# Research on Achievement Outcomes Of Success for All: A Summary and Response to Critics

**Robert E. Slavin**
**Johns Hopkins University**

**Nancy A. Madden**
**Success for All Foundation**

This article was written under funding from the Office of Educational Research and Improvement (OERI), U.S. Department of Education (R-117-40005). Additional funding for the development and evaluation of Success for All and Roots & Wings has been received from New American Schools, the Carnegie Corporation of New York, the Sandler Family Foundation, the Stupski Family Foundation, the Charles A. Dana Foundation, the Pew Charitable Trusts, and the Abell Foundation. However, any opinions expressed are those of the authors and do not necessarily represent the positions or policies of our funders.

For additional articles and other information on Success for All, see our web site, www.successforall.net, or call 1-800-548-4998. Success for All is disseminated by the Success for All Foundation, a nonprofit organization in Baltimore, Maryland.

Success for All is one of the greatest success stories of educational research and reform. Starting in one Baltimore school in 1987, Success for All will be, by September 2000, in more than 1800 schools, serving about a million children. Research on Success for All has involved schools in many districts, including Baltimore, Memphis, Philadelphia, Miami, Houston, Montgomery (AL), Ft. Wayne (IN), Little Rock (AR), Tucson (AZ), Riverside (CA), Modesto (CA), Caldwell (ID), Clarke Co. (GA), St. Mary's Co. (MD), and others, and this research has been carried out by many researchers in several research institutions (see Slavin & Madden, in press; Slavin & Madden, 1999a). Research on Success for All and related designs has taken place in England, Australia, Canada, Mexico, and Israel. All of this research has compared Success for All to matched control schools on standardized measures of reading, both individually-administered and group-administered measures. Not every study has found positive outcomes, but the great majority have, especially when program implementation has been adequate. The studies themselves have been published in some of the most rigorous journals in education. Reviews of research on comprehensive reform designs by the American Institutes of Research (Herman, 1999), and the Thomas Fordham Foundation (Traub, 1999) have identified Success for All as one of two elementary programs (the other is Direct Instruction) with strong, frequently replicated evidence of effectiveness for student achievement.

**Major Elements of Success for All**

Success for All is a schoolwide program for students in grades pre-K to five which organizes resources to attempt to ensure that virtually every student will reach the third grade on time with adequate basic skills and build on this basis throughout the elementary grades, that no student will be allowed to "fall between the cracks." The main elements of the program are as follows:

***Tutors.*** In grades 1-3, specially trained certified teachers and paraprofessionals work one-to-one with any students who are failing to keep up with their classmates in reading. Tutorial instruction is closely coordinated with regular classroom instruction. It takes place 20 minutes daily during times other than reading periods.

***A Schoolwide Curriculum.*** During reading periods, students are regrouped across age lines so that each reading class contains students all at one reading level. Use of tutors as reading teachers during reading time reduces the size of most reading classes to about 20. The reading program in grades K-1 emphasizes language and comprehension skills, phonics, sound blending, and use of shared stories that students read to one another in pairs. The shared stories combine teacher-read material with phonetically regular student material to teach decoding and comprehension in the context of meaningful, engaging stories. In grades 2-6, students use novels or basals but not workbooks. This program emphasizes cooperative learning activities built around partner reading, identification of characters, settings, problems, and problem solutions in narratives, story summarization, writing, and direct instruction in reading comprehension skills. At all levels, students are required to read books of their own choice for twenty minutes at home each evening. Classroom libraries of trade books are provided for this purpose. Cooperative learning programs in writing/language arts are used in grades K-6.

***Preschool and Kindergarten.*** The preschool and kindergarten programs in Success for All emphasize language development, readiness, and self-concept. Preschools and kindergartens use thematic units, language development activities and a program called Story Telling and Retelling (STaR).

***Eight-Week Assessments.*** Students in grades 1-6 are assessed every eight weeks to determine whether they are making adequate progress in reading. This information is used to suggest alternate teaching strategies in the regular classroom, changes in reading group placement, provision of tutoring services, or other means of meeting students' needs.

***Family Support Team.*** A family support team works in each school to help support parents in ensuring the success of their children, focusing on parent education, parent involvement, attendance, and student behavior. This team is composed of existing or additional staff such as parent liaisons, social workers, counselors, and vice principals.

***Facilitator.*** A program facilitator works with teachers to help them implement the reading program, manages the eight-week assessments, assists the family support team, makes sure that all staff are communicating with each other, and helps the staff as a whole make certain that every child is making adequate progress.

Despite this extraordinary body of research, criticism of the effectiveness of Success for All has recently appeared in a few publications, including *Phi Delta Kappan, Education Week,* and the *Educational Researcher*. The authors of these articles, Herbert Walberg, his student Rebecca Greenberg, and Stanley Pogrow, primarily base their criticisms on a subjective conclusion reached in a single study by Richard Venezky (1994). Venezky, a

University of Delaware professor, was brought in by a local foundation to review research on Success for All in its first five schools in Baltimore. Venezky questioned the effects of Success for All because, at the end of fifth grade, Success for All students scored below grade level on the CTBS. Venezky believed that no matter how much the Success for All students exceeded matched control students, the program could not be considered successful if it did not bring children to grade level.

The schools in Venezky's study were among the most impoverished and historically low achieving in a very impoverished and low-achieving district, and had suffered in implementation quality due to a change of superintendents. Still, the Success for All fifth graders performed a full grade equivalent ahead of controls on the individually administered measures and 75% of a grade equivalent higher on CTBS. In addition, Success for All substantially reduced absenteeism and special education placements. These differences were clearly acknowledged by Venezky (1994, p. 12):

> "With all of these caveats, we found the control group studies to be valid and important records of SFA outcomes. The sample of students tested in the SFA schools significantly outperforms those tested in control schools in reading and language arts…We also found the analyses of retention data and of attendance to be valid…

The critics focus on this one study, and mention just two others that were critical of

SFA, on the basis that these are the only third-party evaluations. To justify this tactic, they resort to an *ad hominem* attack ordinarily dismissed in scholarly or scientific debate. Any other studies, they erroneously claim, were done either by Johns Hopkins researchers or by friends or colleagues of the developers. Walberg and Greenberg as well as Pogrow go so far as to question the motives not only of the developers and their colleagues, but also their funding agencies, government agencies, and anyone else with an intellectual or reputational stake in the outcome—in a word, everyone.

In fact, in Walberg's twisted system, only negative data are admissible. Walberg and Greenberg make much of a study of a single school in Charleston, South Carolina (Jones, Gottfredson, & Gottfredson, 1997) done by researchers who were at Johns Hopkins at the time the research was done. This study reported mixed results for a school that was, among other disasters, substantially damaged by Hurricane Hugo. Walberg and Greenberg cited it prominently. Had it been positive, however, they would of course have ignored it as "just another Johns Hopkins study." They ignored many other studies by researchers not at Johns Hopkins because they claimed that these researchers were friends of the developers. Again, however, if the conclusions had been negative, these very same "friends" would have been celebrated. *Ad hominem* reasoning is repugnant when it dismisses research findings solely on the basis that the research confirms the author's expectations. When this then extends to unconnected researchers in different institutions who are declared to be the author's "friends" after the fact, solely on the basis that they produced supportive evidence, *ad hominem* becomes utterly absurd, a cowardly and intellectually empty form of criticism. If evidence were accepted only if it were produced by researchers who initially knew nothing about or disliked the originator of a given

program or theory, the education journals would be empty, or nearly so. Critics have every right to examine and critique research on Success for All or any other program. But widely accepted norms of scientific discourse demand that they address the research itself--the entire body of research, not just the particular studies deemed to be acceptable precisely because they were critical and could therefore be assumed not to have been done by the original author's friends.

The criticism that research on Success for All has been done only by the developers is particularly bizarre because Success for All is exemplary, not deficient, in exactly this regard. Quite in contrast to the situation with most educational programs, Success for All has *primarily* been evaluated by third-party evaluators. Table I shows a list of individuals who have carried out research on Success for All over the years, organized according to the research institutions in which they were employed when their studies were done.

