

Multi-Site Comparison of the Effects of Success for All on Reading Achievement

Lana J. Smith
Steven M. Ross
Jason Casey

UNIVERSITY OF MEMPHIS

This study investigated the effects of the Success for All (SFA) program on elementary school students' reading achievement in 4 cities. The purpose was to conduct an independent examination of the program's effectiveness separate from the program's developers and from the school district in which it was first implemented. The research framework employed a pretest-posttest quasi-experimental design with nonequivalent but systematically matched groups. Results indicated greater achievement for students in SFA classes at 3 of the sites, especially for students achieving below the 25th percentile, relative to students in control classrooms at all 4 sites. However, SFA effects were not as strong and consistent as those obtained in the original studies. This research suggests that SFA can be implemented in sites geographically removed from the developers and apart from their direct supervision, but that continual monitoring and support of the program's quality is needed to ensure success over time.

HOW TO PREVENT THE ACADEMIC FAILURE of many of today's students is a critical question for both researchers and practitioners. Although we know more than we have ever known about how children learn, the basic skills they will need to keep up, and when to intervene to achieve maximum success, translating this knowledge from literacy research into successful practices that capitalize on the research base remains a difficult task. It was precisely this challenge of taking what was known from research and translating it into effective practice that led Robert Slavin and his colleagues at Johns Hopkins University to develop a program called Success for All (SFA), implement it in diverse contexts, and study its effects over time (Slavin, Madden, Karweit, Dolan, & Wasik, 1992).

Further understanding of SFA can be gained by comparing its orientation and strategies relative to those of Reading Recovery, another program that is perhaps more widely recognized by literacy researchers and has been extensively evaluated and employed nationally to assist children experiencing difficulty in reading (Pinnell, 1989). Reading Recovery includes a systematically designed set of procedures for helping the lowest achievers in a class become proficient and independent readers. Such procedures encompass specific strategies for teaching children and recommended reading materials, a staff development program, and administrative systems that coordinate the various program activities. The basic model involves using trained teachers to tutor first graders who are having difficulties learning to read for 30 minutes a day in addition to their regular reading instruction. When the child demonstrates the ability to perform satisfactorily with the regular class, the child is "discontinued" and another child is selected for the program.

A major common element of both Reading Recovery and Success For All is individual tutoring by certified teachers. Another element is the emphasis on developing metacognitive skills and "learning to read by reading." However, critical differences are that Reading Recovery requires more extensive training of tutors, serves first graders only, has longer tutoring sessions (30 minutes versus 20 minutes in SFA), and is less likely to select special students (e.g., those with learning disabilities or who speak English as a second language) for tutoring. Compared to Reading Recovery, SFA places much greater emphasis on integrating tutoring with the classroom curriculum, serves all students in the school (not just those being tutored), and supplements the reading program with family support.

Prior research on Reading Recovery has found fairly consistent positive effects at the end of first grade, with carryover into at least fourth grade (DeFord, Pinnell, Lyons, & Young, 1988; Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994). Recently, however, criticism has been directed to Reading Recovery for attending to a small number of low-achieving students when children with similar needs are not being served (Hiebert, 1994). Increasingly, Reading Recovery schools are responding to the need to supply classroom reading teachers with professional development to reinforce the strategies used in tutoring and to bring the instruction in the two settings into closer alignment.

By comparison, SFA was designed to be a kindergarten through third grade preventative and intensive intervention for students at risk of failing to learn to read. Key elements include individualized tutoring, regrouping across grades into homogeneous reading classes, smaller reading classes, family support, and a comprehensive reading program that progresses from specially designed and sequenced reading materials into basal readers or trade books. Instruction is imbedded in cooperative learning models including partner reading and Cooperative Integrated Reading and Composition (Stevens, Madden, Slavin, & Farnish, 1987).

Early evaluations of SFA demonstrated that the program could be highly successful in increasing reading achievement among very disadvantaged students (Madden, Slavin, Karweit, Dolan, & Wasik, 1993). The initial implementation of SFA at Abbottston Elementary in Baltimore indicated average effect sizes¹ (*ESS*) of .48 for the overall samples and .74 for the low-25% subsample (Slavin, Madden, Karweit, Livermon, & Dolan, 1990). In the second year, effects at Abbottston increased, while those at new schools were positive but relatively weaker (Madden, Slavin, Karweit, Dolan, & Wasik, 1990). Effects continued to increase at Abbottston and the new Baltimore schools in both the third (Madden et al., 1991) and fourth years (Madden et al., 1993). In the latter study, *ESS* across all reading measures averaged .51 in first grade, .60 in second grade, and .57 in third grade. For the low-25% subsample, the respective mean *ESS* were .82, 1.00, and .98.

Developers at Johns Hopkins have continued to monitor the progress of schools and districts that have implemented the program since its inception in 1987 at Abbottston Elementary, and results continue to support its effectiveness in a variety of contexts (Slavin, Madden, Dolan, et al., 1994). However, the overall assessment of impacts is heavily weighted by studies conducted by the developers or at neighboring sites.

Because SFA is lesser known than other reading programs and because the understanding of its components, strategies, and materials are important for interpreting the results of this study, a detailed description of the SFA program is provided.

The Success for All Model

The SFA program was developed on the basis of several assumptions regarding effective interventions for disadvantaged students (Slavin et al., 1990). One assumption was that early intervention to prevent learning problems from occurring would be more influential than remediating already established deficiencies.

1. Effect sizes are calculated by subtracting the control-group mean from the treatment-group mean and dividing by the control-group standard deviation.

Another was that reading and language arts are the most critical curriculum areas for applying special interventions, because the skills taught in these areas provide the foundation for learning in all other subjects. Finally, interventions needed to be comprehensive to include factors outside of the classroom that affect students' readiness, motivation, and opportunity to learn. Such factors include parental involvement, school attendance, and health needs.

Based on these assumptions, the SFA model was designed to include the following major elements: (a) one-on-one tutoring in reading by certified elementary teachers; (b) regrouping of students in grades 1 to 3 from heterogeneous, age-grouped classes into homogeneous, cross-grade ability groups during the language-arts period (approximately 90 minutes); (c) reduced class sizes in language arts as a result of using the reading tutors as reading teachers during that period; (d) frequent assessments (6 to 8 weeks) in which students are tested to determine reading progress and needs for tutoring and/or group changes; (e) a systematic reading program integrating Story Telling and Retelling (STaR), phonics in beginning reading, vocabulary building, comprehension skills, and cooperative learning; (f) a family support team to provide parenting education and assistance for students experiencing personal or health problems; and (g) a program facilitator who works in the school to coordinate and oversee the program. More detailed descriptions of major SFA program features are provided.

Reading Program

The reading program is based on effective practices and research in beginning reading that point to the need for students to learn to read in meaningful contexts while also receiving systematic presentation of word attack skills (Adams, 1990). The context for much of the instruction is provided through models of cooperative learning (Slavin, 1995; Stevens et al., 1987). In most SFA schools, a 90-minute period is scheduled for the reading program activities in all classes. To permit support subjects to be taught throughout the day, the primary grades (K-3) typically are scheduled for reading in one block and the intermediate grades (4-6) in another.

