

An Evaluation of Two Contrasting Approaches for Improving Reading Achievement in a Large Urban District

Karl Skindrud

California State University Dominguez Hills

Russell Gersten

University of Oregon and Instructional Research Group, Signal Hill, CA

Abstract

This independent evaluation of 2 commonly used approaches for accelerating reading achievement and reducing inappropriate special education referrals, Success for All (SFA) and Open Court, was conducted in 12 Title I schools in a large urban district in northern California. To compare the effects of these approaches, we collected data on 936 grade 2 and 3 students over 2 years and 5,694 K through 6 students over 3 years to determine academic and special education enrollment outcomes, respectively. Results supported the prediction that students who used Open Court would outperform those who used SFA on mean SAT9 scores in reading and language but not the prediction that SFA would help students in the bottom quartile of SAT9 score higher or reduce demand for special education services more than Open Court. Neither Open Court nor SFA was associated with reductions in special education enrollment rates, except in Title I schools with the least poverty. A follow-up survey of 17 teachers and an analysis of lesson pacing plans suggested why the teachers saw Open Court as superior on academic outcomes and SFA on social outcomes.

In the past decade, in relatively rapid succession, Congress has supported two quite different approaches to improving reading achievement in low-income, low-achieving schools. Goals of both sets of reform efforts include (a) combining better-quality reading instruction with improved systems for ongoing assessment of student progress and (b) provision of interventions for struggling readers (typically via small-group instruction). Either explicitly or implicitly, the approaches described in the legislation argue that their implementation will result in fewer students from low-income families being inappropriately referred for special education services in reading.

The Comprehensive School Reform Demonstration Program Act of 1998 (CSRD, 1998) provided fiscal incentives for low-achieving schools to adopt whole-school reform models with proven effectiveness. Three years later, Congress enacted the No Child Left Behind Act (P.L. 107-110, 2001); a major part of this legislation was Reading First. Though the goals of Reading First are similar to those of CSRD, their *modus operandi* is different and reflects the effects of two syntheses of research on early reading (National Reading Panel, 2000; Snow, Burns, & Griffin, 1998). Rather than adoption of a school-wide system for teaching reading, language arts, and mathematics, Reading First requires schools to use the empirical knowledge base on reading instruction in its operation. This research base, provided to teachers through ongoing professional development, should be used to guide selection of core reading curricula. In addition, schools are mandated to use valid, reliable screening and progress monitoring to discern which students require intensive early intervention and to provide such intervention.

Whereas Reading First focuses on use of the scientific knowledge base on beginning reading to guide instruction, CSRD stresses better functioning of each school as an organization geared toward instructional improvement. Although in practice there is considerable overlap between the two approaches, there is a significant difference in emphasis. CSRD focuses on the school as a system, whereas Reading First, although not ignoring that change needs to be systematic, highlights the importance of teacher knowledge of research and use of curricula for both instruction and intensive intervention that are based on research. In a sense, these models for school improvement reflect a shift from what was called a management emphasis in the 1990s (American Institutes for Research, 1999) to a curricular emphasis in the 2000s.

The school-wide programs noted in CSRD emphasize classroom and school

management of instruction (e.g., use of progress-monitoring systems, use of cooperative learning groups, flexible groupings across classrooms, no-fault decision making). Among these models are Success for All, the Comer Process, and Accelerated Schools programs. These management models (American Institutes for Research, 1999; Cook, Hunt, & Murphy, 2000) typically restructure critical aspects of schooling but leave the selection of reading curricula to the schools.

Each state has approached Reading First somewhat differently, but all states have developed criteria for adoption of core reading curricula in the primary grades. California has been the most directive, allowing schools to implement only one of two reading curricula, both of which were considered to best represent contemporary research findings.

Given that the more systemic approaches have been advocated for much of the past 20 years (Fullan, 1991; Rowan, Bossert, & Dwyer, 1983), school administrators and faculty have been left to ponder which approach to instructional improvement (management or curricular) will be the best vehicle for improving reading achievement and the costs and benefits of each approach. District administrators also need to consider what evidence there is to support either approach as a long-term strategy for improvement of district-wide reading achievement and/or whether a reasonable combination of the two can be designed and adopted.

This was a dilemma the Sacramento City Schools faced in the late 1990s. To collect evidence on this issue, the district decided to try two approaches and allow an independent evaluation of their effects. The programs selected were among the most popular exemplars of each approach. The first was Success for All (SFA), a school-wide program that emphasizes use of progress-monitoring systems to guide instruction and instructional placements, small class size for reading instruction, intensive early intervention for struggling readers, and use

of cooperative learning principles. The curricular approach was Open Court Reading, a curriculum that attempts to incorporate principles from contemporary scientific research on reading. Unlike SFA, Open Court relies extensively on whole-class instruction, although it does allow teachers to work for a certain length of time during the school day with small groups of students who are struggling. In California, Open Court has benefited from a reduced class size of 20 students, in grades K through 3 statewide, since fall 1998. Thus, SFA emphasizes reorganization of instructional delivery, whereas Open Court focuses on curriculum structure and contemporary reading research.

Nature of the Independent Evaluation

Greenberg and Walberg (1998) noted that both first-party evaluations, conducted by model developers and their colleagues, and second-party evaluations, funded by model developers, create potential conflicts of interest and may be subject to bias. Therefore, Walberg and Greenberg (1999) called for more third-party (disinterested, independent) evaluations of reform models to ensure more objective conclusions and recommendations.

Both the SFA Foundation and California Reading Initiative provided the first author with equal hours of free in-service training in Success for All and Open Court, respectively. However, this study was funded entirely by the Office of Research and Funded Projects at California State University, which has been the first author's sole affiliation since 1972. Because neither author has any affiliation with either SFA or Open Court, financial or otherwise, and neither has ever served as a colleague, classmate, employee, employer, contractor, or student of the developers of SRA or Open Court, we consider this an independent, third-party evaluation. Because of our interest in low-performing students in poverty schools, we intentionally conducted separate analyses

of reading achievement on standardized tests for average students and those in the bottom quartile (i.e., those considered at risk of reading failure).

Overview of the Two Programs

Success for All

The visibility and widespread use of the Success for All program (Slavin & Madden, 2001) is difficult to refute. A total of 1,500 U.S. schools had adopted SFA by 2000 (Slavin & Madden, 2000). Among the 4,648 schools funded by CSRD, SFA was the program most funded, with 431 schools in 2002 (Southwest Educational Demonstration Laboratory, 2003). Historically, SFA, which began in the Baltimore schools in 1987, is the first replicable, transportable, school-wide reading reform model developed for Title I schools in the United States. As shown below, the training, manuals, technical assistance, and program monitoring provided by the SFA Foundation in Baltimore standardize the materials, lessons, and instructional practices nationwide. SFA has served as the model for the CSRD Act and for how to make other school-wide reforms transportable.

Unique components. A distinguishing feature of SFA is that students are grouped homogeneously by reading skills (across classrooms and grades) for 90 minutes of reading/language arts instruction daily, in reading classes of 20–25 students taught outside their homerooms. Students are grouped (and regrouped when appropriate) based on scores on curriculum-based progress tests. Smaller class size is achieved by use of reading specialists and Title I personnel who teach the 90-minute reading classes each morning and provide tutorial services the rest of the day; each student's progress is assessed every 8 weeks, with individual reading tutors assigned to students who are behind. (SFA students receive writing instruction outside their 90-minute reading classes.)

