Comparing Four Literacy Reform Models in High-Poverty Schools: Patterns of First-Grade Achievement

Author(s): Terrence Tivnan and Lowry Hemphill

Source: *The Elementary School Journal*, Vol. 105, No. 5 (May 2005), pp. 419-441

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/10.1086/431885

Comparing Four Literacy Reform Models in High-Poverty Schools: Patterns of First-Grade Achievement

Terrence Tivnan
Harvard University

Lowry Hemphill
Wheelock College

The Elementary School Journal
Volume 105, Number 5

0013-5984/2005/10505-0001$05.00

Abstract

Low-performing districts have sought to raise student achievement through adoption of schoolwide models for the reform of literacy instruction, a trend that has intensified under the Reading First Act. This study examined literacy achievement for first graders in a large urban district that offered its schools a choice of literacy reform models. Sixteen high-poverty schools that had made at least "good" efforts in implementing their chosen reform model were the focus of the investigation. Literacy achievement for 590 children was assessed in fall and spring of first grade, including assessments of word reading, phonemic awareness, vocabulary, reading comprehension, and writing. The models adopted in the district, Building Essential Literacy, Developing Literacy First, Literacy Collaborative, and Success for All, produced similarly strong outcomes in first-grade word reading despite philosophical and practical differences in the models' implementation. Skills related to meaning construction, however, were low for children receiving instruction in all four models, with vocabulary and reading comprehension failing to reach grade-level expectations for most children.

Poor literacy achievement among low-income children has been a persistent challenge for U.S. schools. According to recent national assessments (Donahue, Finnegan, Lutkus, Allen, & Campbell, 2001), 60% of children from high-poverty households score below even a "basic" level in reading at the end of the primary grades. Achievement has remained low for students in large urban districts despite more than 3 decades of federal investment in programs to remediate or more recently *prevent* reading failure (Lutkus & Weiner, 2003).

Many school systems have begun to adopt more systematic approaches to teach-

ing beginning reading as one part of a strategy to improve academic achievement. These approaches include district-developed programs, such as the reading initiatives undertaken in Chicago and New York, as well as school-initiated programs (Taylor & Pearson, 2002). More commonly, however, districts adopt "off-the-shelf" programs such as Success for All, Direct Instruction, and Literacy Collaborative (Duffy-Hester, 1999; St. John, Loescher, & Bardzell, 2003; Wong & Meyer, 2001), a trend that has been strengthened by the provisions of the Reading First Act. The appeal of off-the-shelf programs is that they can substitute for forms of expertise that may be lacking locally, through the provision of new literacy materials and teaching strategies, focused professional development for teachers, and trained coaches to monitor implementation of reform strategies. Such programs also promise to offer students a more consistent experience across grade levels and teachers, and consistency in literacy instruction has been linked to better student outcomes (Mosenthal, Lipson, Torncello, Russ, & Mekkelsen, 2004). Research has demonstrated that nationally disseminated reform models such as Success for All and Literacy Collaborative can result in better outcomes in some schools and districts than traditional reading programs, particularly in the early grades (Borman & Hewes, 2002; Madden, Slavin, Karweit, Dolan, & Wasik, 1993; Smith, Ross, & Casey, 1996; St. John et al., 2003; Williams, Scharer, & Pinnell, 2000). It remains unclear, however, whether such reform models can bring a majority of students to grade-level expectations in schools that serve high concentrations of children in poverty. There is also a limited comparative research base that contrasts the effectiveness of different literacy models when implemented in similar settings.

Literacy Outcomes for Children in High-Poverty Schools.

Despite decades of federal and local investment in remedial and preventive programs, many low-income children continue to struggle to acquire early reading skills. On the most recent national assessments of fourth-grade reading, 54% of children who were eligible for free- or reduced-price lunch scored at the lowest performance level, below basic, whereas only 23% of nonpoverty children scored below basic (Grigg, Daane, Jin, & Campbell, 2003). Achievement of African-American children, a large group in most urban districts, improved only modestly on fourth-grade reading assessments throughout the past decade. In 2002, 60% of African-American fourth graders performed below basic level expectations in National Assessment of Educational Progress (NAEP) reading (Grigg et al., 2003).