In all grades, the reading period starts by reading children's literature to the class and engaging students in a discussion of the story to enhance their comprehension, listening, speaking vocabulary, and knowledge of story structure. The development of basic language skills is emphasized in kindergarten and early first grade through the use of Story Telling and Retelling (STaR), which actively engages students in listening to, retelling, and dramatizing children's literature (Karweit, Coleman, Waclawiw, & Petza, 1990). Extensive use of big books contributes to the development of concepts of print, and receptive and expressive language is further supported by Peabody Language Development Kits (Dunn, Smith, Dunn, & Horton, 1989).

The K-1 reading program is called Reading Roots (formerly Beginning Reading) and is usually begun in the second semester of kindergarten. Students are introduced to letters and letter sounds through a series of active and engaging activities including (a) singing of an alphabet song to coax a puppet named Alphie to come out of his box and bring both a new letter of the day and a tongue-twister saying to reinforce the sound; (b) discriminating objects and pictures that do or do not begin with the sound using whole-class responses of pointing to “yes” or “no” signs posted on the wall; (c) orally repeating a couplet rhyme that guides children’s tracing of the letters on the backs of their classmates or in the air; (d) matching letters with pictures of words beginning with those letters; (e) blending sounds by stretching them out and then saying them quickly; and (f) performing other activities to help learn the discriminations in many ways using different sensory modalities.

The sounds and words practiced are then moved into a context of sentences and stories through specially designed curriculum materials called “shared stories” and “share sheets.” The stories are consumable minibooks containing some material written in small type to be read by the teacher and other material in large type to be read by students. Both the stories and the practice sheets are designed to reinforce the sounds, patterns, and words children have learned, as well as to engage them further in cooperative learning and in early writing. Students progress through the series of lessons leading to the second level of the program, Reading Wings.

Once students have reached the primer reading level, they transition into a form of cooperative learning (Cooperative Integrated Reading and Composition) guided by custom-designed questions and activities (called Treasure Hunts) using the school system’s adopted basal series or trade books. Students engage in partner reading and cooperative learning teams to master vocabulary, story content, and other reading and writing skills. The learning activities are built around story structure, prediction, summarization, vocabulary development, further decoding practice, and story-related writing. Students receive other types of direct instruction in comprehension skills (e.g., finding the main idea and understanding figurative language) and are assigned 20 minutes of homework each night involving reading books of their choice. They share their home readings each week through “Book Club” activities such as presentations, puppet shows, and other formats.

Individual Tutoring

Best-evidence research has indicated that one-to-one tutoring is the most effective intervention strategy for students with reading problems (Wasik & Slavin, 1990). In SFA, certified elementary teachers individually tutor, for 20 minutes each day, students who are having difficulty keeping up with their reading

groups. Priority is given to first graders, with older students served as space permits. An important difference between SFA and traditional Title I pull-out programs is that tutoring in SFA is directly integrated with the reading curriculum. This is accomplished by having the tutors teach a 90-minute reading class and then work one-on-one with children for the rest of the day on the lessons with which they are having the most difficulty. After students have participated in the 90-minute classroom reading instruction, they are pulled out of other subjects (never math or other language-arts blocks) for their individual tutoring. Methods and materials used in regular classroom instruction are essentially those used in tutoring with some adaptations for individual differences among students. A typical sequence of activity would include (a) rereading of familiar story, (b) quick drill over letters and/or words with which student is having difficulty, (c) reading a current story with tutor's guidance, and (d) story-related writing.

Regrouping

Students spend most of the school day in conventional heterogeneous aged-group classes. However, during the daily 90-minute reading period, they are regrouped according to reading performance level into smaller multi-age classes. The philosophy is to avoid the stigma and detrimental effects on learning of all-day tracking of between-class ability grouping (Oakes, 1992), while allowing teachers to adapt instruction to children's needs using whole-class activities and cooperative learning. Importantly, children are able to perform with success in these ability-grouped classes, while also spending most of the school day in regular heterogeneous classes where they can interact with children of different backgrounds and abilities. At 8-week (in some districts, 6-week) intervals, the teachers assess student progress with program-provided instruments covering vocabulary, content, and skills taught during the 8-week period. Using this information, they make needed adaptations in tutoring selections, group assignments, and students' programs.

Family Support Team

The family support team serves two major functions. One is to increase parents' involvement with the school and their children's learning activities to foster more positive school communities (Epstein, 1989). The second is to provide assistance to children when health or home problems interfere with their academic performance (e.g., needing glasses or not attending school regularly). This type of service integration between community and school increases the likelihood that families will make use of needed services and addresses problems that a school cannot solve on its own (Glass, McGaw, & Smith, 1981).

Facilitator

Success for All is coordinated by a part-time or full-time facilitator who works in a shared decision-making role with the principal and school-based management team. Training is provided for individuals by Johns Hopkins staff at yearly training conferences in Baltimore to support this type of leadership and effort in the development of a collaborative school culture – an element that commentators on school restructuring claim is essential (Lieberman & Miller, 1990). Among the facilitator's on-site responsibilities are planning the program, training teachers, assisting teachers who are having difficulty, helping the principal with scheduling, and overseeing the 8-week assessments and associated decisions about tutoring and reading group assignments.

Teacher Training

Training for SFA in the first year involves 3 days per teacher (usually during the summer) and approximately 2 to 4 additional in-service days during the school year. In addition, each class is visited intermittently by the school facilitator for the purposes of monitoring methods and providing feedback. Twice per year, trainers from Johns Hopkins University visit the school and observe most teachers' classes. After the first year, training is reinforced through regular in-services, an annual SFA conference, and implementation checks for the facilitators and John Hopkins University trainers.

Program Costs

The cost of SFA varies depending on the components selected. For a "full" model (the type used at the present sites), the cost is roughly \$800 to \$1,200 per student. However, these costs normally fit within allocated budgets for school-wide Title I programs. Without Title I (or other supplementary) support, the more costly elements of SFA (i.e., tutors and a full-time facilitator) would be difficult for many schools to fund, although they could still use the curriculum, regrouping, and restricted tutoring/facilitator services.

Purpose of the Study

Empirical support for the benefits of SFA is well established in the research literature (Madden et al., 1993; Slavin et al., 1992). The dissemination of that evidence, along with positive anecdotal reports by teachers and administrators at existing SFA sites, has spurred the rapid expansion of SFA to what is presently over 300 schools in districts and states across the United States. Although the evaluations

appear to have been carefully and rigorously performed, that they were conducted by the SFA development team at Johns Hopkins University created a natural interest in replication research.

The acceptability of SFA as a prototype model for at-risk prevention depends to a great extent, especially among the research community, on scrutiny by independent investigators and the extent to which it can be replicated in sites apart from its home base. This paper presents results and conclusions from 2 to 3 years of independent evaluations conducted in four cities beyond the original home of SFA in Baltimore. Accordingly, the present studies were commissioned by the SFA developers at Johns Hopkins University to test the generalizability of SFA benefits when the program is introduced into new settings outside the developers' home city and monitored over time by outside evaluators. Although the developers, through a grant from the Carnegie Foundation, funded the research, they had no involvement with its conduct and no special involvement with the particular sites beyond providing routine training services. Programs in four cities were examined: Memphis, Tennessee (1 school); Fort Wayne, Indiana (2 schools); Montgomery, Alabama (4 schools); and Caldwell, Idaho (2 schools). These programs were selected on the basis of providing diverse geographic and socioeconomic contexts, having available control schools, and expressing interest in participating. The data presented here include 3 years of implementation at the Memphis site (1990 to 1993) and two years at the other three sites (1991 to 1993).