**Table 1**
**Researchers Involved in Studies of Achievement Effects of Success for All**

**Johns Hopkins University**
Robert Slavin
Nancy Madden (now at Success for All Foundation)
Nancy Karweit (now retired)
Barbara Wasik
Lawrence Dolan (now at College Board)
Sam Stringfield
Rebecca Herman (now at American Institutes of Research)
Gary Gottfredson (now at Gottfredson Associates)
Denise Gottfredson (now at Gottfredson Associates)
John Nunnery (now at Memphis City Schools)
Margarita Calderón
Robert Stevens (now at Penn State)
Nancy Yoder

**Success for All Foundation**
Eric Hurley
Anne Chamberlain

**WestEd**
Marcie Dianda (now at National Education Assoc.)
Margaret Livingston
John Flaherty

**College of Notre Dame**
Barbara Livermon

**Abt Associates**
Mary Ann Milsap
Nancy Brigham
Marjorie Levin

**University of North Carolina—Charlotte**
Pamela Nesselrodt (now at Loyola of Chicago)
Eugene Schaffer

**University of Delaware**
Richard Venezky

**Houston ISD**
Joseph Stubbs
Phyllis Hunter
(now at Texas Education Agency)

**Dade County Schools**
Steven Urdegar

**University of Memphis**
Steven Ross
Lana Smith
Jason Casey (now at U. of Pennsylvania)
Marty Alberg
Mary McNelis
Tracey Lewis
Steven Loomis
L. Weiping Wang
Jeanine Rakow

**Univ. of Tennessee, Knoxville**
William Sanders
S. Paul Wright

**Gottfredson Associates**
Elizabeth Jones

**Nottingham University** (England)
David Hopkins
Alma Harris
Mick Youngman
Judith Wordsworth

**Haifa University** (Israel)
Rachel Hertz-Lazarowitz
Bruria Schaedel

**Concordia University** (Canada)
Bette Chambers
(now at Success for All Foundation)
Philip Abrami
Scott Morrison

**Macquarie Univ.** (Australia)
Yola Center
Louella Freeman
Magdalena Mock
Gregory Robertson

There are fifty-two researchers listed in Table 1. Five of them have written primarily about situations in which Success for All failed to consistently improve student achievement. Forty-seven have written primarily about evaluations in which Success for All has improved student achievement. Yet the critics have glorified the five and ignored the forty-seven, in thirteen research institutions beyond Johns Hopkins or the Success for All Foundation. Some were colleagues of Success for All's developers before they began their research, but most were not. Even among those who were, we neither have nor want "friends" who misrepresent data. Perhaps the most important third-party evaluation, carried out using Tennessee state accountability data from Memphis by William Sanders and S. Paul Wright of the University of Tennessee at Knoxville (Ross, Sanders, & Wright, 1998), was done by researchers we've never met.

## Research on Achievement Effects of Success for All

Far more important than the question of who did the research is the nature and findings of the research itself. This article summarizes the research on the achievement effects of Success for All, the research carried out by the many researchers listed in Table 1 (see Slavin & Madden, 1999, in press, for more complete descriptions of the research).

From the very beginning, there has been a strong focus in Success for All on research and evaluation. Longitudinal evaluations of Success for All were begun in its earliest sites, six schools in Baltimore and Philadelphia. Later, third-party evaluators at the

University of Memphis (Steven Ross, Lana Smith, and their colleagues) added evaluations in Memphis; Houston, Texas; Charleston, South Carolina; Montgomery, Alabama; Ft. Wayne, Indiana; Caldwell, Idaho; Tucson, Arizona; Clover Park, Washington; Little Rock, Arkansas; and Clarke County, Georgia. Studies focusing on English language learners in California have been conducted in Modesto and Riverside by researchers at WestEd, a federally-funded regional educational laboratory. Research on Success for All and closely related programs has been carried out by researchers in England, Canada, Australia, Mexico, and Israel. Each of these evaluations has compared Success for All schools to matched comparison schools using either traditional methods or alternative reform models on measures of reading performance, starting with cohorts in kindergarten or in first grade and continuing to follow these students as long as possible (details of the evaluation design appear below). Vagaries of funding and other local problems have ended some evaluations prematurely, but some have been able to follow Success for All schools for many years. Other studies have compared Success for All to a variety of alternative reform models, have compared full and partial implementations of SFA, and have made other comparisons.

## Studies Comparing Success for All to Matched Control Groups

The largest number of studies has compared the achievement of students in Success for All schools to that of children in matched comparison schools using traditional methods, including locally-developed Title I reforms. These studies primarily used individually-administered, standardized measures of reading (see below). The only studies excluded are three studies in which there were pretest differences between Success for All and control groups of more than 30% of a standard deviation.

## Table 2

## Characteristics of Success for All Schools in Experimental-Control Group Comparisons

| District/School | Enrollment | % Free Lunch | Ethnicity | Date Began SFA | Data Collected | Comments |
|---|---|---|---|---|---|---|
| **Baltimore, MD** | | | | | | |
| B1 | 500 | 83 | B-96% W-4% | 1987 | 88-94 | First SFA school; had additional funds first 2 years. |
| B2 | 500 | 96 | B-100% | 1988 | 89-94 | Had additional funds first 4 years. |
| B3 | 400 | 96 | B-100% | 1988 | 89-94 | |
| B4 | 500 | 85 | B-100% | 1988 | 89-94 | |
| B5 | 650 | 96 | B-100% | 1988 | 89-94 | |
| **Philadelphia, PA** | | | | | | |
| P1 | 620 | 96 | A - 60% W-20% B - 20% | 1988 | 89-94 | Large ESL program for Cambodian children. |
| P2 | 600 | 97 | B - 100% | 1991 | 92-93 | |
| P3 | 570 | 96 | B- 1 00% | 1991 | 92-93 | |
| P4 | 840 | 98 | B - 100% | 1991 | 93 | |
| P5 | 700 | 98 | L - 100% | 1992 | 93-94 | Study only involved students in Spanish bilingual program. |
| **Charleston, SC** | | | | | | |
| CS1 | 500 | 40 | B - 60% W-40% | 1990 | 91-92 | |
| **Memphis, TN** | | | | | | |
| MT1 | 350 | 90 | B - 95% W - 5% | 1990 | 91-94 | Program implemented only in grades K-2. |
| MT2 | 530 | 90 | B - 100% | 1993 | 94 | |
| MT3 | 290 | 86 | B - 100% | 1993 | 94 | |
| MT4 | 370 | 90 | B - 100% | 1993 | 94 | |
| **Ft. Wayne, IN** | | | | | | |
| F1 | 396 | 80 | B - 45 % W - 55% | 1991 | 92-94<br>97-98 | |
| F2 | 305 | 67 | B - 50% W - 50% | 1991 | 92-94<br>97-98 | |
| F3 | 588 | 82 | B - 66% W - 34% | 1995 | 97-98 | |
| **Montgomery, AL** | | | | | | |
| MA1 | 450 | 95 | B - 100% | 1991 | 93-94 | |
| MA2 | 460 | 97 | B - 100% | 1991 | 93-94 | |