Children who attend schools with high concentrations of family poverty may be especially vulnerable to underachievement in literacy (Donahue et al., 2001; Grigg et al., 2003; Sutton & Soderstrom, 1999; U.S. Department of Education, 2001). Classrooms in such schools may be bleaker, containing fewer books, and literacy activities may be more rotelike and unchallenging (Barone, 2002; Duke, 2000; Puma, Jones, Rock, & Fernández, 1993). In addition, children in high-poverty schools are less likely to be taught by teachers who are fully certified and who themselves perform well on external assessments (Haycock, 2000).

In contrast with middle-class students, children from low-income families more commonly begin school with a weaker base of language and early literacy skills (Hart & Risley, 1995). Many Hispanic and African-American children, significant groups within urban schools serving children in poverty, may be still acquiring English or may speak other dialects of English, language factors that can delay their early development of English literacy (Carlisle, Beeman, Davis, & Spharo, 1999; Washington, 2001).

Despite multiple risk factors for students and obstacles to meaningful reform of instruction, several large urban districts,

among them San Diego and Houston, have reported gains on state or local literacy assessments from 1995 to 2001. A report by the Council of Great City Schools (Snipes, Doolittle, & Herlihy, 2002) identified key initiatives undertaken by urban districts that have demonstrated improvement in literacy. Along with a strong focus on academic achievement and the adoption of internal accountability measures, the successful urban districts have implemented new reading curricula in the primary grades. Although the districts examined in the council's report raised achievement overall and reduced the gap between racial groups, literacy achievement in some of the highest-poverty schools remained low, and fell as children moved from the early grades through elementary and middle school.

Four Early Literacy Models

Although a combination of accountability measures and investment in reform of literacy instruction has yielded positive results for some urban districts, there remain important questions about the success of such efforts for children at greatest risk of failure. One key concern is whether widely disseminated literacy models offer equivalent benefits for low-income children's literacy progress, particularly for children's development of more challenging literacy skills. A second concern is whether reform models can bring most children to grade level in the highest-poverty schools.

Unlike the districts profiled in the report by the Council of Great City Schools, who selected a single literacy model, Boston, the focus of this study, offered its elementary schools a menu of five nationally disseminated literacy reform models. Schools investigated characteristics of the district-endorsed models and chose a model based on staff discussion and faculty vote. The choice was later limited to four: Building Essential Literacy (BEL, formerly Balanced Early Literacy; Crévola & Hill, 1998; Hill & Crévola, 1999; Hill & Jaggar, 2003), Developing Literacy First (DLF, formely ELIC; Johnson, 2001), Literacy Collaborative (LC; Williams et al., 2000), and Success for All (SFA; Datnow & Castellano, 2000a) (see Appendix for a summary of each model's key features).

SFA uses an approach of scripted instruction with phonetically regular readers, and the three other models, BEL, DLF, and LC, use guided reading with leveled books but differ from each other in other features. BEL and DLF, for example, use outside specialists to provide training for teachers in literacy pedagogy, whereas LC uses a "train the trainers" approach with a designated teacher training her peers (Williams et al., 2000). BEL and LC literacy blocks allocate time for writing as well as reading; DLF and SFA's early-grades "Roots" program makes no requirements for writing time during the literacy block. BEL and LC mandate the use of small leveled readers in first grade, but DLF is more flexible in its recommendations for beginning reading materials. SFA is least flexible in allowing teacher or student choice of reading materials, providing its own Roots readers. BEL and LC both advocate the use of center activities in first grade in which children work to build words, partner-read, or engage in other small-group reading or writing activities while the teacher leads guided reading groups (Crévola & Hill, 1998; Ford & Opitz, 2002; Williams et al., 2000). DLF, in contrast, makes no special requirement about the organization of the classroom while the teacher is conducting guided reading.