Method

Research Design

A common research design was used for all four project sites. In general, the orientation is best described as quantitative evaluation research (Borg, Gall, & Gall, 1993), although some qualitative methods (surveys, interviews, observations) were selectively used at two sites to increase understanding of processes and outcomes. In this paper, we will focus on reporting and synthesizing the quantitative findings from assessments of reading performance.² The quantitative model employed a pretest-posttest quasi-experimental design with nonequivalent (but systematically matched) groups (Wiersma, 1995). Such designs have the disadvantage of being unable to randomly assign students to groups, but generally provide reasonable internal validity, given the use of pretesting, and are highly

2. The qualitative findings, obtained for Memphis and Fort Wayne only, are described in the technical reports by Ross and Smith (1991) for Memphis, and Ross, Smith, Casey, and Johnson (1993) for Fort Wayne.

practical for applied school research in which random assignments cannot be made. Use of a multi-site replicated experiment also provides reliability and generalizability checks that would not be present in a single-site study (see Slavin & Madden, 1993). Procedures and conditions specific to individual sites are described in the separate sections to follow.

Each SFA school was matched with a control school similar in poverty level (percentage of students qualifying for free lunch), historical achievement level, ethnicity, and other factors. At all sites, the control programs were “traditional,” basal-oriented reading approaches using within-class reading groupings. Student cohorts were formed by matching SFA subjects to control counterparts based on pretest scores on the Peabody Picture Vocabulary Test – Revised (PPVT). The rationale for using the PPVT was to test students’ general language abilities before they began participation in SFA and could be influenced by its strategies. This criterion involved testing students at the beginning of kindergarten, before they could read. In some instances where PPVT scores for kindergarten students were not available, random samples were selected from the kindergarten populations at the SFA and control schools.

Except where otherwise noted, the basic method used for the test comparisons was multivariate analysis of variance (MANOVA) performed on posttest raw scores. Overall significance across all test dimensions was evaluated via the multivariate F , and where multivariate significance was found, individual test significances were evaluated via univariate ANOVAS.

Participants

Participants were SFA and control-school primary-grade students ($K-2$) at four sites implementing SFA. The number of students involved by grade, site, and year is provided in Table 1. More specific descriptions of the samples are provided in the separate site-specific sections in Results. General characteristics are as follows: The Memphis site served a 100% African American population in a highly disadvantaged inner-city environment. Over 90% of the children received free or reduced lunch. The Montgomery schools were also inner-city with an African American population of 91% to 99% with over 70% receiving free or reduced lunch. In contrast, the Fort Wayne school served a racially mixed student population (about 60% Caucasian and 40% minority) that was less disadvantaged (about 65% on free or reduced lunch). Caldwell, Idaho served a rural population of mostly Caucasian students with about 15% Hispanics, of which about 45% were receiving free or reduced lunch.

Instruments

Following the Baltimore studies (e.g., Madden et al., 1993) and the evaluation model used by Slavin et al. (1990), reading/language scales were selected from

TABLE 1. Average Effect Sizes for the Overall and Low-25% Samples

Site	Year 1			Year 2			Year 3	
	K	1	2	K	1	2	1	2
<i>Memphis</i>								
Overall								
SFA <i>n</i>	20	45	47	35	26	17	30	15
Control <i>n</i>	23	45	47	35	68	17	33	43
Mean ES	0.97	0.17	0.00	0.08	0.90	0.14	0.38	0.51
Low-25%								
SFA <i>n</i>	—	11	11	9	7	4	7	5
Control <i>n</i>	—	11	11	9	18	5	8	11
Mean ES	— ^a	0.90	0.42	0.00	3.15	1.35	0.11	2.66
<i>Montgomery</i>								
Overall								
SFA <i>n</i>	129	94	—	—	71	—	—	—
Control <i>n</i>	129	94	—	—	96	—	—	—
Mean ES	0.00	0.11	—	—	1.27	— ^a	—	—
Low-25%								
SFA <i>n</i>	33	27	—	—	17	—	—	—
Control <i>n</i>	33	27	—	—	24	—	—	—
Mean ES	-0.08	0.15	—	—	2.28	— ^a	—	—
<i>Fort Wayne</i>								
Overall								
SFA <i>n</i>	73	68	—	—	25	36	—	—
Control <i>n</i>	73	68	—	—	25	36	—	—
Mean ES	0.38	0.47	—	—	0.51	0.44	—	—
Low-25%								
SFA <i>n</i>	19	19	—	—	7	12	—	—
Control <i>n</i>	19	19	—	—	7	12	—	—
Mean ES	0.56	0.60	—	—	0.79	0.79	—	—
<i>Caldwell</i>								
Overall								
SFA <i>n</i>	52	64	—	—	31	32	—	—
Control <i>n</i>	52	64	—	—	31	32	—	—
Mean ES	0.10	0.01	—	—	-0.21	0.12	—	—
Low-25%								
SFA <i>n</i>	14	19	—	—	9	8	—	—
Control <i>n</i>	14	19	—	—	9	8	—	—
Mean ES	1.59	-0.21	—	—	-0.52	0.33	—	—

^aNo sample tested for this population in this year.

SFA = Success for All, ES = effect size.

two nationally standardized reading test batteries. All scales were individually administered in sessions that generally lasted from 20 to 30 minutes. In previous research by the present authors, the scales were found to intercorrelate in the .3 to .6 range with first and second graders, with the lower correlations occurring between the oral reading measure (see below) and the four Woodcock scales; the higher correlations were among the latter four scales. The examiners were students from local colleges in each city. They were all pretrained and received on-site supervision by experienced testers. They had no knowledge about the experimental or control status of the schools and were rotated between schools on different days to avoid any testing bias due to examiner characteristics or abilities. The measurement instruments included:

Woodcock Reading Mastery Test – Revised. Four scales – Letter Identification, Word Identification, Word Attack, and Passage Comprehension (Woodcock, 1987) – measure student knowledge of the alphabet, recognition of common sight words, phonetic synthesis skills, and comprehension in context. Reliability data provided on these scales indicate median split-half reliability coefficients of .94 for Letter Identification, .98 for Word Identification, .94 for Word Attack, and .94 for Passage Comprehension (Woodcock, 1987, p. 99).

Durrell Analysis of Reading Difficulty. The Oral Reading scale presents a series of graded reading passages, which students read aloud, followed by comprehension questions. It was individually administered. The parallel forms reliability was reported by Durrell and Catterson (1980, p. 58) to be .85. Two additional language measures, the *Merrill Language Screening* (Mumm, Secord, & Dykstra, 1980) and the *Test of Language Development* (Newcomer & Hammill, 1988) were used in kindergarten in 1991–1992 only. These instruments were selected because they provide national norms for comparison of student achievement and reflect relative mastery of letter-sound correspondences, word knowledge in isolation and context, and comprehension of connected text in graded story passages.