Adoption of SFA requires an 80% vote of approval by the faculty to ensure "buy-in."

SFA prescribes the core reading curriculum only for grades K and 1; the school selects any commercial reading curriculum it prefers for grades 2 through 6. Use of cooperative learning is a key teaching strategy, and a family support team encourages parent involvement regarding problems with attendance, behavior, eyesight, or homework.

Training and technical assistance during implementation are provided by SFA consultants from a regional SFA office. A school program facilitator provides ongoing mentoring, counseling, and materials for teachers. New teachers receive 3 days of training, and new schools three implementation visits during year 1 and two visits in year 2. Regional staff observe program fidelity across classrooms, provide feedback, and secure improvements.

Supporting evidence. In a review of the evaluation research on SFA, Slavin and Madden (2001) reported (a) mean reading gains averaging 0.5 grade equivalents more per year than each of the control grades 1–5, (b) reading effect size gains for students in the bottom quartile more than twice the mean SFA effect size gains, and (c) a reduction in special education placements of up to 50%, relative to comparison schools.

The data were from multisite, replicated evaluations involving more than 6,000 students in 12 large metropolitan areas at each grade level who began SFA by grade 1 between 1988 and 1999. Students in both the SFA and matched comparison schools were given individually administered standardized tests such as the Woodcock Reading Mastery Test (WRMT). Slavin and Madden (2001) reported that effect sizes on mean performance averaged .5 standard deviations across grade levels. They noted, “consistently, effect sizes for students in the lowest 25% of their grades were particularly positive, ranging from 1.03 in first grade to 1.68 in grade four. Again, cohort-level analysis found statistically significant differences favoring low achievers in SFA” (at all levels) (p. 278).

Prior evaluations. Independent evaluations of SFA provided mixed results in stud-

ies of implementation in Charleston, SC (Jones, Gottfredson, & Gottfredson, 1997) and Baltimore schools (Venezky, 1998). Venezky’s (1998) independent evaluation indicated that SFA students scored significantly better than matched comparison group students, yet only 12.5% of the SFA students read at or above grade level as measured by standardized achievement tests after 6 years in the program. Jones et al. (1997) showed that, although SFA students scored significantly higher than matched controls in kindergarten on letter-word identification and word attack on the WRMT, their scores were significantly lower than those of matched comparison students on the grade 1 SAT.

Slavin and Madden (2001) presented data demonstrating a strong positive relation between the fidelity of a school’s SFA program implementation and reading outcomes and argued that, when SFA implementation is erratic, results can be mixed. They also argued that effects are stronger in mature implementations because it often takes teachers several years to master new approaches to teaching. Data from several sites have supported this assertion (Hall & Hord, 2001).

Critics have called for more independent third-party evaluations of SFA to resolve these issues (Walberg & Greenberg, 1999). In addition, studies should focus on schools provided with reasonable implementation support and those in which implementation is assessed systematically (Charters & Jones, 1974; Gersten, Baker, & Lloyd, 2000; Shadish, Cook, & Campbell, 2002). Finally, effect sizes can and do vary depending on the nature of the comparison group (Gersten, 1985; Hoskyn & Swanson, 2000). Therefore, we compared SFA with another potentially powerful early reading intervention that includes professional development.

Open Court

The Open Court Reading Program was developed in 1962; however, in the early

1990s the first- and second-grade material was revised by Marilyn Adams, based on the findings of her recently completed synthesis of research on learning to read (Adams, 1990). In 1996 reading researchers supplemented and reformatted Open Court to strengthen the use of comprehension strategies in this program that already emphasized decoding. Open Court received a good deal of attention after the publication of a large-scale research study by Foorman, Francis, Fletcher, Schatschneider, and Mehta (1998). The purpose of this study, which was conducted in the Houston schools, was to contrast explicit with implicit phonics and phonemic awareness instruction; yet, upon its release, much attention was focused on the explicit phonics curriculum that yielded superior outcomes. At that time, Open Court was one of a small number of reading series that emphasized explicit phonics and phonemic awareness. Publication of the Foorman article was a factor in district adoptions of Open Court in other urban areas (e.g., Sacramento, Baltimore, Los Angeles, and Long Beach [Helfand, 2000]).

Unique components. Open Court is a comprehensive elementary reading program for grades K through 6. In the primary grades, the program makes no assumptions about students' prior knowledge. Phonemic and print awareness and an understanding of the alphabetic principle are taught systematically. The program provides systematic instruction in the areas of decoding, comprehension, inquiry, investigation, writing, spelling, vocabulary, grammar, usage, mechanics, penmanship, listening, and speaking (SRA/McGraw Hill, 1996, 2000).

In the current study, the district used the 1996 version of the curriculum, Open Court Collections for Young Scholars. The series involved 2 hours of whole-class instruction daily followed by 30 minutes of small-group instruction and/or independent work.

Supporting evidence. To date, there are few independent evaluations of Open

Court in the professional literature. The exceptions are the study by Foorman et al. (1998) and one by Westat (2001) in Baltimore. Foorman et al. (1998) reported that students taught with explicit, systematic phonics (Open Court) performed better than those using embedded phonics (a researcher-developed program), which was superior to implicit phonics (whole language) reading programs, on standardized achievement tests in grades 1 and 2. However, Foorman et al. (1998) used word recognition rather than comprehension of connected text as their primary outcome measure. Furthermore, the comparison conditions were relatively weak and untested in the field (in the case of embedded phonics) or conceptually weak in terms of the research literature (in the case of whole language).

Westat (2001) conducted an independent evaluation for the Maryland Department of Education of the 3-year adoption of Open Court in 102 Baltimore city schools, in grades K through 2. Mean second-grade scores improved from a normal curve equivalent (NCE) of 38 to 42 to 45 from spring 1999 to 2000 to 2001. This is equivalent to effect size (d) growth of .33, an effect that is considered moderate to small. During the same period, however, the statewide growth was 6 NCEs. Thus, Westat (2001) concluded they had no evidence that reading growth with Open Court was significantly different from that obtained with other reading programs. Yet, they were aware that the population in Baltimore was among the neediest in the state. In addition, Open Court was the only reading program in the district associated with consistent student improvement in 2001. Also, no statistical analysis of the data was included, and different reading curricula were used above grade 2, which precluded inferences about program effects.

McRae (2002) reported more positive findings for Open Court in California. He evaluated 293 schools that adopted Open Court in 1998 and followed their progress

relative to that of randomly matched control schools with similar demographics for 3 years. Three-year gain scores showed that Open Court schools outperformed non-Open Court schools 50% to 75% in grades 2 and 3 on the SAT9 test, with the largest differences among low-socioeconomic-status schools. Although the McCrae (2002) evaluation is encouraging, it has not been subjected to peer review, does not include statistical analysis of the data, and is considered a second-party evaluation due to publisher sponsorship.

We believe the differing findings from the Baltimore (Westat, 2001) and California (McCrae, 2002) evaluations are due to differences in fidelity of program implementation. The majority of California Open Court schools received extra funding from the Hewlett-Packard Foundation to ensure pre-service training and weekly classroom supervision from a reading coach (McCrae, 2002). In the Westat (2001) study, only about 52% of Baltimore teachers received "training . . . in the new reading curriculum" (p. 59).