SFA, the fourth model implemented, contrasts most markedly with the other three reform approaches. Teachers in SFA classrooms use a highly structured teacher's manual to carry out instruction, a feature that is not present in the other three models (Datnow & Castellano, 2000a). SFA first graders progress through carefully sequenced, phonetically regular readers that are developed by the SFA organization rather than learning to read with small leveled readers or an eclectic mix of beginning reading materials. Finally, SFA uses a direct

instruction form of pedagogy for teaching word-recognition skills, in which teachers, following a script, explicitly present phonics generalizations and drill children on word patterns (Datnow & Castellano, 2000a). For example, in a first-grade SFA Roots class in Boston, the teacher holds up a letter card and asks, "What letter is this?" Children and teacher recite "Full circle round, up, down, hook left. The sound for 'g' is g-g-g." Children then chorally read words with 'g': gas, mag, tags.

The other three models investigated use versions of "embedded" phonics instruction (Dahl, Scharer, Lawson, & Grogan, 1999; Freppon & McIntyre, 1999) in which phonics generalizations are largely taught informally, as children encounter words with particular patterns in their reading and writing (Foorman, Francis, Fletcher, Shatschneider, & Mehta, 1998). In a first-grade BEL reading group, for example, a child misreads the word "eat" as "at" in the story "Three Billy Goats Gruff." The teacher tells the child that the letters 'ea' make the sound 'ee' and the child continues reading the story.

Research on Literacy Reform Models

The four reform models examined include three, Building Essential Literacy, Literacy Collaborative, and Success for All, whose effects have been examined across multiple school settings. Literacy Collaborative's research on student progress has compared reading scores of second graders in LC schools over time, from the first year of implementation of the reform and for subsequent years. Reading comprehension scores for second graders in the LC schools studied were very low in the first year of implementation, averaging at only the twenty-ninth percentile, an expected finding because schools are typically motivated to adopt a demanding new reading program when student achievement is poor. According to LC's evaluations, average second-grade reading scores at 52 schools that have implemented LC over multiple years rose to the thirty-ninth percentile by the end of the fourth year of implementation, and to the forty-sixth percentile by the end of the fifth year (Scharer, Williams, & Pinnell, 2001; Williams et al., 2000). Consistent with research on other literacy models, second-grade reading achievement was higher for children who remained in the same school from kindergarten through second grade. In addition, achievement improved more markedly in schools with a literacy coordinator who served fewer teachers, and in schools that were able to make more generous provision for tutoring children at risk of failure (Scharer et al., 2001).

Although these results suggested positive accomplishments for the LC model in raising student achievement over time, the research did not report separately the literacy progress of children in very low-income schools where obstacles to improvement may have been more substantial. In addition, although the reported results indicated achievement gains when LC was implemented in a school for multiple years, results aggregated across schools with different characteristics did not demonstrate that a majority of children fully reached grade-level expectations. Similar patterns were reported in a multisite examination of literacy achievement in BEL schools in New York City, Boston, and suburban Chicago. Student achievement rose over 2 years of implementation of BEL instruction but did not reach grade-level benchmarks (Hill & Jaggar, 2003).

Research carried out in many urban districts has suggested that first-grade children in schools that have implemented Success for All out-performed first-grade children in traditional basal reader programs on tests of word identification, word attack, and oral reading (Ross, Smith, & Casey, 1997; Smith et al., 1996). SFA children appeared to continue to make gains in the upper-elementary years; for example, third through fifth graders in SFA classes made greater gains on the Texas state reading assessments than comparison children, al-

though these gains may have resulted in part from lower initial achievement in SFA schools (Hurley, Chamberlain, Slavin, & Madden, 2001). A large group of Baltimore students who received SFA instruction in elementary school experienced lower rates of referral to special education and outperformed comparison students on standard reading achievement tests in eighth grade, although neither SFA students nor comparison students showed average achievement that was at grade-level expectations (Borman & Hewes, 2002). An analysis of progress over time for another cohort of students in Baltimore schools showed generally superior performance for SFA students in comparison with control students but declining rates of growth for both groups across the elementary school years (Venezky, 1998).

In several studies the results for SFA appeared to vary, like those for LC, by important school characteristics. In a large-scale examination of SFA effects in Houston, quality of implementation of the reforms required by SFA was lower, on average, in schools with fewer white students, greater student mobility, and poorer attendance (Nunnery et al., 1997). Like the results of the evaluation of LC outcomes, results for SFA appeared to be strongest in schools that had made the largest investment in tutoring of children at risk of failure and in schools that had released a teacher or administrator full-time who worked over multiple years as the literacy coordinator (Nunnery et al., 1997; Smith et al., 1996).