Program Implementations

Depending on the resources available and particular interests, implementations of SFA may vary across sites with regard to the component strategies included and extensiveness of interventions (Slavin, Madden, Karweit, Dolan, & Wasik, 1994). The programs employed at the four sites examined in this study, however, were highly similar in the sense of being “full models” that implemented all components (e.g., full-time facilitator, one-to-one tutoring, family support, etc.) in the prescribed manner. As a function of participating in the research, each of the SFA schools was visited at least once yearly by Johns Hopkins University training experts and received systematic formative evaluations of their SFA programs. Although some had minor problems in particular years (e.g., a new teacher who was not well trained, too many or too few students in certain reading classes), all

were judged overall to be of good quality. The only more substantive problems that occurred during the period of the research were principal and major staff transfers at the Caldwell school (year 2), facilitator changes in Memphis (years 1 and 3), and undeveloped family support programs at all sites. Although the Caldwell and Fort Wayne programs did not receive school-wide Title I funding (Memphis and Montgomery did), the former schools managed through budget reallocations and creative staffing assignments to offer the full range of strategies. Similarities in philosophies and specific orientations to SFA were further engendered by the involvement of similar training teams (from Johns Hopkins University) in servicing the sites.

Aside from control schools being similar demographically and geographically to the SFA schools, the main criterion for their selection was that they offered “traditional” or “conventional” reading programs, the type that would likely be used at the SFA school had it not adopted the special program. During the year prior to the study, the control programs were all basal driven and delivered in heterogeneous classes, with extensive use of reading groups alternated with seatwork and some whole-class instruction. However, certain control schools initially or over time used certain strategies (e.g., tutoring or regrouping) that were comparable to key elements of SFA. When such practices were considered to be influential, they are discussed below with regard to research design considerations and possible impacts.

Results

Consistent with previous studies of SFA (Madden et al., 1993; Ross & Smith, 1994; Ross, Smith, Casey, & Slavin, 1995; Slavin et al., 1990), effect sizes (*ESS*) were computed to provide descriptive indices of the magnitude of SFA and control-group contrasts on the various outcome measures. Although inferences involving program effects were made on the basis of the MANOVA results, the *ESS*s were reported (even where MANOVA was nonsignificant) to provide useful descriptive information about the size and direction of each effect (see, e.g., rationale by Hedges, Laine, & Greenwald, 1994). As Slavin and Madden (1993) indicated, usage of *ESS* provides the advantage of being able to compare and aggregate results obtained from comparable multiple studies where individual studies may have lacked sufficient power to produce significant results. Results are also separately presented for the student subgroups in each grade who scored in the lowest 25% on the pretest (these students are also part of the “overall sample” analyses).

Summary results for Memphis and the other sites are presented in Table 1. The narrative descriptions following, however, highlight only significant results. Means and standard deviations of SFA and control groups are reported in Tables 2 (kindergarten) and 3 (grades 1 and 2) for significant comparisons only.

TABLE 2. Kindergarten Reading-Test Means and Standard Deviations for Significant MANOVA Comparisons between SFA and Control Students

Site		Letter identification	Word identification	TOLD	Merrill
<i>Memphis</i>					
Year 1 overall					
SFA	<i>M</i>	32.35	5.90	NA ^a	NA ^a
	<i>SD</i>	4.84	4.29		
Control	<i>M</i>	31.05	3.05	NA ^a	NA ^a
	<i>SD</i>	7.40	3.98		
<i>Fort Wayne</i>					
Year 1 overall					
SFA	<i>M</i>	32.43	10.26	15.86	4.45
	<i>SD</i>	4.28	9.82	6.51	0.87
Control	<i>M</i>	29.36	3.15	17.73	4.30
	<i>SD</i>	7.81	4.95	6.42	0.92
<i>Caldwell</i>					
Year 1 overall					
SFA	<i>M</i>	22.98	3.44	13.75	4.21
	<i>SD</i>	12.62	5.35	6.47	1.02
Control	<i>M</i>	25.35	3.48	12.23	4.06
	<i>SD</i>	10.22	8.84	7.31	1.21

^aThis test not given in year 1.

TOLD = Test of Language Development, SFA = Success for All.

Memphis, Tennessee

The SFA program was first implemented in grades K to 2 at an inner-city school in Memphis, Tennessee in the 1990–1991 school year (Ross & Smith, 1991; Ross & Smith, 1994). Thus, it has the longest history of the four cities and an additional year of data. To assess first- and second-grade outcomes, SFA students were individually matched to control students from a school in the same inner-city neighborhood. The two schools were nearly identical in the percentage of students receiving free lunch (over 90%) and in standardized achievement scores obtained during previous years. All SFA control student matches were based on student performance on total language skills scores and total reading scores on the Cali-

TABLE 3. First- and Second-Grade Reading-Test Means and Standard Deviations for Significant MANOVA Comparisons between SFA and Control Students

Site			Word identification	Word attack	Passage comprehension	Oral reading
<i>Memphis</i>						
Grade 1						
Year 2 overall						
	SFA	<i>M</i>	38.65	13.50	18.38	7.23
		<i>SD</i>	7.66	8.65	6.73	3.05
	Control	<i>M</i>	27.59	5.51	11.15	4.81
		<i>SD</i>	13.01	7.26	7.77	3.36
Year 2 low-25%						
	SFA	<i>M</i>	29.86	11.14	11.57	5.14
		<i>SD</i>	8.07	7.73	5.62	1.57
	Control	<i>M</i>	10.78	1.06	3.61	1.00
		<i>SD</i>	5.77	2.24	3.73	1.57
Grade 2						
Year 3 low-25%						
	SFA	<i>M</i>	45.80	7.40	24.40	12.40
		<i>SD</i>	1.30	3.98	1.67	2.61
	Control	<i>M</i>	27.18	1.46	12.09	4.73
		<i>SD</i>	7.10	2.30	5.05	2.57
<i>Montgomery</i>						
Grade 1 overall						
Year 2 overall						
	SFA	<i>M</i>	29.97	9.66	12.16	6.17
		<i>SD</i>	11.01	7.13	6.58	2.61
	Control	<i>M</i>	16.93	3.79	5.54	2.56
		<i>SD</i>	10.18	5.71	5.08	2.44
Grade 1/pair 1						
Year 2 overall						
	SFA	<i>M</i>	29.84	9.70	12.19	6.14
		<i>SD</i>	11.08	6.92	6.93	2.62
	Control	<i>M</i>	16.42	3.93	5.25	2.44
		<i>SD</i>	10.58	5.98	5.08	2.45
Grade 1/pair 2						
Year 2 overall						
	SFA	<i>M</i>	30.50	9.50	12.00	6.29
		<i>SD</i>	11.58	8.22	5.07	2.70
	Control	<i>M</i>	19.67	3.07	7.93	3.20
		<i>SD</i>	7.34	3.97	4.90	2.37
Grade 1						
Year 2 low-25%						
	SFA	<i>M</i>	14.29	5.24	6.88	2.65
		<i>SD</i>	8.40	4.35	4.27	1.69
	Control	<i>M</i>	5.83	0.75	1.79	0.92
		<i>SD</i>	3.64	2.01	1.77	1.02
<i>Fort Wayne</i>						
Grade 1						
Year 1 Overall						