**Table 2 (cont'd)**
**Characteristics of Success for All Schools in Experimental-Control Group Comparisons**

| District/ School | Enrollment | % Free Lunch | Ethnicity | Date Began SFA | Data Collected | Comments |
|---|---|---|---|---|---|---|
| **Caldwell, ID** | | | | | | |
| CI1 | 400 | 20 | W - 80%<br>L - 20% | 1991 | 93-94 | Study compared two SFA schools to Reading Recovery school. |
| **Modesto, CA** | | | | | | |
| MC1 | 640 | 70 | W - 54%<br>L - 25%<br>A - 17%<br>B - 4% | 1992 | 94 | Large ESL program for students speaking 17 languages. |
| MC2 | 560 | 98 | L - 66%<br>W - 24%<br>A - 10% | 1992 | 94 | Large Spanish bilingual program. |
| **Riverside, CA** | | | | | | |
| R1 | 930 | 73 | L - 54%<br>W - 33%<br>B - 10%<br>A - 3% | 1992 | 94 | Large Spanish bilingual and ESL programs. Year-round school. |
| **Tucson, AZ** | | | | | | |
| T1 | 484 | 82 | L - 54%<br>W - 34%<br>B - 69%<br>A - 5% | 1995 | 95-96 | Compared to locally-developed schoolwide projects |
| T2 | 592 | 43 | W - 73%<br>L - 23%<br>B - 1%<br>A - 1% | 1995 | 95-96 | Compared to locally-developed schoolwide projects and Reading Recovery |
| **Little Rock, AR** | | | | | | |
| LR1 | 302 | 73 | B - 80%<br>W - 20% | 1997 | 98-99 | |
| LR2 | 262 | 79 | B - 95%<br>L - 5% | 1997 | 98-99 | |
| **Clarke County, GA** | | | | | | |
| CL1 | 420 | 70 | B - 80%<br>W - 20% | 1995 | 97 | |
| CL2 | 488 | 72 | B - 78%<br>W - 22% | 1995 | 97 | |

Note: SFA = Success for All; ESL = English as a Second Language; B = African American; L = Latino; A = Asian American; W = White

Table 2 summarizes demographic and other data about the schools involved in the experimental-control evaluations of Success for All.

A common evaluation design, with variations due to local circumstances, has been used in most Success for All evaluations carried out by researchers at Johns Hopkins

University, the University of Memphis, and WestEd. Each Success for All school involved in a formal evaluation was matched with a control school that is similar in poverty level (percent of students qualifying for free lunch), historical achievement level, ethnicity, and other factors. Schools were also matched on district-administered standardized test scores given in kindergarten or on Peabody Picture Vocabulary Test (PPVT) scores given by the evaluators in the fall of kindergarten or first grade. The measures used in the evaluations were three scales from the Woodcock Reading Mastery Test (Word Identification, Word Attack, and Passage Comprehension, grades K-6), the Durrell Oral Reading scale (grades 1-3), and the Gray Oral Reading Test (grades 4-7). Analyses of covariance with pretests as covariates were used to compare raw scores in all evaluations, and separate analyses were conducted for students in general and, in most studies, for students in the lowest 25% of their grades.

The figures presented in this article summarize student performance in grade equivalents (adjusted for covariates) and effect size (proportion of a standard deviation separating the experimental and control groups), averaging across individual measures. Neither grade equivalents nor averaged scores were used in the analyses, but they are presented here as a useful summary.

Each of the evaluations in this section follows children who began in Success for All in first grade or earlier, in comparison to children who had attended the control school over the same period. Students who start in the program after first grade were not considered to have received the full treatment (although they are of course served within the schools).

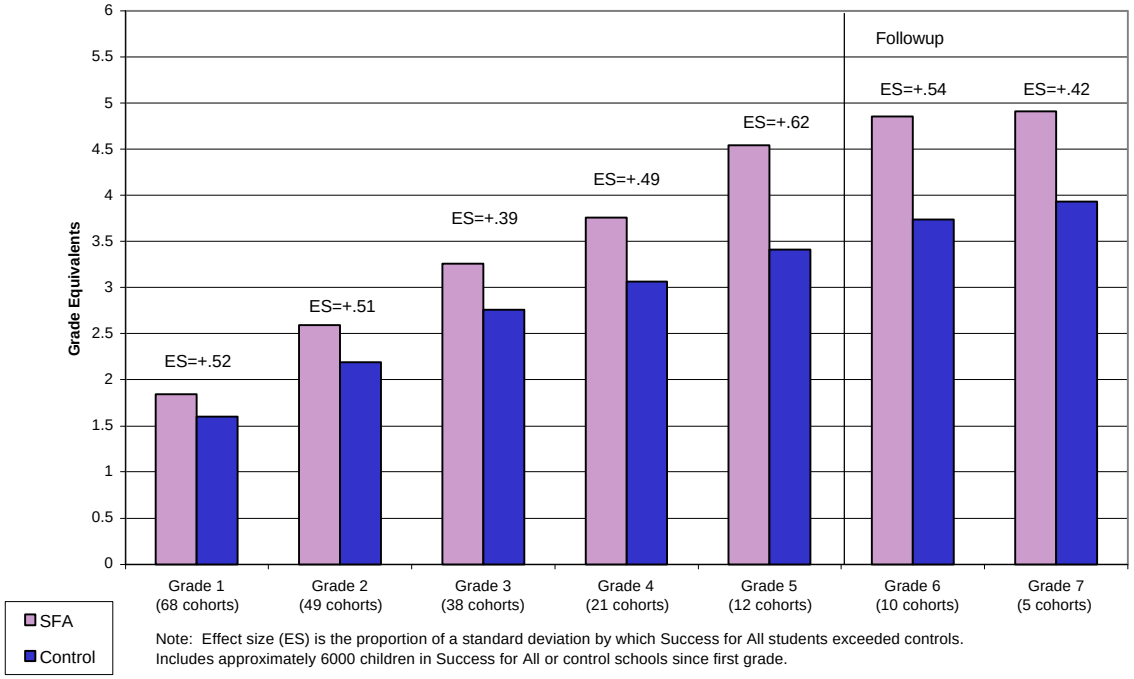
Results for all experimental-control comparisons in all evaluation years are averaged

and summarized in Figure 1 using a method called multi-site replicated experiment (Slavin et al., 1996a,b; Slavin & Madden, 1993).

## Reading Outcomes

The results of the multi-site replicated experiment evaluating Success for All are summarized in Figure 1 for each grade level, 1-5, and for follow-up measures into grades 6 and 7. The analyses compare cohort means for experimental and control schools. A cohort is all students at a given grade level in a given year. For example, the Grade 1 graph compares 68 experimental to 68 control cohorts, with cohort (50-150 students) as the unit of analysis. In other words, each first grade bar is a mean of scores from about 6000 students. Grade equivalents are based on the means, and are only presented for their informational value. No analyses were done using grade equivalents.

**Figure 1**
**Comparison of Success for All and Control Schools in Mean Reading Grade Equivalents and Effect Sizes 1988-1999**




Note: Effect size (ES) is the proportion of a standard deviation by which Success for All students exceeded controls.
Includes approximately 6000 children in Success for All or control schools since first grade.

Statistically significant ($p$=.05 or better) positive effects of Success for All (compared to controls) were found on every measure at every grade level, 1-5, using the cohort as the unit of analysis. For students in general, effect sizes averaged around a half standard deviation at all grade levels. Effects were somewhat higher than this for the Woodcock Word Attack scale in first and second grades, but in grades 3-5 effect sizes (ES) were more or less equivalent on all aspects of reading. Consistently, effect sizes for students in the lowest 25% of their grades were particularly positive, ranging from ES=+1.03 in first grade to ES=+1.68 in fourth grade. Again, cohort-level analyses found statistically significant differences favoring low achievers in Success for All on every measure at every grade level. A followup study of Baltimore schools found that similar positive program effects for the full sample of students continued into grade 6 (ES=+0.54) and grade 7 (ES=+0.42), when students were in middle schools.

**Effects on District-Administered Standardized Tests**

The formal evaluations of Success for All have relied on individually-administered assessments of reading. The Woodcock and Durrell scales used in these assessments are far more accurate than district-administered tests, and are much more sensitive to real reading gains. They allow testers to hear children actually reading material of increasing difficulty and responding to questions about what they have read. The Woodcock and Durrell scales are themselves nationally standardized tests, and produce norms (e.g., percentiles, NCEs, and grade equivalents) just like any other standardized measure.