In one of the few studies that have contrasted the effects of different literacy reform models, researchers found no significant effects of participation in Success for All or Literacy Collaborative on Indiana schools' passing rates on the state's third-grade reading test. Adoption of both SFA and LC, however, was associated with lower rates of referral to special education, and adoption of SFA was associated with lower rates of retention in the primary grades (St. John, Manset, Chung, Simmons, & Musoba, 2000). In a reanalysis of only urban schools participating in the Indiana study, adoption of LC was associated with lower rates of retention in grade, but adoption of SFA was not (St. John et al., 2003).

Building Essential Literacy, Success for All, and Literacy Collaborative contrast with each other in multiple dimensions—for example, SFA's emphasis on explicit phonics teaching and the more embedded approach to word study in BEL and LC (Crévola & Hill, 1998: Dahl et al., 1999). Nonetheless, adoption of a literacy reform model with either of these pedagogical approaches may result in a common set of positive changes for teachers and students: increased allocation of time for reading instruction, new beginning reading materials, regrouping of children into smaller groups with similar needs, and training for teachers in new teaching strategies.

Therefore, rather than simply compare children's progress in standard versus reformed instruction, the most common research design in studies of literacy reform (e.g., Borman & Hewes, 2003; Hill & Jaggar, 2003), it is important to contrast the progress children make when similar reform benefits (professional development, new materials, regrouping of students) are available but when the classroom pedagogies provided by reform models differ. It is also critical to focus on reform outcomes in schools with the greatest risk factors: histories of poor academic achievement, very high concentrations of children in poverty, and student populations that are overwhelmingly African American or Hispanic. Finally, it is crucial to look at children's mastery of grade-level skills and strategies, in addition to examining progress over time.

Characteristics of Literacy Development

An assessment of children's literacy progress in reformed instruction should be informed by an understanding of several key characteristics of early literacy development. One important characteristic is literacy's complexity. Literacy growth cannot be

reduced to progress within a single skill but rather is evident across an array of skills whose development proceeds somewhat independently and with different social and environmental supports (Mason, Stewart, Peterman, & Dunning, 1992; Snow, 1983). Essential areas of literacy for children in the primary grades include word reading, word attack, phonemic awareness, writing, and reading comprehension (Armbruster, Lehr, & Osborn, 2001). Receptive vocabulary is also an important component of early literacy, especially affecting children's development of reading comprehension (Dickinson & Tabors, 2001; Mason et al., 1992). Because reform models differ in focus and pedagogy and thus may not provide equally strong support for achievement across all domains of literacy, it is critical to assess student progress in multiple aspects of literacy.

A second important characteristic of early literacy development is that children begin formal instruction at different points. Some children enter high-poverty elementary schools having attended preschools that provide rich opportunities to engage with books, whereas classmates may begin first grade with no preschool or even kindergarten experience. Kindergarten literacy programs can offer a strong foundation for beginning reading instruction or, alternatively, may not expose students to rich vocabulary or to understandings about letter-sound correspondences (Barone, 2002). Beginning literacy skills may also vary for first graders as a result of different forms and levels of home support for literacy. Some low-income students start first grade with considerable home exposure to books and with rich oral language experiences; other students lack this base (Britto & Brooks-Gunn, 2001). Because multiple factors can affect variation in children's early literacy skills, even within a low-income sample, it is essential to account for different starting points in assessing children's literacy growth.

A third key aspect of early literacy development is the powerful influence of child characteristics such as family income level, home language, gender, and race/ethnicity on student progress. These factors are associated with children's initial skill levels as they begin first grade and with their rates of progress as school literacy instruction begins (Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Mason et al., 1992). In examining student achievement over time, therefore, it is important to attempt to control at the child level for the possible effects of family income, home language, gender, and ethnicity.

Our examination of first-grade achievement within four literacy reform models was shaped by the following questions:

1. Taking into account child characteristics and varying starting points, what are the effects of participation in a particular early literacy model on low-income students' acquisition of literacy skills?
2. Are there differences in the models' effectiveness for different components of literacy (e.g., decoding skills, vocabulary, reading comprehension, writing)?
3. Do children reach challenging grade-level benchmarks at the end of first grade?