Purpose of the Study

As we noted, Walberg and Greenberg (1999) called for more third-party evaluations of education-reform models in general, and SFA in particular. The study reported in this article was designed as an independent evaluation of these two widely advocated approaches for improving reading achievement. One reason we compared the two reading programs (SFA and Open Court) as packaged by their developers and enhanced by state practices is that schools generally want to know what the outcomes are when such programs are implemented as developed and/or supplemented. There is also a growing demand for programs to provide unbiased, scientific evidence of their performance in the field as called for by the What Works Clearinghouse of the U.S. Department of Education (2005). Hence, we compared these two programs, although daily time allocations of the programs are not

closely matched in the Sacramento City Unified School District (SCUSD).

This district seemed an excellent venue for such a comparative evaluation because it (a) simultaneously implemented both reading programs district-wide in the fall of 1997, (b) pretested all students in fall 1997 before implementation; and (c) obtained funding for reading coaches in all elementary schools for both approaches; and (d) schools in the evaluation sample achieved implementation typical of developer-supervised sites on measures designed by program developers. Also, (e) because all 59 elementary schools used one of the two programs, we were able to find closely matched comparison schools.

Method

Design

We primarily used a quasi-experimental design, with two samples of schools carefully matched on demographics and prior reading achievement. Data were analyzed separately in each of two cohorts starting with reading programs at the second- (Cohort 1) and third-grade levels (Cohort 2) over 2 years for students with both mean and lowest-quartile scores on standardized tests. We also examined special education enrollment rates per school in grades K through 6 for schools using the two models over a 3-year period. Two years after data collection, a follow-up teacher survey was conducted to compare the perceptions of a subset of 17 teachers, who were trained and experienced in both SFA and Open Court, on the most effective components of each reading reform. Because by fall of 1997 all 59 SCUSD elementary schools had been mandated to adopt SFA or Open Court as part of the initiation of district-wide reforms, we could not find a set of schools to serve as a pure control group.

Our analyses had three objectives and predictions. The first was a comparison of achievement growth in reading and language on a standardized group achievement test. We also were interested in whether program outcomes for students in the low-

est quartile of each school were stronger or weaker than those for the whole school population. Third, we looked at the effect of both reading programs on special education enrollment rates school-wide. We predicted that Open Court would excel on the first objective and SFA on the second and third objectives because of Open Court's curriculum focus and SFA's homogeneous grouping, cooperative learning, and tutoring emphases. Note, however, that in objective 3 we were limited to an interrupted time-series design. This design reveals only associational, not causal, relations. Objective 3 has been retained because of the increasing interest, under NCLB and Reading First, in whether models for improving classroom reading also reduce the rate of special education placements in the learning disabilities category.

Participants

We ranked all 59 SCUSD regular elementary schools on the district's Title I poverty criteria: (percentage free and reduced-price lunch plus percentage on AFDC [Aid to Families with Dependent Children] / 2). Each of the four SFA schools was matched with the two Open Court schools ranked just above and below it on the Title I poverty criteria. (One of the five SFA schools in SCUSD was excluded because it began reading reforms in fall 1995, 2 years early, giving it an unfair advantage.) The schools selected ranged from 58% to 86% on district poverty criteria. The three samples drawn from these 12 schools were the (1) academic outcomes sample ($N = 936$, grades 2 and 3), (2) special education enrollment sample ($N = 5,694$, grades K–6), and (3) follow-up teacher survey sample ($N = 17$, grades 1–6 teachers in four SFA schools that converted to Open Court after 2000). Each sample is described in the Academic Outcomes Sample and/or Measures sections below.

Academic outcomes sample. This student reading sample consisted of 936 second- (Cohort 1) and third-grade (Cohort 2) students continuously enrolled in the 12

schools studied from fall 1997 through spring 1999. Students in grades K and 1 were not included in the study because they were not routinely tested at that time. Also, data collection was limited to 2 years due to annual average cumulative student mobility (15%) and missed tests (6%) increasing sample attrition to unacceptable levels.

ITBS at pretest: The mean NCE Reading Total pretest scores on the Iowa Tests of Basic Skills (ITBS) for the four SFA and eight Open Court schools were 35.9 and 30.1, and standard deviations were 22.0 and 16.4, respectively, when data from grades 2 and 3 were combined. The difference of 5.8 NCE scores had an effect size (d) of .30 in favor of the SFA group. Consequently, we used analysis of covariance to statistically equate the SFA and Open Court samples.

English proficiency: The mean percentages of students in each of the four SFA and eight Open Court schools, labeled English language learners (ELLs) in spring 1999 (California Department of Education, 1999), were 38.0% and 42.4%, respectively, with a mean difference of 4.4%. This difference favored SFA with a lower ELL rate. However, a t -test of this difference was not significant ($t = 0.37$, $df = 10$; $p = .72$). The effect size (d) of the difference was quite small, .014, indicating no practical differences on English proficiency.

Sample attrition: We removed 29% of the cases where students tested in fall 1997 were not enrolled for either of the two post-tests. This action removed students who had either moved into or out of a treatment (SFA) or comparison (Open Court) school during the study. We also excluded an additional 12% of the students who were absent on any day of the three test administrations and were not provided with an additional testing occasion. The mean student attrition rates over 2 years for the SFA and Open Court schools, respectively, were identical at 41% each. These are moderate attrition levels for most urban schools in our experience. The mean NCE pretest reading achievement scores for excluded

SFA and Open Court students were 18.3 and 19.7, with standard deviations of 19.3 and 19.1, respectively. A *t*-test of the difference between means ($t = .94$, $df = 676$; $p = .35$, *ns*) ruled out sample bias between the reading achievement scores of SFA and Open Court students removed from the sample.

However, the SFA and Open Court means on the pretest for the general population prior to subject attrition ($N = 1,614$) were significantly lower than the sample means after removal (28.49 vs. 35.87 with $d = +.33$, and 25.74 vs. 30.09 with $d = +.25$, for the population and sample means within the four SFA and eight Open Court schools, respectively). This does not affect sample bias, but it does prohibit generalization of results from the more mobile and absent students.

Grade-level retention: Including retained students in year 2 of a study has been shown to inflate the bottom 20% of test scores (Maddahian, 2002). In addition, if retention rates differ between samples, results will differentially inflate sample test scores. Thus, to prevent bias, we removed all retainees, which averaged only 2%, across the reading population. It is interesting to note that, first, a disproportionate number of retainees (21 of 34, 62%) fell in the bottom quartile of the study sample; and, second, in the bottom quartile, there was a disproportionate number of retainees in the Open Court schools (i.e., 17 of 21, 81% [vs. 67%] and 4 of 21, 19% [vs. 33%] in the eight Open Court and four SFA schools, respectively). We attribute this trend to SFA's homogeneous skill grouping across grade levels during reading periods, which accommodates low achievers. Fortunately, retention removal prevented bias here, which would have favored the Open Court schools. The final reading sample size, after purging all "leavers," "absentees," and "retainees," was 936 of 1,648, or 57% of the original student population.

Teacher characteristics and teacher-pupil ratios. Approximately 110 grade 2, 3, and 4

teachers participated in this study from fall 1997 to spring 1999 (36 teachers in the four SFA and 74 teachers in the eight Open Court schools). We compared teachers in the two sets of schools on several commonly used indicators of teacher quality to see if they differed significantly across reforms: (a) teaching experience, (b) certification status, and (c) formal education. If so, these factors could confound the results of reading reforms with those due to teacher characteristics. In addition, we examined the teacher-to-pupil ratios in the two sets of schools. We used DataQuest (California Department of Education, 1998) to access the information for the 1997–1998 school year.