However, educators usually want to know the effects of innovative programs on the

kinds of group-administered standardized tests they are usually held accountable for. There are hundreds of test score reports from individual Success for All schools showing dramatic gains on standardized tests, and these are the types of data so often used by other program developers, including Stanley Pogrow (1990), to support their programs. However, such evaluations have no scientific validity, both because they have no comparison groups (test scores may have been rising in the entire district or state) and because such score gain data are usually reported for selected schools that happened to make gains in a given year (see Slavin & Fashola, 1998).

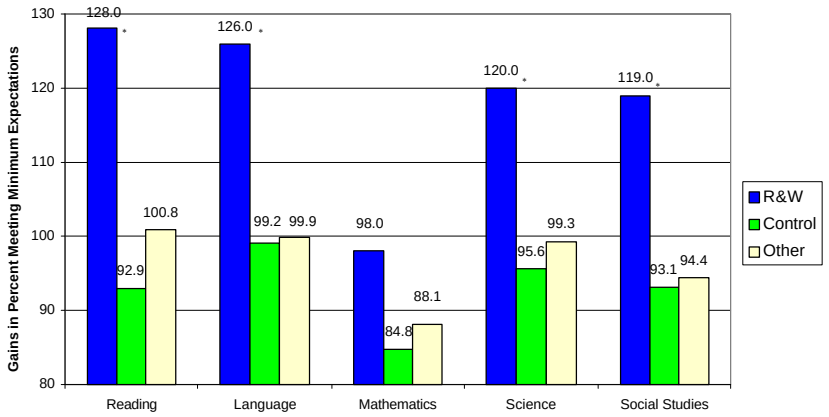
District test score data can produce valid evaluations of educational programs if comparison groups are available. To obtain this information, researchers have often analyzed standardized or state criterion-referenced test data comparing students in experimental and control schools. The following sections briefly summarize findings from these types of evaluations.

**Memphis, Tennessee**

One of the most important independent evaluations of Success for All/Roots & Wings is a study carried out by researchers at the University of Tennessee-Knoxville for the Memphis City Schools (Ross, Sanders, & Wright, 1998). William Sanders, the architect of the Tennessee Value-Added Assessment System (TVAAS), who was not familiar with any of the developers of the programs he evaluated, carried out the analysis. The TVAAS gives each school an expected gain, independent of school poverty levels, and compares it to actual scores on the Tennessee Comprehensive Assessment Program (TCAP).

TVAAS scores above 100 indicate gains in excess of expectations; those below 100 indicate the opposite. Sanders compared TVAAS scores in eight Memphis Success for All schools to those in (a) matched comparison schools, and (b) all Memphis schools.

**Figure 2**
**Percent of Expected Gain on TVAAS for Roots & Wings, Control, and Other Memphis Schools, Grades 2-5, 1997**




* Roots & Wings scores significantly higher than those of controls or others ($p<.05$).
Tennessee Value-Added Assessment System (TVAAS) scores analyzed by William Sanders for the Tennessee State Department of Education. See Ross, Sanders, & Wright, 1998.

Figure 2 summarizes the results for all subjects assessed. At pretest, the Success for All schools were lower than both comparison groups on TVAAS. However, after two years of implementation, they performed significantly better than comparison schools in reading, language, science, and social studies. While some schools had implemented aspects of WorldLab, the social studies/science component used in some Success for All schools, none had implemented MathWings, the math program. Despite this, even math scores nonsignificantly favored the Success for All schools.

A third-year evaluation found that Success for All schools averaged the greatest gains and highest levels on the TVAAS of six restructuring designs (Co-nect, Accelerated Schools, Audrey Cohen College, ATLAS, and Expeditionary Learning), as well as

exceeding controls, averaging across all subjects and averaging data from the second and third implementation years (Ross, Wang, Sanders, Wright, & Stringfield, 1999).

The importance of the Memphis study lies in several directions. First, it is a completely independent evaluation that involved state assessment scores of the kind used in most state accountability systems. Second, it shows carryover effects of a program focused on reading, writing, and language arts into science and social studies outcomes.

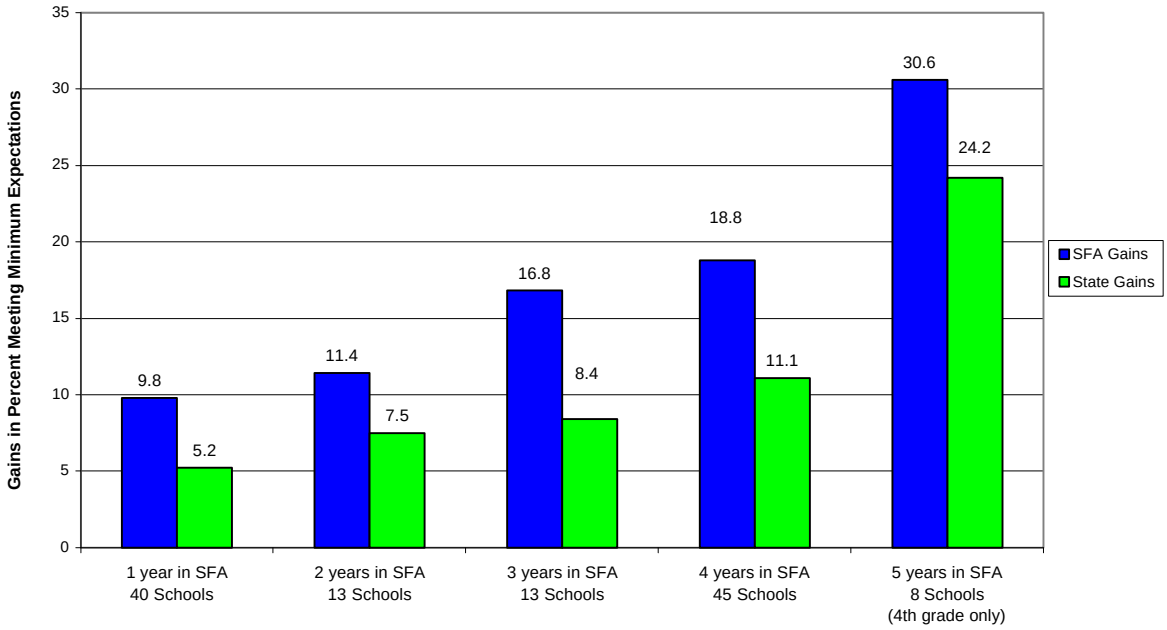
An earlier study of Success for All schools in Memphis (by Ross, Smith, & Casey, 1995) also showed positive effects on the TCAP. This was a longitudinal study of three Success for All and three control schools. On average, Success for All schools exceeded controls on TCAP reading by an effect size of +0.38 in first grade and +0.45 in second grade.

**State of Texas**

The largest study ever done to evaluate achievement outcomes of Success for All was recently completed by Hurley, Chamberlain, Slavin, & Madden (2000). Using data available on the Internet, Hurley et al. compared every school that ever used Success for All anywhere in the State of Texas during the period 1993-1998 (n=119 schools). Gains in these schools on the percent of students passing the Texas Assessment of Academic Skills (TAAS) reading measures were compared for grades 3-5 in the SFA schools and for the state as a whole; in each case, gains from the year before program inception to 1998 were compared. (Changes in testing procedures made 1999 scores non-

comparable). Figure 3 shows the overall results, which indicates greater gains for Success for All schools than for the rest of the state for every cohort. Analyzing school means, the differences are highly significant ($p < .001$; ES = +0.60).