TABLE 3. CONTINUED

<i>Site</i>		<i>Word identification</i>	<i>Word attack</i>	<i>Passage comprehension</i>	<i>Oral reading</i>
SFA	<i>M</i>	35.04	12.60	16.37	6.74
	<i>SD</i>	10.63	7.43	8.07	4.25
Control	<i>M</i>	28.00	7.90	13.91	4.68
	<i>SD</i>	14.70	7.91	9.31	3.83
Year 1 low-25%					
SFA	<i>M</i>	28.16	9.05	9.84	5.05
	<i>SD</i>	10.02	5.37	6.18	3.29
Control	<i>M</i>	18.53	4.68	8.11	2.84
	<i>SD</i>	12.78	5.76	7.13	3.35
Year 2 overall					
SFA	<i>M</i>	39.88	17.08	19.24	7.68
	<i>SD</i>	13.07	8.10	7.13	3.30
Control	<i>M</i>	30.60	10.76	17.12	6.32
	<i>SD</i>	11.57	9.51	8.36	4.11
Grade 2					
Year 2 overall					
SFA	<i>M</i>	51.06	20.39	27.44	12.78
	<i>SD</i>	9.67	9.44	6.81	5.16
Control	<i>M</i>	41.92	15.53	23.78	11.11
	<i>SD</i>	14.64	9.12	10.10	7.04
<i>Caldwell</i>					
Grade 1					
Year 1 overall					
SFA	<i>M</i>	35.30	15.45	19.19	6.31
	<i>SD</i>	12.39	8.15	13.41	3.36
Control	<i>M</i>	37.77	13.88	17.27	6.98
	<i>SD</i>	10.28	7.37	7.87	3.68
Year 1 low-25%					
SFA	<i>M</i>	28.26	12.21	13.21	5.11
	<i>SD</i>	11.49	6.06	7.53	3.00
Control	<i>M</i>	35.05	10.11	12.90	5.84
	<i>SD</i>	8.18	6.15	6.03	1.95
Year 2 overall					
SFA	<i>M</i>	22.68	11.52	12.06	3.68
	<i>SD</i>	14.36	7.72	8.16	3.02
Control	<i>M</i>	29.42	10.90	12.71	5.32
	<i>SD</i>	13.27	8.27	8.72	4.78
Year 2 low-25%					
SFA	<i>M</i>	15.78	6.00	9.44	2.44
	<i>SD</i>	13.66	4.03	7.92	3.43
Control	<i>M</i>	29.22	9.56	12.44	5.33
	<i>SD</i>	17.03	8.68	10.38	5.00
Grade 2					
Year 2 overall					
SFA	<i>M</i>	53.94	24.44	29.19	14.19
	<i>SD</i>	8.71	6.76	6.21	6.63
Control	<i>M</i>	54.4	21.81	26.28	15.38
	<i>SD</i>	78.10	8.66	7.39	7.46

SFA = Success for All.

ifornia Achievement Test (CAT), a state-mandated assessment administered to district students each spring semester. At the kindergarten level, no prior achievement test scores were available for use in matching. Therefore, random samples were selected from the kindergarten populations at the SFA and control schools.

Kindergarten

Dependent variables for the kindergarten analysis in year 1 were Letter Identification, Word Identification, and Oral Reading. Results (SFA $n = 20$, control $n = 22$) were highly consistent with pronounced SFA effects. For the overall sample in year 1, the dependent-samples MANOVA was significant, $F(3, 38) = 3.20$, $p = .03$. Univariate tests showed significant SFA advantages on Word Identification only, $F(1, 40) = 5.00$, $p = .03$, ($ES = .72$). A significantly higher percentage of SFA than control students (35% versus 3%) successfully read the initial paragraph on the Oral Reading test, $\chi^2(1) = 5.05$, $p < .05$. Kindergarten outcomes, however, were not significant in year 2. In year 3, kindergarten outcomes were not assessed, based on our feeling that the kindergarten program comparisons were highly influenced by the amount of emphasis given by the school's curriculum (regardless of program) to teaching reading prior to first grade.

First Grade

Posttests for first grade in year 1 were the same as for kindergarten, but the MANOVA performed on the overall sample was not significant ($p = .19$), whereas the overall MANOVA for students who had scored in the lowest 25% in reading on the previous-year CAT test only approached significance ($p = .08$).

For posttesting in year 2, the Letter Identification test was replaced by the Passage Comprehension test on the basis of the former appearing to be too easy and not discriminating enough for students at this level. The MANOVA for first grade (second year in SFA) was significant, $F(4, 42) = 3.10$, $p = .03$. Effect sizes were larger than those obtained for kindergarten, averaging .90 across all tests. The two largest effects were 1.10 for Word Attack and .93 for Passage Comprehension. Univariate comparisons, using ANOVA, showed significant advantages for SFA on Word Identification, $F(1, 45) = 5.22$, $p = .03$; Word Attack, $F(1, 45) = 10.66$, $p = .01$; and Passage Comprehension, $F(1, 45) = 6.44$, $p = .02$. Oral Reading advantages approached significance, $F(1, 45) = 3.84$, $p = .06$.

Parallel analyses conducted for the low-25% subsamples (SFA $n = 7$, control $n = 18$), $F(4, 9) = 5.06$, $p = .02$, also indicated extremely high ESS, averaging 3.15 overall. Significant univariate effects, all favoring the SFA group, were obtained for Word Identification, $F(1, 12) = 16.45$, $p < .01$; Word Attack, $F(1, 12) = 14.79$, $p < .01$; and Oral Reading, $F(1, 12) = 12.24$, $p < .01$.

In year 3, results for first grade (SFA $n = 30$, control $n = 33$) indicated a directional SFA program advantage on all of the tests, but the overall MANOVA was not significant ($p = .09$), nor was the MANOVA significant for the low-25% subsample.

Second Grade

At the second-grade level, neither the overall sample nor low-25% subsample MANOVAs were significant for years 1 and 2. In year 3, the MANOVA effect for second graders (SFA $n = 15$, control $n = 43$) approached significance, $F(4, 53) = 2.43$, $p = .06$. The overall mean ES was $.51$. Results for the low-25% subsample were more striking, $F(4, 11) = 9.53$, $p < .001$, with significant group differences (all $ps < .01$) favoring SFA on all tests: Word Identification ($ES = 2.62$), $F(1, 14) = 32.68$; Word Attack ($ES = 2.59$), $F(1, 14) = 14.16$; Passage Comprehension ($ES = 2.44$), $F(1, 14) = 27.40$; and Oral Reading ($ES = 2.98$), $F(1, 14) = 30.31$. The mean ES was 2.66 .

Montgomery, Alabama

Success for All was implemented at four elementary schools in Montgomery in 1991–1992. Both the program and evaluation procedures were fundamentally the same as those described for the Memphis program. Each school was matched to a comparable school having similar demographic characteristics and prior achievement on standardized tests. During year 1, the evaluation encompassed all four schools at grade K but only two schools at grade 1. The reason for excluding two schools from the grade 1 evaluation was that their first graders had been in SFA as kindergartners in the previous year and thus could no longer be pre-tested to achieve matches with control counterparts without possible contamination from SFA effects.

Kindergarten

In year 1, the matched MANOVA design at the kindergarten level consisted of four school pairs (Pair 1 $n = 39$, Pair 2 $n = 34$, Pair 3 $n = 19$, Pair 4 $n = 48$). The overall MANOVA was not significant for either the overall sample or for the low-25% subsample. Kindergarteners were not assessed in year 2.