Method

Characteristics of the Study Schools

Schools that undertake significant curriculum reforms typically engage in a multiyear process, including initial training of teachers, trying out new teaching strategies, and gradual incorporation of the reforms into daily work with children (Datnow & Castellano, 2000b). Because we sought to examine student achievement in schools that had reached a "mature" phase of the literacy reform process, schools selected for the study were in their third to fifth year of implementation of the reform model when the study began. Because there is compelling evidence that uncommitted leadership or poor uptake of reforms by teachers can blunt their effects (Cooper, 1998; Payne, 1998; Useem, Christman, Gold, & Simon,

1997), we included only schools that in the judgment of district leaders were making at least "good" efforts to carry out the changes demanded by each reform model. Finally, because there is evidence that effects of literacy reforms may vary in schools with different demographic characteristics (Nunnery et al., 1997), we chose schools that all met definitions of high-poverty status, with 75% or more of their students receiving free or reduced-price lunch, and whose student bodies were predominantly African American and Hispanic.

It is important to note that schools in the district chose to adopt their reform model and accumulated several years of experience with the model rather than having an intervention randomly assigned to them. Thus, the design of the study was of necessity quasi-experimental rather than randomized. As Pressley and other observers have noted, instructional interventions that require considerable teacher commitment and expertise are particularly difficult to study in a randomized experiment (Pressley, 2003). The design of this study also included no nonintervention control group because all schools in the district were implementing one of the four literacy models. Rather, the study contrasted child achievement when different types of literacy reforms were implemented (Table 1).

The 16 study schools were located in seven neighborhoods of Boston, including largely Latino East Boston and the African-American neighborhoods of Dorchester, Roxbury, and Mattapan. Because of Boston's "controlled choice" desegregation plan, all study children did not necessarily live in the neighborhood surrounding their elementary school but were bused within larger districts that cut across broad sections of the city. The proportion of white children in the study schools ranged from 3% in the most racially isolated school to 32% in the most racially mixed school.

Children in all the study schools were overwhelmingly poor: the proportion of children eligible for free or reduced-price lunch ranged from 78% to 90%. Both larger and smaller schools were represented in the sample, reflecting the range of elementary school sizes in Boston. The smallest school enrolled 132 children, grades K–5; the largest enrolled 800 students. Three of the smaller schools included only one first-grade classroom, whereas the three largest schools included five first-grade classrooms.

Virtually all first-grade teachers in the study schools were fully certified, and most had completed master's degrees or additional coursework beyond the bachelor's. The range of teaching experience among the teachers was considerable; 10 had 5 years experience or less (and three were first-year teachers), and 13 had more than 30 years experience. The mean years of teaching experience for the first-grade teachers in English-medium classrooms was 18.

Table 2 displays school characteristics for the four schools implementing each of the four literacy models. The 16 schools were similar to each other in their overwhelmingly low-income and minority student bodies but showed some variation in relative proportions of African-American and Hispanic students. Proportions of Hispanic students were higher, on average, in

Table 1. Sample Size Information

	Building Essential Literacy (BEL)	Developing Literacy First (DLF)	Literacy Collaboration (LC)	Success for All (SFA)	Entire Sample
Schools	4	4	4	4	16
Homeroom teachers	11	8	10	10	34
Children	160	157	145	128	590

TABLE 2. Student Demographic and Prior-Year Achievement Data (Percentages) for Schools Using Each Literacy Model

Model/School	African American	Hispanic	Asian	White	Free-Lunch Eligible	At Reading Levels 3 and 4
Building Essential Literacy:						
B1	22	37	9	32	88	71
B2	51	9	13	26	80	33
B3	56	25	16	3	89	33
B4	16	64	7	14	85	21
Mean	36	34	11	19	86	40
Developing Literacy First:						
D1	48	18	5	29	83	63
D2	90	3	1	5	82	36
D3	49	15	11	26	82	39
D4	76	18	4	3	88	25
Mean	66	13	5	16	84	41
Literacy Collaborative:						
L1	52	26	3	18	78	35
L2	74	18	1	7	85	50
L3	81	12	<1	7	82	16
L4	35	47	3	16	85	33
Mean	60	26	2	12	83	34
Success for All:						
S1	24	70	1	6	87	29
S2	86	13	1	1	90	46
S3	6	61	3	30	84	30
S4	53	21	2	23	79	64
Mean	42	41	2	15	85	42