There was no significant difference in teaching experience ($t = .70$, $df = 10$; $p = 42$, *ns*). Teachers in both sets of schools were fairly experienced; respective means were 13.8 years for SFA and 10.25 for Open Court. Credential status was coded: full credential (with student teaching) = 3, emergency or intern credential (without student teaching) = 2, no credential/waiver = 1. Mean credential status in fall 1997 was virtually identical for the two samples: 2.87 for SFA and 2.89 for Open Court ($t = -.6$, $df = 8.6$; $p = .58$, *ns*). These data indicated that between 11% and 13% of the teachers were on emergency licenses; the rest were fully certified. Mean educational level was coded doctorate = 5, MA + 30 hours coursework = 4, MA = 3, BA + 30 hours = 2, BA = 1. Again, differences were not significant. Mean education level was 2.23 for SFA and 2.14 for Open Court teachers ($t = .53$, $df = 10$; $p = .61$, *ns*). Comparisons of teacher-to-pupil ratios in 1997–1998 were 1 to 21.5 in SFA classrooms and 1 to 22.5 in Open Court classrooms ($t = .83$, $df = 10$; $p = .42$, *ns*). None of these teacher characteristics or teacher-pupil ratios was significantly different across the treatment samples. However, where there were differences, they tended to favor the SFA schools.

Measures

Academic outcomes. The academic outcome measures were the Reading and Lan-

guage subtests of the Stanford Achievement Test, 9th edition (SAT9) from spring 1998 and 1999. Again, we included the SAT9 language subtests to assess transfer of reading effects to total language (mechanics, usage, content, and organization) on the SAT9 (Harcourt Brace, 1996).

Special education services demands. For the analysis of special education outcomes, we used the total K through 6 enrollment of 5,694 students in 1998 across the 12 schools. We had the SCUSD special education division tabulate the number of students during each school each year who had (a) initial psychological services reports, (b) initial placements in special education learning disability (LD) classes or resource programs, and (c) official enrollment in the school's special education resource program. This search covered the 5- to 6-year period between fall 1994 (or 1995 for *c* only) and spring 2000. A 2- to 3-year baseline before the programs began was compared with a 3-year intervention period after programs began. For example, special education resource program enrollment as a percentage of school enrollment included all K through 6 students assigned to the resource program (for mild disabilities) in each school, divided by the total school enrollment. The percentages before and after the onset of reading reforms were then compared to assess possible treatment effects in an interrupted time-series design.

Follow-up survey of teachers experienced in both SFA and Open Court. In fall 2002, 2 years after completion of this evaluation, the first author returned to the district to survey a subset of teachers who had by that time been trained and taught in both the SFA and Open Court programs. This survey assessed teachers' perceptions of seven variables: (1) characteristics of respondents, (2) school progress, (3) time allocations, (4) management strategies, (5) lesson presentation techniques, (6) curriculum structure, and (7) implementation support. The survey concluded with four open-ended questions on teachers' opinions re-

garding (8) significant program differences, and causes of (9) academic, (10) special education, and (11) social outcomes. The purpose was to obtain teachers' insights into which components of the two reading programs (SFA and Open Court) were more responsible for differences in the academic and special education outcomes determined earlier (and any unmeasured outcomes). (See the Follow-up Teacher Survey results in Skindrud & Gersten, 2005, App. B.)

Implementation

Success for All. The materials and guidelines for instruction in grades 1, and 2 through 4, respectively, used in the SFA classrooms in the Sacramento City Schools from 1997–2000, the span of this study, can be found under Reading Roots (grade 1), and Reading Wings (grades 2–4) in Slavin, Madden, Dolan, and Wasik (1996, chap. 2). The professional development procedures and the roles of the school SFA facilitator and the regional consultants in maintaining program integrity are also outlined in Slavin et al. (1996, chap. 7), as are the SFA tutoring procedures (chap. 3). The visiting SFA consultants' observed rates of implementation fidelity in the four SCUSD SFA schools were then compared to the national norms for "total average" implementation of SFA, as explained under "SFA Fidelity" below and described in Skindrud and Gersten (2005, App. C).

Both the Roots and Wings levels of the SFA program require teachers to provide "additional writing instruction . . . outside of the 90-minute reading block using SFA writing or a comparable program" (SFA Implementation Visit Record Form pp. 11 and 16, SFA Foundation, 1999). All four SCUSD SFA schools were rated as doing so in 1998 and 1999 by their SFA training consultants. However, SCUSD SFA schools frequently conducted spelling and grammar, as well as SFA writing and tutoring, outside the 90-minute SFA reading block because all the SCUSD Open Court schools were spending

150 minutes daily on reading and language arts instruction.

Open Court. Teacher materials included *Open Court Collections for Young Scholars* (SRA/McGraw Hill, 1996), 43 sound spelling cards, a teacher's guide, and transition and review, or phonics review, designed to provide supplemental instruction in decoding skills for students in grades 2 and 3 who had no experience in Open Court. However, Sacramento Unified chose a more intense 2-year Open Court transition plan for grades 2 and 3 in their Title I schools only. Because these students were starting Open Court "midstream" in fall 1997, a condensed sequence of grades 1, 2, and 3, or 1, 2, 3, and 4 program levels was started in fall 1997 for grades 2 or 3, respectively. At the end of this 2-year "catch-up" (fall 1997–fall 1999) period, Title I students who started programs in grades 2 and 3 were finally placed in grade-appropriate materials.

During the second year, the district used informal reading assessments every 6 to 8 weeks at the end of each unit. Five areas were assessed: (a) word or oral reading fluency for 1 minute, (b) spelling, (c) reading comprehension, (d) vocabulary, and (e) punctuation and grammar. Action plans were written for students who had not reached unit criteria, but these did not include tutoring.

Staff support included a full- or half-time reading coach at each Open Court school depending on school size; there were 4 days of basic grade-level training for new Open Court teachers in year 1 followed by 4 days of advanced grade-level training in year 2. Open Court experts met monthly with reading coaches and administrators to resolve problems and refine supervision skills.

Results

We begin with a description of implementation fidelity. We then present findings on differential achievement patterns in reading and language. This is followed by the analyses of students in the lowest quartile and special education referrals. Finally, the teacher survey responses are presented.

Implementation Fidelity

SFA. We obtained the implementation ratings for all four SFA schools in the study, using the SFA Implementation Visit Record Form from the SFA Foundation (1999) in Baltimore. SFA training consultants from outside the school district conducted the observations within one semester of the end of the first 2 years of model adoption. The respective teacher observation ratings in each of the four SFA schools were 113, 113, 81, and 92, averaging 99 points, where 100 points were what SFA found to be mean levels of implementation by SFA national norms. Thus, these four schools represented typical levels of SFA implementation with full SFA monitoring. (See "Sample SFA Implementation Visit Form," 1999 revision, in App. C of Skindrud & Gersten, 2005.)

Open Court. The school reading coaches collected implementation observations in Open Court schools, but we were denied access to these data due to a confidentiality agreement with the teachers' union. However, the district's Open Court reading coordinator (S. Van Vleck, personal communication, Nov. 15, 2000, and Feb. 8, 2002) reported that, although there were a variety of initial implementation problems the first semester (fall 1997), none of the eight Open Court schools in this study had implementation problems by the end of year 2 of the reforms (spring 1999). The missing data on Open Court implementation are a possible confound but are not a major problem given this study's outcomes.

