded differently? Could they have handled the challenges they faced in other ways? Where did the system break down? Perhaps more professional development and support for principals and other school leaders in how to deal with challenges would have helped. In their review of research on 24 comprehensive school reforms, in which SFA was identified as one of the three most successful, Herman et al. (2000) found that key aspects of faithful model implementation were teacher buy-in, professional development, school leadership, district support, parent and community involvement, and technical assistance and resources. Unfortunately, our data do not reveal the extent to which these requirements were present. However, the comments of both the principal and assis-

tant principal in one school suggested less than total “buy-in”:

I’ve learned to accept change. Every 5 years we change reading programs anyway, so each time this is supposed to be a cure, when I know nothing is a cure for all. . . . Any program you implement properly . . . if you think it’s so good, you’ll see the improvement.

I can use a reading program, and if I don’t believe in it as a teacher it will not work. I don’t care what the program is, because I say that teaching reading is eclectic because you pull from everywhere to help that child learn to read. . . . This program gives the structure, and any teacher can add to it and be as creative as possible.

This study has its limitations. Although we spent considerable time observing in each classroom and developed an extensive dataset, our primary concern was not judging the effectiveness of SFA instruction, nor are we experts on SFA implementation. Rather, we viewed the observations of, and interviews about, SFA as a subset of the data and applied to them the same analytic procedures we used with the larger body of data regarding the referral and decision-making processes by which students are placed in special education programs. Also, we based our findings on our observational and interview data but did not attempt to relate these to students’ standardized test results or grades. We consider our findings regarding SFA to be sound in terms of offering an accurate description of significant aspects of the program’s implementation in these schools, but we do not argue that our data indicate clear outcomes of the program. However, we do know that 1 year after our study this school district discontinued SFA implementation, in part because student outcomes were not as expected.

In summary, our data raise interesting questions about SFA implementation that merit further exploration. How feasible is it to expect that SFA (or any packaged reading

program) can be implemented without requiring changes in response to the local context? When schools have difficulty implementing various aspects of the program, what assistance might help? How much of a difference would it make to provide additional support for those students who continue to struggle? What would happen if expert teachers were empowered to draw on their expertise and make refinements in the program? Then a follow-up question would be, Which aspects of the model must be kept intact and which can be altered?

In light of the popularity of SFA and other school-wide initiatives designed to meet the needs of students considered at risk, it is important for teachers and researchers to be aware of the challenges that those who attempt to implement such programs in inner-city schools face. We believe that in high-need schools such as these, success for all rather than just for some will take more than the wholesale implementation of a packaged reading program; it will require the thoughtful and flexible implementation of a program or programs by those who are experts not only in literacy and the model but also in the local context.

Appendix

Refined Set of Codes for Classroom Observations (Brief Version)

Demographics:

- teacher:
 - name
 - description (ethnicity, gender, approximate age)
 - information about reference to CST
- students:
 - number
 - gender
 - ethnicity
 - special education
 - potential target students

Classroom:

- desk arrangement
- organization
- posted rules
- posted classroom management system (e.g., chart with pockets and

green, yellow, and red cards for each student)

- what's on the walls (e.g., bulletin boards, decorations, students' work)
- other

Teacher:

- location
 - in front of students
 - circulating around the room
 - seated at his/her desk
 - other
- instructional activity
 - whole-class instruction
 - working with a small group of students
 - providing assistance to individual students
 - other
- classroom management
 - procedures (including routines and transitions)
 - expectations (e.g., teacher stressed that class time is to be used productively)
 - encouragement/support (e.g., "I like how William is sitting so beautifully")
 - desists (e.g., tells students to put their heads on their tables)
 - ignoring of students' behavior
 - demeanor (e.g., friendly, respectful, angry, businesslike, "in charge")

Student:

- location
 - at seats
 - sitting on the floor by the teacher
 - at the computer
 - at centers
 - other
- activity
 - noncompliant behavior (in relation to what seems to be expected in the classroom; e.g., kicking another student)
 - listening/watching during whole-class lesson
 - working on an assignment
 - reading a book
 - responding to teacher's questions
 - working at the computer
 - helping another student
 - other
- grouping configuration
 - working independently
 - working with a classmate
 - working in a group
- demeanor (e.g., excited, confused, energized, engaged in the task)

Lesson:

- content
 - subject
 - topic
- product (i.e., assignment students are expected to complete)
- equipment/materials:
 - overhead projector
 - dry-erase board
 - computers
 - visual aids
 - manipulatives(e.g., unifix cubes)

Other:

- visiting students (from another classroom):
 - location
 - activity
- interruptions
 - announcements over the intercom
 - visitor to the classroom

Researcher:

- location
- activity (e.g., helping a student)
- impressions
- level of comfort

References

- Borman, G. D., & Hewes, G. M. (2001). *The long-term effects and cost-effectiveness of Success for All*. Baltimore: Center for Research on the Education of Students Placed at Risk.
- Clarke, P. A. (2001). *Analysis of Success for All and school development programs and their effects on reading comprehension*. Unpublished master's thesis, Kean University, Union, NJ.
- Fashola, O. S., & Slavin, R. E. (1998). Schoolwide reform models: What works? *Phi Delta Kappan*, *79*, 370–379.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.
- Harry, B., & Klingner, J. K. (2006). *Why are so many students in special education? Understanding race and disability in schools*. New York: Teachers College Press.
- Harry, B., Sturges, K., & Klingner, J. K. (2005). Qualitative data analysis: Mapping the process. *Educational Researcher*, *34*(2), 3–13.
- Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity*. Washington, DC: National Academy press.
- Herman, R., Carl, B., Lampron, S., Sussman, A., Berger, A., & Innes, F. (2000). *What we know about comprehensive school reform models*. Washington, DC: American Institutes for Research.
- Joyce, B. R. (1999). The great literacy problem and Success for All. *Phi Delta Kappan*, *81*(2), 129–131.
- Klingner, J. K., Ahwee, S., Pilonieta, P., & Mendendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children*, *69*, 411–429.
- Klingner, J. K., Arguëlles, M. E., Hughes, M. T., & Vaughn, S. (2001). Examining the school-wide "spread" of research-based practices. *Learning Disability Quarterly*, *24*(4), 2221–2234.
- Klingner, J. K., Sturges, K., & Harry B. (2003). Conducting ethnographic classroom observations of literacy instruction. In S. Vaughn & K. Briggs (Eds.), *Reading in the classroom: Systems for observing teaching and learning* (pp. 145–177). Baltimore: Brookes.
- Klingner, J. K., Vaughn, S., Arguëlles, M. E., Hughes, M. T., & Ahwee, S. (2004). Collaborative strategic reading: "Real world" lessons from classroom teachers. *Remedial and Special Education*, *25*, 291–302.
- Klingner, J. K., Vaughn S., Hughes, M. T., & Arguëlles, M. E. (1999). Sustaining research-based practices in reading: A 3-year follow-up. *Remedial and Special Education*, *20*, 263–274.
- LeCompte, M. D., & Preissle, J. (1993). *Ethnography and qualitative design in education research*. (2d ed.). San Diego: Academic Press.
- Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity. *Educational Researcher*, *31*(1), 3–12.
- Levitt, J. (2000). *An interim evaluation of operation safety net: Three-year report*. Miami: Miami-Dade County Public Schools, Office of Evaluation and Research.
- Madden, N. A., Slavin, R. E., Christie, N., & Karpouzis, K. P. (2001). *Success for All alignment with Stanford Achievement Test: Ninth edition*. Washington, DC: Office of Educational Research and Improvement.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* [On-line]. Available: <http://www.nichd.nih.gov/publications/nrp/smallbook.pdf>
- Pogrow, S. (1998). What is an exemplary program and why should anyone care? A reaction to Slavin and Klein. *Educational Researcher*, *27*(7), 22–28.
- Pogrow, S. (1999). Rejoinder: Consistent large gains and high levels of achievement are the best measures of program quality: Pogrow responds to Slavin. *Educational Researcher*, *28*(8), 24–26.

- Pogrow, S. (2002). Success for All is a failure. *Phi Delta Kappan*, *83*(6), 463–468.
- Pressley, M., & El-Dinary, P. B. (1997). What we know about translating comprehension-strategies instruction research into practice. *Journal of Learning Disabilities*, *30*, 486–488.
- Slavin, R. E. (1999). Rejoinder: Yes, control groups are essential in program evaluation: A response to Pogrow. *Educational Researcher*, *28*(3), 36–38.
- Slavin, R. E., & Madden, N. (1999). Effects of bilingual and English as a Second Language adaptations for Success for All on the reading achievement of students acquiring English. *Journal of Education for Students Placed at Risk*, *4*, 393–416.
- Slavin, R. E., & Madden, N. A. (2001a, April). *Reducing the gap: Success for All and the achievement of African-American and Latino students*. Paper presented at the annual meeting of the American Educational Research Association, Seattle.
- Slavin, R. E., & Madden, N. A. (2001b). *Success for All and Comprehensive School Reform: Evidence-based policies for urban education*. Washington, DC: Office of Educational Research and Improvement.
- Slavin, R. E., Madden, N. A., Dolan, L., & Wasik, B. (1996). Success for All: A summary of research. *Journal of Education for Students Placed at Risk*, *1*, 41–76.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*. Newbury Park, CA: Sage.
- Stringfield, S., Millsap, M., & Herman, R. (1997). *Special strategies for educating disadvantaged children: Results and policy implications*. Washington, DC: U.S. Department of Education.
- Urdegar, S. (2000). *Evaluation of the Success for All program: 1998–1999*. Miami: Miami-Dade County Public schools, Office of Evaluation and Research.
- Walberg, H. J., & Greenberg, R. C. (1999). Educators should require evidence. *Phi Delta Kappan*, *80*, 132–135.
- Wells, L. R., Blendinger, J., & Greene, D. (2000, November). *An investigation of the Success for All reading program*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Bowling Green, KY.