NOTE.—Table entries are the percentages of each school population falling into the categories. These entries thus represent information for entire schools and not just the first-grade students in our study. Achievement levels reported here are the percentages of third graders in the year preceding the study who scored at levels 3 or 4 ("solid" or "superior" performance) on the Stanford 9 Reading Test.

the schools implementing BEL and SFA, in part because of the greater availability of Spanish-medium materials for these models. Past reading achievement of third graders showed some variability among schools using the same model, but no significant differences were evident in the average past achievement for groups of schools implementing each model.

Participants

In September, all of the first graders in the 16 study schools were invited to participate in the evaluation. Close to 95% received parental permission to engage in the project's activities. In May, when end-of-first-grade achievement data were collected, complete beginning and end-of-grade assessments were available for 590 children enrolled in first-grade classes using English as the medium for literacy instruction and for 102 children receiving literacy instruction in Spanish. Because of noncomparable norming groups for the English and Spanish versions of the standardized assessments administered, the analyses below present only information for children who were instructed in English.

With parent consent we collected demographic data from district records concerning study children's ethnicity, eligibility for free or reduced-price lunch, and home language (Table 3). Among the 590 first graders receiving literacy instruction in English, 52% were African American, 27% were Hispanic, 15% were white, and 6% were Asian American. Seventy-five percent received free lunch, 12% qualified for reduced-price lunch, and 13% paid for lunch. Eighty-six percent of the children receiving instruction in English identified English as their home language, 7% spoke Spanish at home, and

TABLE 3. Demographic Information (Percentages) on First-Grade Sample

	Building Essential Literacy (BEL)	Developing Literacy First (DLF)	Literacy Collaborative (LC)	Success for All (SFA)	Entire Sample
Gender:					
Male	46	56	50	51	51
Female	54	44	50	49	49
Ethnicity:					
African American	42	64	60	42	52
Hispanic	23	17	24	45	27
White	21	14	13	10	15
Asian	13	5	1	3	6
Lunch status:					
Free	82	64	74	83	75
Reduced-price	9	15	12	11	12
Paid	9	20	15	6	13
Home language:					
English	79	92	94	79	86
Spanish	11	2	4	12	7
Other	10	6	1	9	7

7% spoke another home language, such as Haitian Creole, Brazilian Portuguese, or Vietnamese.

Literacy Assessments

In October, children participated in individual assessments of word reading, phonemic awareness, and oral language, skills that are foundational for beginning literacy. Two subtests of the Woodcock-Johnson Diagnostic Reading Battery (WDRB; Woodcock, 1997) were administered: the word and letter identification test, which presents children with a list of upper- and lowercase letters and sight words to read; and the word attack test, which presents children with a list of decodable nonsense words. These subtests of the Woodcock-Johnson have been used in other comparisons of first-grade reading achievement for students receiving literacy instruction within different literacy programs (e.g., Smith et al., 1996). Children also took the Yopp-Singer phonemic awareness test (Yopp, 1995), which asks children to segment one-syllable words presented orally into their component phonemes, and the Peabody Picture Vocabulary Test-III (PPVT-III; Dunn & Dunn, 1997), a widely used test of receptive vocabulary.

All of the individually administered literacy assessments were repeated in May of first grade along with a group-administered reading comprehension test, the Gates-MacGinitie Primary 1 (GMRT-4; MacGinitie & MacGinitie, 2000), and a locally developed writing task. The writing task asked children to generate a written description of a photograph depicting children playing a clapping game. Descriptions composed by the children were scored for the use of sentence conventions, phonetic and conventional spelling, elaborated content, and genre features. Scores for each of these dimensions of writing were summed into a composite score. Reading comprehension and writing assessments were not attempted in the fall of first grade because of children's limited independent reading and writing skills.