Patterns of Student Achievement

Mean results of the 2-year, longitudinal analysis are presented in Tables 1 and 2. The covariance-adjusted means are displayed separately for Cohorts 1 and 2, starting programs in grades 2 and 3, respectively. Two by two analyses of covariance (ANCOVAs), with one between-subjects factor (reading programs), one within-subjects factor (year of test), and ITBS scores for fall 1997 as a covariate (Keppel & Zedeck, 1989), were conducted by combining the SAT9 scores

TABLE 1. Covariance-Adjusted NCE Reading Scores for Success for All and Open Court Reading Programs for Two Cohorts of Students in the Sacramento City Schools, 1997–1999

Cohort/Grade	Test	Success for All				Open Court				<i>d</i>
		<i>N</i>	<i>M</i>	<i>SD</i>	%	<i>N</i>	<i>M</i>	<i>SD</i>	%	
Cohort 1:		142				292				
Pre-2	ITBS		30.5	21.7	18		30.5	17.0	18	
2	SAT9		37.2	16.8	27		44.3	17.1	39	+ .41
3	SAT9		38.6	18.5	29		43.9	16.5	39	+ .30
Cohort 2:		152				350				
Pre-3	ITBS		33.1	22.1	21		33.1	15.9	21	
3	SAT9		38.8	21.6	30		37.8	15.2	28	– .05
4	SAT9		40.8	23.0	33		42.8	17.3	37	+ .10

NOTE.—ITBS = Iowa Tests of Basic Skills.

TABLE 2. Covariance-Adjusted NCE Language Scores for Success for All and Open Court Reading Programs for Two Cohorts of Students in the Sacramento City Schools, 1997–1999

Cohort/Grade	Test	Success for All				Open Court				<i>d</i>
		<i>N</i>	<i>M</i>	<i>SD</i>	%	<i>N</i>	<i>M</i>	<i>SD</i>	%	
Cohort 1:		142				286				
Pre-2	ITBS		28.7	24.0	16		28.7	14.7	16	
2	SAT9		34.9	20.5	24		40.6	19.3	33	+ .29
3	SAT9		40.6	18.1	33		47.5	16.6	45	+ .40
Cohort 2:		142				241				
Pre-3	ITBS		30.4	22.3	18		30.4	16.3	18	
3	SAT9		39.9	21.0	32		38.6	15.4	29	– .07
4	SAT9		42.7	19.9	36		45.4	17.8	41	+ .14

across the four SFA and eight Open Court schools in the study. We used normal curve equivalent (NCE) scores on the SAT9 as the dependent measure for the analysis. Separate analyses were performed for Cohort 1 and Cohort 2 reading scores in Table 1 and then for students within each cohort for the language scores in Table 2. We repeated these four ANCOVAs for the bottom-quartile scores in Tables 3 and 4. The mean NCE scores were converted to percentile ranks in each table for ease of interpretation. We made Bonferroni corrections to adjust *p* levels for multiple comparisons. (For a complete summary of each of these four ANCOVAs on mean and bottom-quartile academic outcomes, see Skindrud & Gersten, 2005, App. D.)

Mean reading and language measures.

The 2 × 2 ANCOVA on SAT9 reading scores showed a significant main effect, $F(1,$

431) = 31.6, $p < .001$, favoring Open Court for Cohort 1, and a significant interaction, $F(1, 499) = 7.8, p < .005$, for Cohort 2. Simple effect results favored Open Court in the second year of the program. Thus, in two of the four instances, Open Court students' reading achievement was significantly better than that of similar students in SFA. For Cohort 1, effect sizes were .41 for second grade and .30 for third grade, in the low to moderate range. For Cohort 2, effect sizes were minimal: –.05 for grade 3 and +.10 for grade 4. This suggests that beginning Open Court in second grade was an effective means of improving reading achievement, whereas beginning in third grade led to minimal effects.

The 2 × 2 ANCOVA for language scores demonstrated a significant main effect between programs for Cohort 1, $F(1, 425) = 25.3, p < .001$, indicating that Open Court

TABLE 3. Covariance-Adjusted Scores (in NCE units) in Total Reading for Bottom Quartile of Success for All and Open Court Reading Program from Two Cohorts of Students in the Sacramento City Schools, 1997–1999

Cohort/Grade	Test	Success for All				Open Court				<i>d</i>
		<i>N</i>	<i>M</i>	<i>SD</i>	%	<i>N</i>	<i>M</i>	<i>SD</i>	%	
Cohort 1:		35				73				
Pre-2	ITBS		9.7	9.7	3		9.7	7.6	3	
2	SAT9		25.8	5.9	12		33.6	13.7	22	+.73
3	SAT9		25.4	14.2	12		34.6	13.1	23	+.67
Cohort 2:		38				88				
Pre-3	ITBS		12.2	9.4	4		12.2	7.7	4	
3	SAT9		21.8	12.7	9		25.4	10.6	12	+.31
4	SAT9		24.4	11.9	11		29.7	13.0	17	+.43

TABLE 4. Covariance-Adjusted Scores (in NCE units) in Total Language for Bottom Quartile of Success for All and Open Court Reading Program from Two Cohorts of Students in the Sacramento City Schools, 1997–1999

Cohort/Grade	Test	Success for All				Open Court				<i>d</i>
		<i>N</i>	<i>M</i>	<i>SD</i>	%	<i>N</i>	<i>M</i>	<i>SD</i>	%	
Cohort 1:		39				75				
Pre-2	ITBS		9.1	6.8	3		9.1	7.9	3	
2	SAT9		22.5	11.3	10		29.8	16.3	17	+.52
3	SAT9		29.5	10.2	16		38.3	14.5	29	+.52
Cohort 2:		36				61				
Pre-3	ITBS		9.3	8.9	3		9.3	8.1	3	
3	SAT9		28.8	12.3	16		29.6	12.8	17	+.06
4	SAT9		31.8	12.5	19		33.0	14.6	21	+.09

students achieved higher at both year-end assessments. For Cohort 2, we found a significant interaction between program and grade level, $F(1, 380) = 9.0, p < .003$. Simple effects revealed that the difference was only significant for the second year of implementation. Effect sizes were moderate for Cohort 1 (.29, .40) and weak for Cohort 2 (–.07 for the first year, .14 for the second year), consistent with the effects for mean reading scores above.

Statewide mean gains. According to California Department of Education (2002) reports, during this 2-year period, the statewide grade 2 to 4 student cohort's mean reading gains were +6 national percentile ranks (NPRs) on the SAT9. In contrast, the gains for the students in Cohort 1 during these years were approximately 11 NPRs for SFA and 21 NPRs for Open Court. We have no way of precisely equating these scores because the students in this school district were pretested on the Iowa Tests of

Basic Skills and then posttested both years on the SAT9, and the National Research Council (1999) found that these two tests are not equivalent in terms of norms. Nonetheless, our data suggested that the gains in both sets of schools surpassed the state average.

Bottom-quartile reading and language scores. Tables 3 and 4 present results only for the bottom quartile of each school; the subsamples were derived from ITBS scores collected prior to the start of the study. The 2×2 ANCOVA produced a significant main effect between programs in bottom-quartile reading for both Cohort 1, $F(1, 105) = 15.6, p < .001$, and Cohort 2, $F(1, 123) = 5.5, p < .02$, and for bottom-quartile language for Cohort 1, $F(1, 111) = 14.8, p < .001$, favoring Open Court. This is a somewhat stronger pattern of effect sizes than we found for the full reading sample (i.e., Cohort 1 = +.73 and +.67; Cohort 2 = +.31 and +.43).