First grade

Year 1 results were not significant for either the overall sample or the low-25%

subsample. A follow-up descriptive study of the control schools (see Bond, Ross, & Smith, 1993) indicated that these schools (which received curriculum guidance from the same Title I coordinator who helped to bring SFA to the district) were using many of the essential SFA program elements, such as tutoring, re-grouping, reduced class sizes, and language development programs (e.g., STaR and Peabody) in beginning reading.

Based on these findings, the two original SFA schools that lacked suitable control schools were eliminated from the evaluation in year 2. Two other SFA schools, matched to control schools that were predominately using traditional reading curricula and support strategies, were maintained (to be called Pairs 1 and 2 below). Only first grade was examined because of the establishment of new pairings. Success for All Control Pair 2, however, included 14 surviving matched subjects who had been pretested in kindergarten. Pair 1 (SFA $n = 57$, control $n = 82$) showed significant program differences, $F(4, 133) = 17.90, p = .001$, and large, significant effects (all $ps < .01$) favoring SFA students on all of the tests (mean $ES = 1.28$): Word Identification, $F(1, 136) = 51.80$; Word Attack, $F(1, 36) = 27.36$; Passage Comprehension, $F(1, 136) = 46.12$; and Oral Reading, $F(1, 136) = 72.03$. For Pair 2, the MANOVA $F(4, 24) = 3.08, p < .04$. Success for All advantages were significant (all $ps < .01$) on all tests: Word Identification, $F(1, 27) = 9.19$; Word Attack, $F(1, 27) = 7.37$; Passage Comprehension, $F(1, 27) = 6.91$; and Oral Reading, $F(1, 27) = 10.74$. For the low-25% subsamples, the overall MANOVA yielded $F(4, 36) = 7.50, p < .01$. Significant univariate effects, both favoring SFA, were obtained on Word Identification, $F(1, 17) = 4.28, p < .05$, and on Word Attack, $F(1, 17) = 4.54, p < .05$. The mean ES , however, for the combined Pairs 1 and 2 low-25% subsamples was 2.28 (see Table 1).

Fort Wayne, Indiana

Two elementary schools in Fort Wayne, Indiana implemented SFA in grades K to 6. Unlike Memphis and Montgomery, these schools were racially mixed and located in lower- to middle-income neighborhoods. Kindergarteners and first graders were pretested and posttested in the first year and then as first and second graders during the second year using the same instruments as the Memphis and Montgomery sites.

Kindergarten

The independent-samples MANOVA performed on the kindergarten overall sample ($n = 73$ pairs) yielded a significant program effect, $F(5, 67) = 2.30, p = .04$. Univariate tests showed that SFA students were superior to control students on the Merrill test only, $F(1, 71) = 7.86, p = .01$. The overall mean ES across all tests was .38 (see Table 1). For the low-25% subsamples, the MANOVA result only approached significance, $F(5, 14) = 2.65, p = .069$.

First Grade

The first-grade MANOVA results ($n = 68$ pairs) indicated a significant program effect, $F(4, 63) = 6.45, p < .01$ (mean $ES = .47$). Univariate tests showed the SFA schools to be superior to the control schools on Word Identification, $F(1, 66) = 10.84, p < .001$; Word Attack, $F(1, 66) = 18.08, p < .01$; and Oral Reading, $F(1, 66) = 9.93, p < .01$. The SFA advantage on Passage Comprehension approached significance ($p = .052$). Results for the low-25% sample ($n = 19$ pairs) also significantly favored SFA, $F(4, 13) = 3.73, p < .05$ (mean $ES = .60$). Two univariate tests, both significantly favoring SFA, were on Word Identification, $F(1, 17) = 4.28, p < .05$, and Word Attack, $F(1, 17) = 4.54, p < .05$. In year 2, however, program effects were not significant for either the overall or low-25% samples.

Second Grade

The MANOVA on second grade conducted in year 2 revealed a significant effect favoring SFA, $F(4, 31) = 3.40, p = .02$ (mean $ES = .44$). Again, SFA students performed better than control students, but significant effects occurred on Word Identification ($ES = .62$) only, $F(1, 34) = 8.20, p < .01$.

Caldwell, Idaho

Success for All was implemented in grades K and 1 of one elementary school during 1991–1992 in the Caldwell, Idaho community. In 1992–1993, the Caldwell district opened a new school and moved the principal and part of the staff from the original SFA implementation, making both schools SFA schools. These two sites were the first rural SFA schools to be evaluated. All evaluation procedures and instruments were consistent with the other three sites.

Kindergarten

The kindergarten research design included the core tests used in other cities, but added the PPVT as a posttest to determine whether differences in language vocabulary occurred at the end of the year. Kindergarten outcomes were generally positive, although slight negative effects occurred on Letter and Word Identification. The MANOVA yielded $F(5, 47) = 2.63, p < .05$, with an average ES of .10. Univariate analyses revealed a significant difference only on the PPVT posttest, $F(1, 51) = 8.52, p < .01$ ($ES = .41$). The analyses of the low-25% sample of grade K ($n = 14$ pairs) revealed no significant differences on either the overall MANOVA or the univariate analyses (mean $ES = 1.59$). The small number of pairs obviously limited the power of the formal analyses.

First Grade

For grade 1 in year 1 ($n = 64$ pairs), the MANOVA was significant, $F(5, 59) = 2.69$, $p < .05$ (mean $ES = .01$). Follow-up univariate analyses, however, yielded no significant effects on any test. Results from the low-25% sample ($n = 19$ pairs) were negative, with a mean ES of $-.21$, $F(5, 14) = 4.53$, $p < .05$. Univariate tests favored the control sample on Word Identification only ($ES = -.83$), $F(1, 18) = 6.48$, $p < .05$.

In year 2, the MANOVA for first graders ($n = 31$ pairs) was highly significant, $F(4, 27) = 12.50$, $p < .01$ (mean $ES = -.21$). Univariate analyses indicated that control students achieved significantly better than SFA students on Word Identification ($ES = -.51$), $F(1, 30) = 5.51$, $p < .05$. For low-25% first graders ($n = 9$ pairs) the overall MANOVA was also significant, $F(4, 5) = 9.80$, $p = .01$ (mean $ES = -.52$). Univariate results indicated significant advantages for the control students on Word Identification ($ES = -.79$), $F(1, 8) = 10.01$, $p < .01$; and Oral Reading ($ES = -.58$), $F(1, 8) = 9.26$, $p < .05$.

Second Grade

The MANOVA for second grade ($n = 32$ pairs), conducted in year 2, was highly significant, $F(4, 28) = 3.24$, $p = .03$, revealing overall advantages for SFA (mean $ES = .12$). Success for All students performed significantly better than control students on Passage Comprehension ($ES = .39$), $F(1, 31) = 4.76$, $p < .05$. The MANOVA revealed no significant results for the low-25% subsample ($n = 8$ pairs), $F(4, 4) = 2.18$, $p = .24$ (mean $ES = .33$).