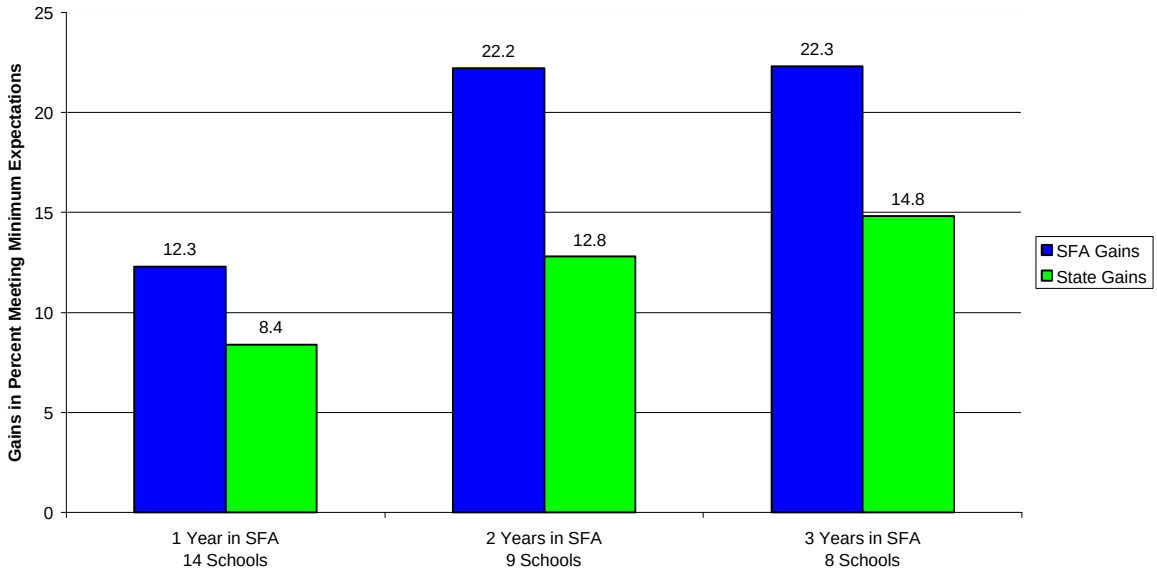
**Figure 3**
**Gains From Preimplementation Year to 1998**
**Success for All vs. Texas Means**
**Texas Assessment of Academic Skills**
**Reading, Grades 3,4,5**




The TAAS has been criticized for having a ceiling effect, giving the appearance of significantly reducing the gap between minority and white students. The Success for All analysis shown above may reflect this problem, as Success for All schools are far more impoverished than the state average (students receiving free lunches are 85% of those in SFA schools and 45% in the state as a whole). However, if there is a ceiling effect it exists primarily among white students, who averaged 94.1% passing in 1998. African-American students across the state averaged 81.8% passing, and Hispanic

students averaged 79.6% passing. Partly to address this possible ceiling effect, Hurley et al. (2000) compared scores for African-American and Hispanic students in Success for All schools and those for similar students in the state as a whole for 1995-1998 (years when state scores were available by ethnicity). Figures 4 and 5 show these

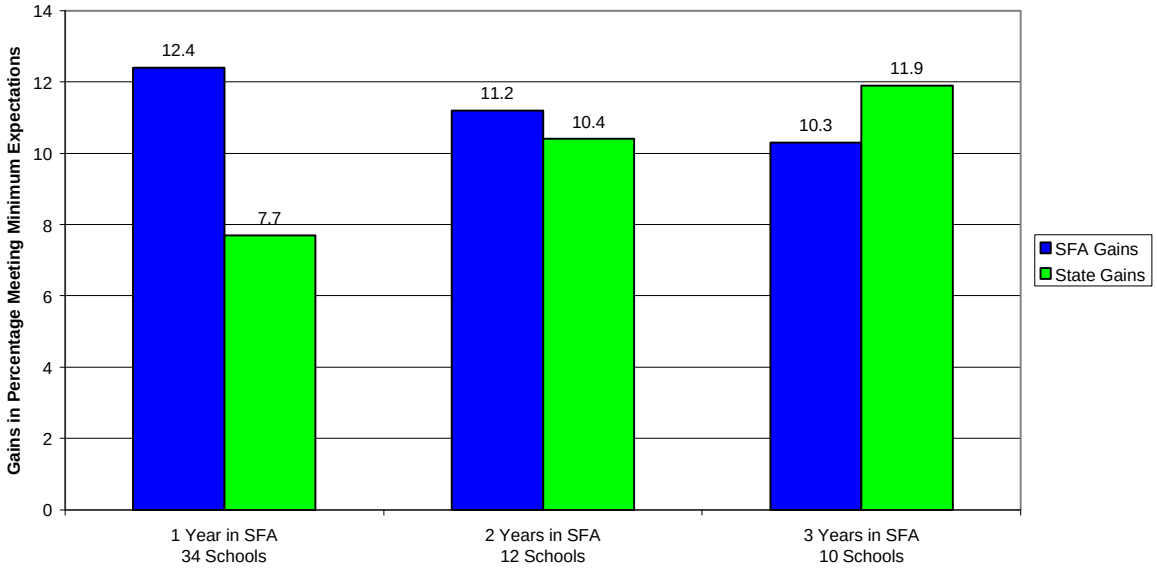
**Figure 4**
**Gains From Preimplementation Year to 1998**
**Success for All Schools vs. State of Texas**
**Texas Assessment of Academic Skills**
**African-American Students, Grades 3-5**




results.

As Figure 4 shows, African-American students in Success for All schools were closing the gap with white students much faster than were other African-American students. For example, SFA African-American students advanced from 63.3% passing in 1995 to 86.2% passing in 1998, while other African-American students only gained from 64.2% passing to 78.9% passing. Patterns were not quite as clear for Hispanic students (Figure 5), but in two of the three cohorts, Hispanic students in SFA gained more on percent passing TAAS than did Hispanic students elsewhere in the state. Combining

**Figure 5**
**Gains From Preimplementation Year to 1998**
**Success for All Schools vs. State of Texas**
**Texas Assessment of Academic Skills**
**Hispanic Students, Grades 3-5, Reading**




across cohorts, African-American scores gained significantly more in SFA schools than in the state ($p<.05$), while Hispanic scores gained marginally faster ($p<.07$).

What is particularly important about the Texas analyses is that they involve all 119 schools that ever used Success for All in Texas during 1995-1998. There is no "cherry picking," selection of schools that happened to have more gains. Further, although the analyses were carried out by researchers at the Success for All Foundation, they used data that are readily available on the Internet, so anyone with an Internet account and a list of schools (which SFAF will provide) can replicate them.

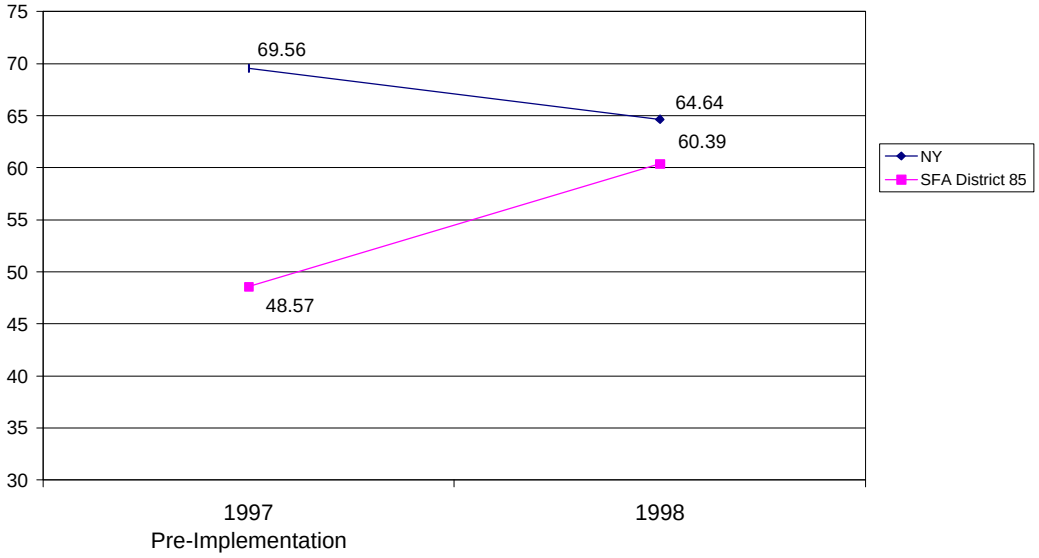
**New York City**

Another study using data from the Internet evaluates schools in the Chancellor's District (District 85) in New York City. This is a "district" composed of schools whose achievement levels were so low that they were taken from their community districts (New York City has 32 community districts) and assigned to a special city-wide district in which they received additional resources and assistance as well as additional accountability pressure; if the schools did not show improvement, they could be closed down or reconstituted. Chancellor's District schools were strongly encouraged to take on Success for All, and over time, all of them have voted in favor of Success for All.