TABLE 5. Special Education Resource Program Enrollment in Title I Schools with Most and Least Poverty Before and After Implementation of Reading Programs in 1998

Poverty Level and Reading Program/School	Enrollment	Free and Reduced Lunch (%)	Resource Program Enrollment as % of School Enrollment					rho	p	
			Before Implementation (Rank)	1996	1997	1998	1999			2000
Schools with most poverty (N = 3):										
Open Court	676	86	2.1 (1)	3.3 (2)	4.3 (4.5)	3.4 (3)	4.3 (4.5)	3.4 (3)	.9	.04
SFA	351	86	3.7 (1)	4.3 (2)	5.4 (4)	6.8 (5)	4.6 (3)	6.8 (5)	.9	.04
Open Court	567	84	3.8 (2)	3.2 (1)	4.2 (3)	4.8 (4)	5.8 (5)	4.8 (4)	.8	.10
Schools with least poverty (N = 3):										
Open Court	398	59	4.5 (5)	3.5 (4)	3.3 (3)	1.5 (1)	3.0 (2)	1.5 (1)	-1.0	.001
SFA	441	58	5.7 (4.5)	5.7 (4.5)	5.0 (3)	4.5 (1.5)	4.5 (1.5)	4.5 (1.5)	-.95	.02
Open Court	258	58	4.3 (5)	2.7 (3)	3.9 (4)	2.3 (1)	2.3 (2)	1.6 (1)	-.9	.04

NOTE.—Total of n = 12 Title I schools; SFA = Success for All.

The academic outcomes for the bottom quartile, generally favoring Open Court, were contrary to our prediction and to the developer's claims. Bottom-quartile students using Open Court significantly exceeded the reading and language gains of SFA students in Cohort 1 (from grades 2 to 3) and the reading gains of SFA students in Cohort 2 (from grades 3 to 4).

Special Education Resource Program Enrollments (K–6)

We compared each school's demand for special education services with its baseline rate just prior to reading reforms. Comparing each school to itself helps control for context effects. Neither SFA nor Open Court was associated with changes in any school's initial psychological evaluation or initial special education placement rates from baseline to reform. However, both SFA and Open Court were associated with significant reductions in the percentage of students enrolled in special education resource programs in the three sample schools with the least poverty. Both reforms were also associated with significant increases in the proportion of students enrolled in resource rooms among the three schools with the most poverty. (For all special education outcomes, see Skindrud & Gersten, 2005, App. E.)

Teacher Survey

We conducted a follow-up teacher survey in the district 5 years after the initiation of district reading programs. In fall 2002, we distributed the 32-item survey to one teacher at each grade level, 1 through 6, in what had been the four SFA schools participating in this evaluation ($N = 24$ teachers altogether). When these four schools had completed their 3-year contract with SFA, they converted to Open Court. These conversions complied with new statewide reading/language arts curriculum adoptions in January 2002. California did not include SFA as one of the grade K through 6 adopted reading programs. Switching to

Open Court was necessary for a school to qualify for additional state literacy funds. Seventeen of these 24 SCUSD teachers responded to this survey for a 71% return rate. The teachers, who had at least 2 years of experience each with SFA and Open Court, attributed the superior academic outcomes with Open Court to a variety of factors, not just to greater time allocations to reading. These included more rapid pacing of reading skills and more consistent review of phonemic awareness, decoding, and reading fluency in grades K through 6; more specific reading comprehension and writing programs, grades K through 6; and more extensive use of decodable and authentic readers. They attributed the equal effects of SFA and Open Court on special education outcomes to different features of each program: to SFA's greater accommodation of student differences (e.g., grouping students according to their skills), and to Open Court's greater effect on students with bottom-quartile test scores. Teachers also felt that SFA excelled on social outcomes due to its emphases on cooperative learning, teamwork, conflict resolution, and social reinforcement strategies. Sixty percent rated SFA higher, 5% rated Open Court higher, and 30% rated both the same on social outcomes. These teachers generally saw SFA as the more socially effective and Open Court as the more academically effective reform model (see Skindrud & Gersten, 2005, App. B).

Discussion

Analyses indicate that Open Court produces significantly better achievement outcomes than SFA in both reading and language SAT9 scores when the programs are started in grade 2. Effects are moderate. For students who begin in third grade, effects are small and only present in fourth grade. The superior performance of Open Court over SFA on standardized test scores is consistent with the mixed (Jones et al., 1997; Venezky, 1998) results of SFA and superior (Foorman et al., 1998) results of Open Court

relative to other core reading programs in prior independent evaluations. The delayed mean gain when Open Court was initiated in third grade may be due primarily to the intensive transition plan used in the Title I schools in the sample.

The strongest effects for the Open Court program are for students in the lowest quartile. Effect sizes in reading in Cohort 1 are a strong .73 in second grade and .67 in third grade for students in the lowest quartile, yet a modest .41 and .30 for the full sample. For Cohort 2 reading, effects are a moderate .31 for third grade and .43 for fourth grade for the lowest quartile, yet we obtained only nonsignificant effects of $-.05$ and $+.10$ for fourth grade for the whole sample. This is somewhat surprising in that SFA stresses cooperative learning and other techniques to increase active engagement in the mornings and places greater emphasis on afternoon tutoring and other interventions for lagging students. In addition, Open Court implementation relies heavily on large blocks of whole-class instruction, with all students working on grade-level curriculum. Yet, in the lowest quartile, the effects are consistently larger for students enrolled in Open Court.

These findings support the interpretation that the quality of the curriculum tools teachers use may be critical to students' success, especially for lower-achieving students, because this is one of the major differences between the programs from grades 2 through 6. Interestingly, a recent comparison of SFA and Open Court by district evaluators in the Los Angeles City Schools (Maddahian, 2002) also showed that Open Court students outperformed SFA students, "especially for low initial performing students in grade 2 across schools of similar poverty levels" (p. 9).

Special Education Resource Program Enrollment Rates

It is clear that these core reading reforms, implemented under close supervision, do not generally reduce demands for

special education services in typical Title I, urban schools. What is different about the six of 36 cases in this study where there were noticeable decreases (or increases) in special education services per school during reading reforms? First, these associations only occur with special education resource program enrollment rates where disabilities are the mildest (i.e., learning disabilities). Second, these core reading reforms are only associated with reductions in resource program enrollments in Title I schools that have moderately low poverty levels (i.e., free-lunch rate below 58%) and with increases in resource program enrollment rates in schools with high poverty levels (i.e., free-lunch rates above 84%). Third, both reforms have similar outcomes, suggesting that the results are not random. These trends, along with the disproportionately lower grade-level retention rates in SFA schools in this study, suggest that (a) core reading reforms under the best of circumstances may help poor or at-risk students but generally not students with both severe poverty and identified learning disabilities; (b) students handicapped by both learning disabilities and severe poverty will be identified earlier than most, due to the 6- to 8-week progress tests routinely used in SFA and Open Court; and (c) the new Individuals with Disabilities Education Improvement Act (IDEA) of 2004 (2005) rules, which allow response to intervention (RTI) models for special education enrollment, may be necessary to reduce the proportion of students with learning disabilities in most Title I schools.