Overall Comparison of Four Sites

Kindergarten ES s for each site's first year of implementation ranged from .97 in Memphis to .00 in Montgomery (see Table 1). It is interesting to note that for Memphis, the only site where kindergarten results were collected for 2 years, there was great disparity in results from the first year to the second (.97 ES to .08 ES). During the first year of implementation, the SFA program's more academic emphasis represented a fairly substantial deviation from the traditional kindergarten curriculum taught at the control school. However, during the second year, the control school had shifted its emphasis to a more academic curriculum including direct teaching of alphabetic and sound principles in addition to acquiring a Writing to Read computer program in which all kindergarten children participated. Effect sizes for the low-25% subgroups are generally higher than those for the overall sample and ranged from a low of $-.08$ at the Montgomery site to a substantial 1.59 in Caldwell.

Our results, overall, are equivocal regarding SFA's effects in kindergarten

when the control school is also employing a quality academically focused program (e.g., the school noted previously in year 2). Fluctuations in SFA effects within or across sites would be least surprising at the kindergarten level given several factors. First, the teaching of actual reading skills to SFA children does not occur until the second half of the year, with the introduction of the Reading Roots program. Second, outcomes from the SFA control school comparison are largely dependent on the degree of emphasis the control school places on academic learning objectives as opposed to developmental concerns. Regardless of the quality of the curriculum, high concentration on reading instruction is likely to produce observable success on posttest measures assessing beginning skills. For example, in evaluating a phonics-based program, Bond et al. (1992) found very high benefits in kindergarten, but smaller effects in higher grades.

Outcomes at the first-grade level are of key importance in view of SFA's primary focus of giving all students a strong start in reading. Of particular note are the second year results at each site. It is during the second year of implementation, when teachers have had a year's experience with the program and have received children from kindergarten with the early SFA experiences in language, storytelling, and graphophonic principles, that the true impact of the program can be seen. With the exception of Caldwell, all sites realized increased benefits of program during the second year of implementation, with *ESS* ranging from .51 to 1.27 for the overall sample. Similar to the findings for kindergarten, for the low-25% first graders the *ESS* were even stronger, ranging from .79 to 3.15. These larger effects are consistent with SFA's philosophy of providing all possible interventions to help children most at risk of failure (Slavin et al., 1990). Accordingly, the low-25% achievers tend to be the primary recipients of tutoring and family support services.

Given the above interpretations, however, an anomalous outcome would appear to be the drop in both the overall and low-25% sample *ESS* for the Memphis site in the third year. Although cross-site results suggest that SFA is fairly robust, despite site-specific variations in its implementation, certain elements seem likely to have a greater impact than others on program effects. In this sense, it is significant that the Memphis program's veteran facilitator retired at the end of the second year and was not replaced for the third. In the absence of a facilitator, program fidelity and efficacy would be expected to decrease over time.

Results for the overall second-grade samples were fairly strong in Memphis ($ES = .51$) and Fort Wayne ($ES = .44$) but weaker in Caldwell ($ES = .12$). For the low-25% sample, the strongest effects occurred in Memphis ($ES = 2.66$) and Fort Wayne ($ES = .79$), with more moderate results in Caldwell ($ES = .33$).

Discussion

The purpose of this study was to examine the effects of the SFA program on

reading achievement in locations apart from the school district in which it was developed and first implemented. Results for Memphis, Montgomery, and Fort Wayne are generally supportive of SFA benefits for students in all grade levels, with stronger effects for the lowest performing students. Some inconsistency of results and variance within and between sites suggests that it may be unrealistic to expect outcomes to be identical across all applications. Programs differ in teacher training, school conditions, and types of students. Another key factor affecting evaluation outcomes is the appropriateness of the control programs used for evaluation comparisons. In basic research, it is reasonable to establish a pure control condition that is identical to the experimental condition, but does not receive the “treatment.” In contrast, in applied research with schools, every school has a treatment in the form of the curriculum program being implemented during the year. As our longitudinal research has shown, such programs continually evolve with staff changes and exposure to new ideas, and even more critically, may even become influenced within a district by the experimental program (in this case, SFA). This latter situation occurred in Montgomery and compromised the original SFA control school comparison.

In the case of Caldwell, a different control school problem occurred – the absence in a small district of an available control school that was highly similar to the SFA school in neighborhood location and types of students served. Outcomes at the Idaho site have been the most inconsistent of all site comparisons and are particularly puzzling. Observations of the SFA implementation by Johns Hopkins University trainers indicated that the school was implementing the program with high fidelity. Given that the control school is a high-quality school that provided special reading programs, and that it is located in a more suburban, less rural area of the region, it is reasonable to question the validity of the SFA control comparison. The results for the overall second-grade sample, and especially the second-grade low-25% subsample, however, do reflect the general pattern of findings at other sites and suggest the cumulative benefits of SFA over time. That is, the longer a lower achiever is in SFA, the greater the effects.

Other types of dynamics from school to school probably had additional extraneous effects. In some instances, a clear competitive spirit naturally arose between school staffs and principals of the matched schools as well as between schools that were starting up new programs at the same time, as in Fort Wayne and Montgomery. Principal and major staff transfers, as in Caldwell and to some degree in Fort Wayne, and facilitator changes (Memphis, years 1 and 3; Caldwell, year 2), could conceivably have had major impact on program implementations. Other seemingly influential differences from one site to another were inner-city (Memphis and Montgomery), urban (Fort Wayne), and rural (Caldwell) environments; and background characteristics of students (100% African American and highly disadvantaged in Memphis and Montgomery; less disadvantaged and more culturally diverse in Fort Wayne and Caldwell).

Based on these considerations and from the achievement results of this

study, it seems fair to infer that SFA can typically be replicated successfully at distant locations from, and with limited monitoring by, the program developers. Success for All contains the characteristics most often reflected in exemplary reading programs including systematic procedures, comprehensive intervention strategies, frequent diagnosis, facilitator leadership, monitoring components, and extensive and ongoing staff development (Samuels, 1981). These elements appear to be consistent with other literacy interventions (e.g., Reading Recovery) that have been associated with high levels of literacy attainment (Hiebert, 1994).

Overall, of the many multivariate comparisons ($f = 31$) across grade levels, subsample types, and sites reported in this paper, close to half ($f = 14$) significantly favored SFA, whereas only three (all in Caldwell) indicated advantages for the control program. Yet, these results are not as strong and consistent as those obtained in the original series of SFA evaluations performed by the developers (Madden et al., 1993; Slavin et al., 1990). Variations in implementation quality and particularly in achievement outcomes both within and across sites suggest that expectations regarding SFA's potential to enhance reading achievement at a given school must be realistic. As with any other comprehensive educational program, the various elements must be effectively operationalized and integrated, teachers must be well trained, and principals and school staffs must be supportive of the program philosophy and invested in ensuring its quality or "high reliability" (Stringfield, 1993). A key challenge in this regard is maintaining the integrity and fidelity of strategies as school administrators and teachers and district policies change over time. As SFA expands nationally, the attention that individual schools and districts can receive from the developers will decrease. The associated need will be stronger local monitoring and quality control of the programs. This area is already being addressed by the developers through the establishment of regional training centers and the placements of local expert trainers, funded by the school districts, to work with the individual schools.

Although measurements used in these studies were individually administered tests that are sensitive to all aspects of reading (comprehension, fluency, word attack, and word identification), there may be broader definitions of literacy that would be beneficial to examine in programs such as SFA and Reading Recovery (e.g., listening comprehension, oral language development, writing ability, and student perception of self as reader). In addition, other studies investigating the effects of the individual components of the program on literacy acquisition and reading achievement will help schools determine appropriate choices for implementation, especially when resources are limited and a full model would not be feasible for schools to implement.