Figure 6 shows the first-year gains for all six Chancellor's District schools that began Success for All in 1997. Unfortunately, as in Texas, a change in testing procedures made it impossible to track schools from pretest to the present.

Figure 6 shows data on the percentage of students performing at or above the state reference point on the New York State Pupil Evaluation Program (PEP) in Reading for the Success for All schools and for the entire city. As is clear from the figure, these schools started far below the New York City mean. However, after one year, they were nearly equal to the city mean. Again, our staff carried out these analyses, but any researcher with an Internet account and a list of schools could replicate them.

**Figure 6**
**Percent of Students at or Above State Reference Point (SRP)**
**for Pupil Evaluation Program (PEP) Test**
**District 85 Success for All Schools vs. NYC**




## Special Strategies

A study of ten innovative programs was commissioned by the U.S. Department of Education as part of *Prospects*, the national longitudinal evaluation of Title I (Stringfield, Millsap, Herman, Yoder, Brigham, Nesselrodt, Schaffer, Karweit, Levin, & Stevens, 1997). Some of the programs were locally developed, some used targeted designs (e.g., Reading Recovery), and four used comprehensive designs: Success for All, Comer's School Development Project, Paideia, and the Coalition of Essential Schools. All participating schools were followed over a three-year period on the CTBS. Only two of the ten programs, Success for All and the Comer model, showed significantly greater achievement gains than other schools.

**Baltimore**

A longitudinal study in Baltimore from 1987-1993 collected CTBS scores on the original five Success for All and control schools. On average, Success for All schools exceeded control schools at every grade level. The differences were statistically and educationally significant. By fifth grade, Success for All students were performing 75% of a grade equivalent ahead of controls (ES=+0.45) on CTBS Total Reading scores (see Slavin, Madden, Dolan, Wasik, Ross, & Smith, 1994).

**International Evaluations of Success for All Adaptations**

Several studies have assessed the effects of adaptations of Success for All in countries outside of the United States. These adaptations have ranged from relatively minor adjustments to accommodate political and funding requirements in Canada and England to more significant adaptations in Mexico, Australia, and Israel.

The Canadian study (Chambers, Abrami, & Morrison, in press) involved one school in Montreal, which was compared to a matched control school on individually-administered reading measures. Results indicated significantly better reading performance in the Success for All school than in the control school, both for special needs students (a large proportion of the SFA students) and for other students. Similarly, a study of five SFA schools in Nottingham, England found that Success for All students gained more in reading than did students in a previous cohort, before the program was

introduced (Hopkins, Youngman, Harris, & Wordsworth, 1999; Harris, Hopkins, Youngman, & Wordsworth, in press).

A school in Juarez, Mexico, across the border from El Paso, Texas, implemented the Spanish adaptation of Success for All, Éxito Para Todos (Calderón, in press). This study showed substantial gains relative to an earlier cohort for the experimental schools.

Because of language and cultural differences, the most significant adaptation of Success for All was made to use the program in Israel with both Hebrew-speaking children in Jewish schools and Arabic-speaking children in Israeli Arab schools, all in or near the northern city of Acre. The implementation involved community interventions focusing on parent involvement, integrated services, and other aspects in addition to the adapted Success for All model. In comparison to control groups, Success for All first graders performed at significantly higher levels on tests of reading and writing (Hertz-Lazarowitz, in press).

Finally, Australian researchers created a simplified adaptation of Success for All, which they called SWELL. SWELL uses instructional procedures much like those used in Success for All, but uses books adapted for the Australian context. Only the early grades are involved, schools do not have full-time facilitators or family support programs, and they may or may not provide any tutoring. Two studies of SWELL found positive effects of the program on reading performance in comparison to control groups and to Reading Recovery schools (Center, Freeman, & Robertson, in press; Center, Freeman,

Mok, & Robertson, 1997).

The international studies of programs adapted from Success for All have importance in themselves, of course, but also indicate that the principles on which Success for All are based transfer to other languages, cultures, and political systems. In addition, they provide third-party evaluations of Success for All in diverse contexts, strengthening the research base for Success for All principles and practices.

## Quality and Completeness of Implementation

Not surprisingly, effects of Success for All are strongly related to the quality and completeness of implementation. In a large study in Houston, Nunnery, Slavin, Ross, Smith, Hunter, and Stubbs (in press) found that schools implementing all program components obtained better results (compared to controls) than did schools implementing the program to a moderate or minimal degree.

A Memphis study (Ross, Nunnery, Smith, & Lewis, 1997; Ross, Smith, & Nunnery, 1998) compared the achievement of eight Success for All schools to that of four schools using other restructuring designs, matched on socioeconomic status and PPVT scores. Each pair of SFA schools had one school rated by observers as a high implementer, and one rated as a low implementer. In the 1996 cohort, first grade results depended entirely on implementation quality. Averaging across the four Woodcock and Durrell scales, every comparison showed that high-implementation SFA schools scored higher than their comparison schools, while low-implementation SFA schools scored lower (Ross et al.,

1996). However, by second grade, Success for All schools (high as well as low implementers) exceeded comparison schools, on average.

A Miami study (Urdegar, 1998) evaluated Success for All, two integrated learning systems computer programs (CCC and Jostens), and Reading Mastery, on the Stanford Achievement Test's Reading Comprehension scale. None of the programs was associated with higher achievement gains than matched controls. However, buy-in procedures were not followed, a change of superintendents led to a withdrawal of support, and program implementation was very poor in the Success for All schools, particularly in that there were few or no tutors in most schools. Also, a pretest, given eight months before the posttest, was used as a covariate, even though the programs had been used for several years in most schools. The pretest is likely to reflect some or all of the program's impact over time, making the analysis of covariance difficult to interpret.

An early study by a separate team of Johns Hopkins researchers also found mixed outcomes in a study with serious implementation problems. This study, in Charleston, South Carolina, compared one school to a matched control school. However, the researchers failed to obtain the required 80% vote in favor of the program, implementation was very poor, and Hurricane Hugo ripped the roof off of the school, closing it for two months and disrupting it for many more. Despite this, most kindergarten and first grade measures favored Success for All, and retentions in grade were significantly diminished. However, second and third grade measures did not favor the Success for All school (Jones, Gottfredson, & Gottfredson, 1997).

**Comparisons With Other Programs**

A few studies have compared outcomes of Success for All to those of other reform model designs.

As noted earlier, a study of six restructuring designs in Memphis on the Tennessee Value Added Assessment System (TVAAS) found that Success for All schools had the highest absolute scores and gain scores on the TVAAS, averaging across all subjects (Ross et al., 1999).

A study in Clover Park, Washington, compared Success for All to Accelerated Schools (Hopfenberg & Levin, 1993), an approach that, like Success for All, emphasizes prevention and acceleration over remediation, but unlike Success for All does not provide specific materials or instructional strategies to achieve its goals. In the first year of the evaluation, the Success for All and Accelerated Schools programs had similar scores on individually administered reading tests and on a writing test (Ross, Alberg, & McNelis, 1997). By second grade, however, Success for All schools were scoring slightly ahead of Accelerated Schools in reading, and significantly ahead in writing (Ross, Alberg, McNelis, & Smith, 1998).