Impressions of Program Accommodations

There are several possible explanations why SFA appeared to accommodate retainees more than Open Court did in this study. First, Open Court program lessons generally move faster than SFA lessons in the grade 1 materials by covering all 43 sound-symbol cards in half a year, and, in SCUSD, by covering grade levels 1, 2, and 3, or 1, 2, 3, and 4, in the first 2 years of Open

Court in the two cohorts, respectively. Second, the whole-class, heterogeneous grouping of students for 80% of reading time makes the Open Court program less “friendly” than SFA to at-risk students in regular classrooms. The saving grace for at-risk students in Open Court is the “independent work time,” the last 30 minutes daily. This is the only class time during which Open Court clusters students for small-group instruction or individual reading tasks. By contrast, SFA always homogeneously groups students by reading skills (across grade levels and classrooms), assigns students to pairs or cooperative learning teams of four or five, and provides individual tutoring as needed. This makes it easier to mainstream students with special needs into SFA classrooms. Third, SFA adds motivational strategies (e.g., class celebrations, cheers, team competitions, and other incentives). This gives at-risk students a home in the classroom and peer support from the beginning. One can imagine the teacher appeal of the more supportive management emphasis of SFA. One question remains: Are SFA accommodations for at-risk students only temporary? More research is necessary to clarify this issue.

Two Fundamental Questions

This study raises two questions. First, why is Open Court superior to SFA on most academic outcomes? Some will say it is because of the extra time spent on reading/language arts by Open Court. However, equally likely explanations are the clear differences in curriculum control, content, and lesson pacing. SFA has much less control over curriculum content and lesson pacing than Open Court. Open Court emphasizes explicit teaching of phonemic awareness, decoding, and comprehension in grades K–6. SFA emphasizes intensive instruction in homogeneous classes and use of cooperative learning in grades K through 6 (and tutoring for slow primary students), but only controls reading curriculum for such students in grades K and 1. As we argue below,

instructional quality may be more critical than instructional quantity. Analysis of the lesson-pacing plans for both Open Court and SFA (Skindrud & Gersten, 2005, App. A) reveals the following differences, with SFA writing being done in the homeroom: (a) Increased instructional time in Open Court (150 minutes daily vs. 90–150 minutes daily in SFA). (b) Heterogeneous versus homogeneous skill grouping in Open Court and SFA, respectively. (c) Intensive emphasis on decoding and phonemic awareness instruction in Open Court, two to three times the daily rate of SFA in our analysis of typical grades 2 through 4 lesson-pacing plans. (d) Linguistic control of readers and decoding instruction: Open Court prescribed specific decodable and/or anthology readers for grades K through 6; SFA only prescribed these readers for grades K and 1, leaving random selection of the grades 2 through 6 reading curriculum to each district as noted above. Furthermore, SFA had less control over the 60 minutes of supplementary writing, spelling, and/or grammar instruction in home classrooms. (e) Programming of critical decoding skills: Open Court was more “cumulative” (approximately daily), whereas SFA was more “spiral” (approximately weekly) in its programming of skills (e.g., daily blending by sounds, syllables, words, or sentences in decodable readers at progressive grade levels in Open Court vs. only weekly, and at grades K and 1, in SFA). Implementation of comprehension instruction was more balanced across the two reforms and thus less likely to explain differing outcomes.

In conclusion, there are too many differences in curriculum control of content, sequencing, pacing, and delivery between the programs in this study to target any one factor as the sole cause of the reading outcomes. All the factors noted above are worthy of consideration. The importance of curriculum content and quality of implementation can be seen in the results of a recent observational study of Open Court (2000, 2002 editions) conducted by Slayton,

Oliver, and Burley (2003). This classroom observational study included 44 grade 2 and 44 grade 3 teachers randomly selected from the Los Angeles Unified School District during the first 3 years of district-wide Open Court adoption. Among the findings were that quality of instruction was directly related to student reading outcomes. Slayton et al. (2003) stated, "as the quality of teacher pedagogy improved [on program standards] so did students' scores on both . . . reading and language . . . when we controlled for ethnicity, language program, meal program, teacher credential and teachers' years of experience" (p. 104). This was especially true for African American and English-only students and for teachers without credentials ($r = .34, p < .01$) and with either fewer than 4 years of experience ($r = .33, p < .01$) or more than 17 years ($r = .46, p < .01$).

Slayton et al. (2003) concluded, "We did find that the amount of time spent on reading/language arts . . . was not a significant predictor of student achievement" (p. 26), and that "the quality of pedagogy provided, during [Open Court] Reading and Responding instructional time, was a better predictor of student achievement" (p. 26) in spite of wide variation in the time teachers spent on reading/language arts instruction (from 79 to 200 minutes daily). In our own research (Gersten, Baker, Haager, & Graves, 2005) we found a strong relation between quality of classroom instruction and student growth in reading achievement; correlations were .73 for first-grade reading.

Common complaints about Open Court Reading are that it takes too much time and reduces the time for social studies and science. Defenders of Open Court argue that there is social studies content in Open Court and that improving reading in the lower grades will improve access to content in the upper grades. Some hope that more efficient approaches to reading and language arts instruction will eventually improve or replace the current editions of Open Court.

A second question this study raises is

how schools can reduce inappropriate special education placements due to reading failure. Recall that neither of these core reading programs was associated with dramatic decreases in special education referrals. Beginning reading reforms in grades 2 or 3 appears to be too late to lead to a significant reduction in special education placements in this study. However, where core reading reforms are insufficient, special interventions for struggling readers may be necessary to significantly reduce inappropriate special education referrals. Numerous authors (Coyne, Kame'enui, Simmons, 2001; Mathes et al., 2005; O'Connor, Notari-Syverson, & Vadasy, 1996; Shaywitz, 2003; Snow et al., 1998; Torgesen et al., 2001; Vellutino et al., 1996) have noted how core reading programs can be supplemented with reading interventions for 15 to 40 minutes a day to help struggling readers. Early reading interventions, in grades K through 2, are now called for in special education legislation (IDEA, 2005). This may reduce inappropriate special education referrals more than either SFA or Open Court alone, as originally conceptualized by their program developers.

Limitations

Field evaluations often lack internal validity, but have fewer problems with external validity. This quasi-experimental field evaluation is no exception. Concerning internal validity, we evaluated and controlled the most feasible demographic confounds. However, the following remain as limitations: (a) lack of access to objective measures of Open Court implementation fidelity, (b) lack of control over instructional time and reading curriculum content in SFA in grades 2 through 4, and (c) evaluation of special education outcomes was limited to an associational (vs. causal) research design. With regard to external validity, generalization of our results is most relevant to Title I schools with 58% to 86% poverty rates, 14% to 47% English language learners, implementation fidelity typical for schools with

developer-recommended training and supervision, and schools starting both reading reforms in grades 2 and 3 with Open Court transition plans to bridge the grade 1 through 4 gap in beginning reading skills.