Research that continues to explore the effectiveness of programs across widely diverse elementary-school contexts and over extended time is crucial for restructuring initiatives. Increasing schools' abilities to ensure that disadvantaged children will have opportunity to learn and be successful in school hopefully is the inevitable result.

References

- Adams, M.J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Bond, C., Ross, S., Smith, L., Nunnery, J., Goldstein, R., & Bowie, R. (1992). *Analysis of the impact of Sing, Spell, Read, and Write on reading/language arts achievement of primary grade children*. Memphis, TN: Memphis State University, Center for Research in Educational Policy.
- Bond, C.L., Ross, S.M., & Smith, L.J. (1993). *Montgomery, Alabama Success for All program control school visitation study*. Memphis, TN: Memphis State University, Center for Research in Educational Policy.
- Borg, W.R., Gall, J.P., & Gall, M.D. (1993). *Applying educational research* (3rd ed.). White Plains, NY: Longman.
- DeFord, D., Pinnell, G.S., Lyons, C., & Young, P. (1988). *Reading Recovery: Vol. 9. Report of the follow-up studies*. Columbus, OH: Ohio State University.
- Dunn, L.M., Smith, J.O., Dunn, L.M., & Horton, K.B. (1989). *Peabody language development kits*. Circle Pines, MN: American Guidance Service.
- Durrell, D., & Catterson, J. (1980). *Durrell analysis of reading difficulty*. New York: Psychological Corporation.
- Epstein, J.L. (1989). Effects of teacher practices of parent involvement on student achievement in reading and math. In S. Silvern (Ed.), *Literacy through family, community, and school interaction*. Greenwich, CT: JAI Press.
- Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Hedges, L.V., Laine, R., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23, 5-9.
- Hiebert, E.H. (1994). Reading Recovery in the United States: What difference does it make to an age cohort? *Educational Researcher*, 23, 15-25.
- Karweit, N.L., Coleman, M.A., Waclawiw, I., & Petza, R. (1990). *Story Telling and Retelling (STaR): Teacher's manual*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.
- Lieberman, A., & Miller, L. (1990). Restructuring schools: What matters and what works. *Phi Delta Kappan*, 71, 759-764.
- Madden, N.A., Slavin, R.E., Karweit, N.L., Dolan, L.J., & Wasik, B.A. (1990, April). *Success for All: Effects of variations in duration and resources of a schoolwide elementary restructuring program*. Paper presented at the meeting of the American Educational Research Association, Boston.
- Madden, N.A., Slavin, R.E., Karweit, N.L., Dolan, L.J., & Wasik, B.A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal*, 30, 123-148.
- Madden, N.A., Slavin, R.E., Karweit, N.L., Dolan, L.J., Wasik, B.A., Shaw, A., Mainzer, K.L., Leighton, M., Petza, R., & Haxby, B. (1991, April). *Success For All: Third-year outcomes*. Paper presented at the meeting of the American Educational Research Association, Chicago.
- Mumm, N., Secord, W., & Dykstra, K. (1980). *Merrill language screening test*. New York: Psychological Corporation.
- Newcomer, P. & Hammill, D. (1988). *Test of language development*. Austin, TX: Pro-Ed.
- Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher*, 21(4), 12-21.
- Pinnell, G.S. (1989). Reading Recovery: Helping at-risk children learn to read. *Elementary School Journal*, 90, 161-182.

- Pinnell, G.S., Lyons, C.A., DeFord, D.E., Bryk, A.S., & Seltzer, M. (1994). Studying the effectiveness of early intervention approaches for first-grade children having difficulty in reading. *Reading Research Quarterly*, 29, 8-39.
- Ross, S.M., & Smith, L.J. (1991). *Final report: 1990-1991 Success for All program in Memphis*. Memphis, TN: Memphis State University, Center for Research in Educational Policy.
- Ross, S.M., & Smith, L.J. (1994). The impact of using the Success for All model on reading achievement, teacher adjustment, and restructuring at an inner-city school. *Elementary School Journal*, 95, 121-138.
- Ross, S.M., Smith, L.J., Casey, J., & Johnson, B. (1993). *Final reports 1992-93: Success for All program in Fort Wayne, Indiana*. Memphis, TN: Memphis State University, Center for Research in Educational Policy.
- Ross, S.M., Smith, L.J., Casey, J., & Slavin, R.E. (1995). Increasing the academic success of disadvantaged children: An examination of alternative early intervention programs. *American Educational Research Journal*, 32, 773-800.
- Samuels, S.J. (1981). Characteristics of exemplary reading programs. In J.T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews*. Newark, DE: International Reading Association.
- Slavin, R.E. (1995). *Cooperative learning: Theory, research, and practice* (2nd. ed.). Boston: Allyn & Bacon.
- Slavin, R.E., & Madden, N.A. (1993, April). *Multi-site replicated experiments: An application to Success for All*. Paper presented at the meeting of the American Educational Research Association, Atlanta, GA.
- Slavin, R.E., Madden, N.A., Dolan, L.J., Wasik, B.A., Ross, S., & Smith, L. (1994). Whenever and wherever we choose: The replication of Success for All. *Phi Delta Kappan*, 75, 639-647.
- Slavin, R.E., Madden, N.A., Karweit, N.L., Dolan, L.J., & Wasik, B.A. (1992). *Success for All: A relentless approach to prevention and early intervention in elementary schools*. Arlington, VA: Educational Research Service.
- Slavin, R.E., Madden, N.A., Karweit, N.L., Dolan, L.J., & Wasik, B.A. (1994). Success For All: A comprehensive approach to prevention and early intervention. In R.E. Slavin, N.L. Karweit, & B.A. Wasik (Eds.), *Preventing early school failure*. Needham Heights, MA: Allyn & Bacon.
- Slavin, R.E., Madden, N.A., Karweit, N.L., Livermon, B.J., & Dolan, L. (1990). Success for All: First-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal*, 27, 255-278.
- Stevens, R.J., Madden, N.A., Slavin, R.E., & Farnish, A.M. (1987). Cooperative Integrated Reading and Composition: Two field experiments. *Reading Research Quarterly*, 22, 433-454.
- Stringfield, S. (1993, November). *Attempts to enhance students' learning: A search for valid programs and highly reliable implementation techniques*. Paper presented at the meeting of the American Evaluators Association, Dallas, TX.
- Wasik, B.A., & Slavin, R.E. (1990, April). *Preventing early reading failure with one-to-one tutoring: A best-evidence synthesis*. Paper presented at the meeting of the American Educational Research Association, Boston.
- Wiersma, W. (1995). *Research methods in education: An introduction* (6th ed.). Needham Heights, MA: Allyn & Bacon.
- Woodcock, R.W. (1987). *Woodcock reading mastery tests-revised*. Circle Pines, MN: American Guidance Service.

Manuscript received: April 28, 1995

First revision requested: June 26, 1995

Final revision received: February 7, 1996

Accepted for publication: February 12, 1996

J L R

Journal of Literacy Research
Copyright © 1996 National Reading Conference Inc.
All rights reserved. Printed in the USA.

N R C