Two studies compared Success for All to schools using Reading Recovery. In one, in rural Caldwell, Idaho, first graders scored somewhat better in SFA than in the

Reading Recovery schools (ES=+17), but there were no differences in scores between students tutored in SFA and those tutored in Reading Recovery (Ross, Smith, Casey, & Slavin, 1995). In an Arizona study, Ross, Nunnery, & Smith (1996) compared urban first graders in schools using SFA, Reading Recovery, or a locally-developed Title I schoolwide project. Results strongly favored SFA over both schools (ES=+0.68 for Reading Recovery, +0.39 for the locally developed model), and even the tutored students performed far better in SFA than in Reading Recovery schools (ES=+2.79).

**Success for All and English Language Learners**

Six studies have evaluated adaptations of Success for All with language minority children (see Slavin & Madden, 1999b). Three of these evaluated *Éxito Para Todos* ("Success for All" in Spanish), the Spanish bilingual adaptation, and three evaluated a program adaptation incorporating English as a second language strategies.

**Bilingual Studies.** One study compared students in *Éxito Para Todos* to those in a matched comparison school in which most reading instruction was in English. Both schools served extremely impoverished, primarily Puerto Rican student bodies in inner-city Philadelphia. Not surprisingly, *Éxito Para Todos* students scored far better than control students on Spanish measures. More important was the fact that after transitioning to all-English instruction by third grade, the *Éxito Para Todos* students scored significantly better than controls on measures of English reading.

An evaluation of *Éxito Para Todos* in California bilingual schools was reported by Livingston and Flaherty (1997), who studied three successive cohorts of students. On Spanish reading measures, *Éxito Para Todos* students scored significantly higher than

controls in all grades, 1-3. A large study in Houston compared limited English proficient (LEP) first graders in 20 schools implementing *Éxito Para Todos* to those in 10 control schools (Nunnery, Slavin, Madden, Ross, Smith, Hunter, & Stubbs, in press). As an experiment, schools were allowed to choose Success for All/*Éxito Para Todos* as it was originally designed, or to implement key components. Medium-implementation schools significantly exceeded their controls on all measures (mean ES=+0.24). Low implementers exceeded controls on the Spanish Woodcock Word Identification and Word Attack scales, but not on Passage Comprehension (mean ES=+0.17).

One additional study evaluated Bilingual Cooperative Integrated Reading and Composition (BCIRC), which is closely related to *Alas Para Leer*, the bilingual adaptation of Reading Wings. This study, in El Paso, Texas, found significantly greater reading achievement (compared to controls) for English language learners in grades 3-5 transitioning from Spanish to English reading (Calderón, Hertz-Lazarowitz, & Slavin, 1998).

**English as a Second Language (ESL) Studies.** Three studies have evaluated the effects of Success for All with English language learners being taught in English. In this adaptation, ESL strategies (such as total physical response) are integrated into instruction for all children, whether or not they are limited in English proficiency. The activities of ESL teachers are closely coordinated with those of other classroom teachers, so that ESL instruction directly supports the Success for All curriculum, and ESL teachers often serve as tutors for LEP children.

The first study of Success for All with English language learners took place in Philadelphia. Students in an Asian (mostly Cambodian) Success for All school were

compared to those in a matched school that also served many Cambodian-speaking children. Both schools were extremely impoverished, with nearly all children qualifying for free lunches.

At the end of a six-year longitudinal study, Success for All Asian fourth and fifth graders were performing far ahead of matched controls. On average, they were 2.9 years ahead of controls in fourth grade (median ES=+1.49), and 2.8 years ahead in fifth grade (median ES= +1.33). Success for All Asian students were reading about a full year above grade level in both fourth and fifth grades, while controls were almost two years below grade level. Non-Asian students also significantly exceeded their controls at all grade levels (see Slavin & Madden, 1999b).

The California study described earlier (Livingston & Flaherty, 1997) also included many English language learners who were taught in English. Combining results across three cohorts, Spanish-dominant English language learners performed far better on English reading measures in Success for All than in matched control schools in first and second grades.

An Arizona study (Ross, Nunnery, & Smith, 1996) compared Mexican American English language learners in two urban Success for All schools to those in three schools using locally-developed Title I reform models and one using Reading Recovery. Two SES school strata were compared, one set with 81% of students in poverty and 50% Hispanic students and one with 53% of students in poverty and 27% Hispanic students. Success for All first graders scored higher than controls in both strata.

The effects of Success for All for language-minority students are not statistically significant on every measure in every study, but the overall impact of the program is

clearly positive, both for the Spanish bilingual adaptation, *Éxito Para Todos*, and for the ESL adaptation. What these findings suggest is that whatever the language of instruction may be, student achievement in that language can be substantially enhanced using improved materials, professional development, and other supports.

**Success for All and Special Education**

The data relating to special education-related outcomes clearly support the program's effects. One of the most important outcomes in this area is the consistent finding of particularly large effects of Success for All for students in the lowest 25% of their classes. While effect sizes for students in general have averaged around +0.50 on individually administered reading measures, effect sizes for the lowest achievers have averaged in the range of +1.00 to +1.50 across the grades (Slavin, 1996). In the longitudinal Baltimore study, only 2.2% of third graders averaged two years behind grade level, a usual criterion for special education placement. In contrast, 8.8% of control third graders scored this poorly. Baltimore data also showed a reduction in special education placements for learning disabilities of about half (Slavin et al., 1992). A study of two Success for All schools in Ft. Wayne, Indiana found that over a two-year period, 3.2% of Success for All students in grades K-1 and 1-2 were referred to special education for learning disabilities or mild mental handicaps. In contrast, 14.3% of control students were referred in these categories (Smith, Ross, & Casey, 1994).

Taken together, these findings support the conclusion that Success for All both reduces the need for special education services (by raising the reading achievement of very low achievers) and reduces special education referrals and placements.

## Policy Implications

There is no magic in education. No program works everywhere, and outcomes of any program depend on the quality, completeness, and appropriate application of the program. However, it would be astonishing if Success for All were *not* effective when fully implemented. The elements of the program are themselves based on rigorous research comparing schools using various practices to those in matched or randomly assigned control schools. Ironically, the program includes almost all of the elements identified in a synthesis of meta-analyses by Walberg himself (Walberg, 1986). In one sense, the contribution of the Success for All project is not primarily in the demonstration that the program works; it would be surprising if that were not true. The real contribution is in demonstrating that an effective program composed of elements that are themselves based on high-quality research can be scaled up to serve a large enough set of schools to matter at the policy level. The Texas data, as well as the Memphis and New York City data presented above, are particularly important in this regard in demonstrating that even aggregating state accountability data from more than a hundred schools, Success for All produces significantly greater gains than other schools. From a research perspective, the studies that followed individual children over time on individually-administered measures are better indicators than the state assessment data of the effects of Success for All on reading achievement and other outcomes. However, it is also essential to demonstrate effects on the measures for which schools are held accountable, and to show that the program does not lose effectiveness as it is disseminated on a very large scale.

The policy implications of the research on Success for All, and of the widespread

dissemination of the program, are potentially profound. The ability to affect student achievement in high-poverty Title I schools on a substantial scale means that there is little excuse for doing less. It is by no means appropriate to require schools to adopt Success for All or any other particular program, and we have never condoned this; we require a positive vote by secret ballot of at least 80% of all teachers before we will agree to work with a school. However, it is appropriate to provide start-up funding to help schools adopt from among a range of effective programs. This is precisely what happened in the New Jersey *Abbott* case where the New Jersey Supreme Court required schools in the 28 highest-poverty urban districts to select a proven comprehensive model. Success for All was identified as the "presumptive model" for elementary schools, but other models were also offered. The same is true of the Comprehensive School Reform Demonstration (CSRD), which provides grants of at least $50,000 for up to three years to help schools adopt proven, comprehensive models.

The CSRD grants and the New Jersey Abbott decision, among other more local policy decisions along similar lines, are harbingers of genuine change in school reform. For the first time ever, serious funding is being attached to evidence of effectiveness for school change models that affect the entire school. The potential here is revolutionary. It is now possible to contemplate setting in motion a process of research, development, evaluation, and dissemination that will truly transform our schools.