References

- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- American Institutes for Research. (1999). *An educator's guide to school reform*. Arlington, VA: Educational Research Service. (Also available on-line at <http://www.ers.org>)
- California Department of Education. (1998). DataQuest data on "Teacher credentials and experience" and "Staff education" in 1998, and "English learners by CA schools" in 1999 [On-line]. Available: <http://data1.cde.ca.gov/dataquest/>
- California Department of Education. (1999). DataQuest data on "English learners by CA schools" in 1999 [On-line]. Available: <http://data1.cde.ca.gov/dataquest/>
- California Department of Education. (2002). Standardized testing and reporting (STAR) [On-line]. Available: <http://star.cde.ca.gov>
- Charters, W. W., Jr., & Jones, J. E. (1974). *On neglect of the independent variable in program evaluation: Project MITT*. Eugene: University of Oregon.
- Comprehensive School Reform Demonstration Program (CSRDP) Act of 1998. (1998). [On-line]. Available: <http://www.wested.org/csrd/aboutcsrd.html>
- Cook, T., Hunt, H., & Murphy, R. (2000). Comer's school development program in Chicago: A theory-based evaluation. *American Educational Research Journal*, *37*(2), 535–597.
- Coyne, M. D., Kame'enui, E. J., & Simmons, D. C. (2001). Prevention and intervention in beginning reading: Two complex systems. *Learning Disabilities Research and Practice*, *16*, 62–72.
- Foorman, B., Francis, D., Fletcher, J., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, *90*, 37–55.
- Fullan, M. (1991). *The new meaning of educational change* (2d ed.). New York: Teachers College Press.
- Gersten, R. (1985). Structured immersion for language minority students: Results of a longitudinal evaluation. *Educational Evaluation and Policy Analysis*, *7*(3), 187–196.
- Gersten, R., Baker, S., Haager, D., & Graves, A. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners: An observational study. *Remedial and Special Education*, *26*, 197–206.
- Gersten, R., Baker, S., & Lloyd, J. (2000). Designing high-quality research in special education: Group experimental design. *Journal of Special Education*, *34*(1), 2–18.
- Greenberg, R., & Walberg, H. (1998). The Diogenes effect: Why program evaluations fail. In K. Wong (Ed.), *Advances in educational policy: Perspectives on the social functions of schools* (Vol. 4, pp. 167–178). Greenwich, CT: JAI.
- Hall, G. E., & Hord, S. M. (2001). *Implementing change: Patterns, principles, and potholes*. Boston: Allyn and Bacon.
- Harcourt Brace. (1996). *Stanford test of academic skills: Directions for administering complete/basic battery multiple choice* [of Stanford Achievement Test, 9th ed.]. Publication no. 015875218X. San Antonio, TX: Harcourt Brace Educational Measurement.
- Helfand, D. (2000, January 20). Decades later, frustrated father is phonics guru. *Los Angeles Times*, pp. A1, A16.
- Hoskyn, M., & Swanson, H. L. (2000). Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *School Psychology Review*, *29*(1), 102–119.
- Individuals with Disabilities Education Improvement Act (IDEA) of 2004. (2005). Additional procedures for evaluating children with specific learning disabilities. *Federal Register*, *70*(118), June 21, 2005, pp. 35802–35803.
- Jones, E., Gottfredson, G., & Gottfredson, D. (1997). Success for some: An evaluation of the "Success for All" program. *Evaluation Review*, *21*(6), 643–670.
- Keppel, G., & Zedeck, S. (1989). *Data analysis for research design: Analysis of variance and multiple regression/correlation approaches*. New York: W. H. Freeman.
- Maddahian, E. (2002). *A comparative study of second-grade students' reading, language and spelling gain scores for the LAUSD reading programs*. Program Evaluation and Research Branch (Publication no. 113). Los Angeles: Los Angeles Unified School District [On-line]. Available: http://notebook.lausd.net/portal/page?_pageid=33,103101&_dad=ptl&_schema=PTL_EP
- Mathes, P., Denton, C., Fletcher, J., Anthony, J., Francis, D., & Schatschneider, C. (2005). The

- effects of different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, **40**(2), 148–182.
- McRae, D. J. (2002). *Test score gains for Open Court schools in California: Results from three cohorts of schools* [On-line]. Available: <http://www.sraonline.com/index.php/home/ocrr/1115>
- National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- National Research Council. (1999). *Uncommon measures: Equivalence and linkage among educational tests*. Washington, DC: National Academy Press.
- No Child Left Behind Act of 2001, P.L. 107-110 (2001). [On-line]. Available: <http://www.ed.gov/legislation/ESEA02/> (see Part B, Subparts 1 and 2, Reading First, etc.).
- O'Connor, R. E., Notari-Syverson, A., & Vadasy, P. F. (1996). Ladders to literacy: The effects of teacher-led phonological activities for kindergarten children with and without learning disabilities. *Exceptional Children*, **63**(1), 117–130.
- Rowan, B., Bossert, S. T., & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, **12**(4), 24–31.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
- Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. New York: Knopf.
- Skindrud, K., & Gersten, R. (2005). *An independent evaluation of two prominent reading programs in the Sacramento City Schools: Academic and special education outcomes* (Technical Report 2005A). Copies available: Research and Funded Projects, Weiss Hall D445, California State University Dominguez Hills, 1000 East Victoria Street, Carson, CA 90747.
- Slavin, R., & Madden, N. (2000). Roots and Wings: Effects of whole-school reform on student achievement. *Journal of Education for Students Placed at Risk*, **5**(1 & 2), 109–136.
- Slavin, R., & Madden, N. (2001). *One million children: Success for all*. Thousand Oaks, CA: Corwin.
- Slavin, R., Madden, N., Dolan, L., & Wasik, B. (1996). *Every child, every school: Success for all*. Thousand Oaks, CA: Corwin.
- Slayton, J., Oliver, D., & Burley, K. (2003). *Evaluation of the district reading plan 2002–2003* (Publication no. 170). Program Evaluation and Research Branch, Los Angeles Unified School District [On-line]. Available: http://notebook.lausd.net/portal/page?_pageid=33,103101&_dad=ptl&_schema=PTL_EP
- Snow, C. S., Burns, S. M., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Southwest Educational Demonstration Laboratory. (2003). CRS awards databases, 1998–2002 [On-line]. Available: <http://www.sedl.org/csrd/awards.html>
- SRA/McGraw Hill. (1996). *Open court collections for young scholars* (1996 ed.). [On-line]. Available: <http://www.aft.org/pubs-reports/downloads/teachers/seven.pdf>
- SRA/McGraw Hill. (2000). *Open court reading 2000* (2000 ed.). [On-line]. Available: <http://www.sraonline.com/index.php/home/curriculum/solutions/reading/ocr2000/ocr2000longcopy/233>
- Success for All Foundation. (1999). *Success for All implementation visit record form*. Unpublished form by SFA Foundation, Baltimore, MD.
- Torgesen, J. K., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, **34**, 33–58.
- U.S. Department of Education. (2005). What works clearinghouse [On-line]. Available: Retrieved 3-04-05 at: <http://www.whatworks.ed.gov/>
- Vellutino, F., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers. *Journal of Educational Psychology*, **88**, 601–638.
- Venezky, R. (1998). An alternative perspective on "Success for All." In K. Wong (Ed.), *Advances in educational policy: Perspectives on the social functions of schools* (Vol. 4, pp. 145–165). Greenwich, CT: JAI.
- Walberg, H., & Greenberg, R. (1999). The Diogenes factor. *Phi Delta Kappan*, **81**, 127–128.
- Westat. (2001). *Report on the final evaluation of the city-state partnership: New Baltimore City Board of School Commissioners and the Maryland State Department of Education* (chap. 4, pp. 21–26, and chap. 5, p. 59). Unpublished report by Westat, 1650 Research Blvd., Rockville, MD 20850.

Copyright of Elementary School Journal is the property of University of Chicago Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.