What both Pogrow and Walberg seem to be afraid of is in fact coming to pass. Title I is increasingly emphasizing comprehensive reform models, especially for schoolwide projects (in which at least 50% of students qualify for free lunches). An emphasis on evidence and comprehensiveness in Title I is anathema to Pogrow, whose HOTS program

has neither. It is anathema to Walberg, who wants to destroy Title I in favor of vouchers and charters. Yet research-based, comprehensive reform could be the salvation of millions of children in Title I schools. Instead of continuing to have Title I primarily support remedial programs or classroom aides, neither of which have much support in research, Title I schools could increasingly use programs that are well worked out, well researched, and capable of working with hundreds or thousands of schools with quality and integrity. The same process could have equally profound impacts on bilingual and English as a second language policies and on special education policies, as effective, well-evaluated, replicable programs become available in these areas as well. Today's models and today's research will surely be improved upon in the future with better models and better research; the comprehensive school reform movement is still very young. Let a thousand ideas bloom as long as there are scientifically valid evaluations of each involving experimental-control comparisons, or equally rigorous designs. It is possible to criticize Success for All or any other program, but how could any rational person oppose the *process* of developing, evaluating, and disseminating effective programs to high-poverty schools? It is entirely appropriate, indeed necessary, for federal agencies and other funders to support every step of the R & D process, from initial research to scale-up to third-party evaluation to broad dissemination. To oppose this, particularly by picking at perceived flaws in a tiny subset of studies on just one of many programs, merely sows doubt and misinformation about what is arguably the most promising development in educational reform in the 21st century.

**References**


Calderón, M. (in press). Success for All in Mexico. In R.E. Slavin & N.A. Madden (Eds.), *Success for All: Research and reform in elementary education*. Mahwah, NJ: Erlbaum.

Calderón, M., Hertz-Lazarowitz, R., & Slavin, R.E. (1998). Effects of Bilingual Cooperative Integrated Reading and Composition on students making the transition from Spanish to English reading. *Elementary School Journal, 99,* (2), 153-165.

Center, Y., Freeman, L., Mok, M., & Robertson, G. (1997, March). *An evaluation of Schoolwide Early Language and Literacy (SWELL) in six disadvantaged New South Wales schools*. Paper presented at the annual meeting of the American Educational Research Associaton, Chicago

Center, Y., Freeman, L., & Robertson, G. (in press). A longitudinal evaluation of the Schoolwide Early Language and Literacy Program (SWELL). In R.E. Slavin & N.A. Madden (Eds.), *Success for All: Research and reform in elementary education*. Mahwah, NJ: Erlbaum.

Harris, A., Hopkins D.,Wordsworth, J., & Youngman, M. (in press). The implementation and impact of Success for All in English schools. In R.E. Slavin & N.A. Madden (Eds.), *Success for All: Research and reform in elementary education*. Mahwah, NJ: Erlbaum.

Herman, R. (1999). *An educator's guide to schoolwide reform.* Arlington, VA: Educational Research Service.

Hertz-Lazarowitz, R. (in press). Success for All: A model for advancing Arabs and Jews in Israel. In R.E. Slavin & N.A. Madden (Eds.), *Success for All: Research and reform in elementary education.* Mahwah, NJ: Erlbaum.

Hopfenberg, W.S., & Levin, H.M. (1993). *The Accelerated Schools resource guide*. San Francisco: Jossey-Boss.

Hopkins, D., Youngman, M., Harris, A., & Wordsworth, J. (1999). Evaluation of the initial effects and implementation of Success for All in England. *Journal of Research in Reading, 22,* (3), 257-270.

Hurley, E., Chamberlain, A., Slavin, R.E., & Madden, N.A. (2000). *Effects of Success for All on TAAS Reading: A Texas statewide evaluation*. Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Jones, E.M., Gottfredson, G.D., & Gottfredson, D.C. (1997). Success for some: An evaluation of the Success for All program. *Evaluation Review, 21* (6), 643-670.

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

Nunnery, J., Slavin, R.E., Ross, S.M., Smith, L.J., Hunter, P., & Stubbs, J. (in press). An assessment of Success for All program component configuration effects on the reading achievement of at-risk first grade students. *American Educational Research Journal.*

Pogrow, S. (1990, January). Challenging at-risk students: Findings from the HOTS program. *Phi Delta Kappan, 71* (5), 389-397.

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

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

Ross, S.M., Nunnery, J., & Smith, L.J. (1996). *Evaluation of Title I Reading Programs: Amphitheater Public Schools. Year 1: 1995-96.* Memphis: University of Memphis, Center for Research in Educational Policy.

Ross, S.M., Nunnery, J., Smith, L., & Lewis, T. (1997). *Evaluation of Success for All programs in Memphis City Schools: 1996-1997.* Memphis: University of Memphis, Center for Research in Educational Policy.

Ross, S.M., Sanders, W.L., & Wright, S.P. (1998). *An analysis of Tennessee Value Added Assessment (TVAAS) performance outcomes of Roots & Wings schools from 1995-1997.* Memphis: University of Memphis.

Ross, S.M., Smith, L.J., & Casey, J.P. (1995). 1994-1995 *Success for All program in Ft. Wayne, IN: Final Report.* Memphis: University of Memphis, 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.

Ross, S.M., Smith, L.J., & Nunnery, J.A. (1998, April). *The relationship of program implementation quality and student achievement.* Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Ross, S.M., Wang, L.W., Sanders, W.L., Wright, S.P., & Stringfield, S. (1999). *Two- and three-year achievement results on the Tennessee Value-Added Assessment System for restructuring schools in Memphis*. Memphis: University of Memphis, Center for Research in Educational Policy.

Slavin, R.E. (1996). Neverstreaming: Preventing learning disabilities. *Educational Leadership, 53* (5), 4-7.

Slavin, R.E. & Fashola, O.S. (1998). *Show me the evidence: Proven and promising programs for America's schools*. Thousand, Oaks, CA: Corwin.

Slavin, R. E., & Madden, N. A. (1993, April). *Multi-site replicated experiments: An application to Success for All.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Slavin, R.E., & Madden, N.A. (1999a). *Success for All/Roots & Wings: 1999 summary of research on achievement outcomes*. Baltimore: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

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

Slavin, R.E., & Madden, N.A. (in press). *One million children: Success for All*. Thousand Oaks, CA: Corwin.

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

Slavin, R.E., Madden, N.A., Dolan, L., Wasik, B.A., Ross, S.M., Smith, L.J. & Dianda, M. (1996a). Success for All: A summary of research. *Journal of Education for Students Placed at Risk, 1*, 41-76.

Slavin, R.E., Madden, N.A., Dolan, L., & Wasik, B.A. (1996b). *Every child, every school: Success for All.* Newbury Park, CA: Corwin.

Smith, L. J., & Ross, S. M., & Casey, J. P. (1994). *Special education analyses for Success for All in four cities*. Memphis: University of Memphis, Center for Research in Educational Policy.

Stringfield, S.C. (1998, August 5). Science, cynicism, and Diogenes' double-edged lamp. *Education Week*, August 5, 1998.

Stringfield, S., Millsap, M.A., Herman, R., Yoder, N., Brigham, N., Nesselrodt, P., Schaffer, E., Karweit, N., Levin, M., & Stevens, R.J. (1997). *Special strategies studies final report*. Washington, DC: U.S. Department of Education.

Traub, J. (1999). *Better by design? A consumer's guide to schoolwide reform*. Washington, DC: Thomas Fordham Foundation.

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

Venezky, R.L. (1994). *An evaluation of Success for All: Final report to the France and Merrick Foundations*. Unpublished paper, University of Delaware.

Walberg, H.J. (1986). Synthesis of research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd edition, pp. 214-229). New York: Macmillan.

.