Characteristics of Literacy Instruction

For 2 hours in the fall and 2 hours in the spring, literacy instruction was observed in each study classroom, using an observation system modeled on the CIERA Classroom Observation Scheme (Tayor, Pearson, Peterson, & Rodriguez, 2003). Research assistants with experience as primary-grade teachers or as reading specialists recorded narrative

field notes and coded at 5-minute intervals for the predominant literacy focus (e.g., phonics, discussion of word meaning, choral reading), text focus (e.g., word list, basal reader, trade book), and instructional grouping (e.g., whole class or partners). Observers scanned the classroom at the end of each 5-minute observation block, counting children who were engaged or not engaged in the ongoing lesson activity, such as partner reading or listening to the teacher modeling writing conventions. Observers used a four-point scale for engagement, ranging from 1 (few or no students on task) to 4 (everyone on task). High interrater agreement was obtained when two coders jointly observed the same classroom—in the range of .80 or higher for literacy focus, text focus, instructional grouping, and student engagement levels.

Results

Literacy Achievement in First Grade

Table 4 presents fall of first grade literacy achievement scores for children receiving instruction within each of the four literacy models. The pattern for overall beginning of first grade achievement represented average-to-good early word reading skills, as evidenced by mean scores that were at or above grade level on the two subtests of the WDRB, letter-word identification and decoding, but comparatively weak scores on the Yopp-Singer test of phonemic awareness. Children's mean scores on the Yopp-Singer in the fall of first grade reflected the ability to segment two-phoneme words such as /d/ /o/ but not to segment words with three or more phonemes. Vocabulary scores were very low at the beginning of first grade, averaging only at the nineteenth percentile.

Some contrasts in average performance for children in the different literacy models were evident at the beginning of first grade, reflecting the combined effects of differences in the effectiveness of schools' kindergarten literacy programs, different rates of participation in kindergarten at study schools, and differences in individual chil-

TABLE 4. Test Scores in Fall of First Grade, by Literacy Model

	Literacy Model				ANOVA Results		
	Building Essential Literacy (BEL)	Developing Literacy First (DLF)	Literacy Collaborative (LC)	Success for All (SFA)	*F*	*p*	Contrasts
Letter-word identification:							
Raw score mean	15.9	17.4	17.9	14.6	6.08	.009	D, L > B, S
Standard deviation	4.9	6.0	5.6	5.3			
Standard score mean	99	103	106	93			
Percentile rank	46	58	66	31			
Word attack:							
Raw score mean	1.1	1.9	1.8	1.5	2.27	.133	
Standard deviation	5.9	5.9	6.1	6.9			
Standard score mean	99	104	104	102			
Percentile rank	47	61	61	56			
Phonemic awareness:							
Raw score mean	6.6	7.9	8.0	7.1	.63	.607	
Standard deviation	6.6	6.5	6.6	7.3			
Vocabulary:							
Raw score mean	73.8	76.3	77.0	67.1	3.53	.0485	B, D, L > S
Standard deviation	17.7	17.2	17.8	18.4			
Standard score mean	86.3	88.5	88.9	81.9			
Percentile rank	18	23	23	12			

NOTE.—Table entries are raw score means and standard deviations and, where available, the corresponding standard scores and percentile ranks based on test publisher information for nationwide norms. National norms are not available for the test of phonemic awareness.

dren's home language and literacy experiences. Analysis of variance demonstrated that children in BEL and SFA schools scored significantly lower, on average, in fall of first grade word reading than children in DLF and LC schools. Children in SFA schools also scored lower in fall of first grade vocabulary than children in schools implementing the other three models. In our analyses of possible model differences in rates of student progress during first grade, we planned to adjust for the relatively small differences in children's initial skill levels. We used the fall letter-word identification scores and the fall vocabulary scores to create a composite variable representing children's skill levels in the fall of first grade, and this measure of initial skill level was used in our subsequent analyses of student growth.

Table 5 presents the literacy achievement scores for the study children in May of first grade, after children had participated in the first-grade literacy program offered by their school. In most domains of literacy, gains were evident between the beginning and end of first grade. The entire sample of children moved from the fiftieth percentile in word and letter identification at the beginning of first grade to the sixty-first percentile at the end of the year. Average word-attack skills also showed an increase in percentile rank across the first-grade year, rising to the fifty-eighth percentile.

Phonemic segmentation was another skill area where study children showed substantial progress across the first-grade year. In May, children were able to segment about 10 more words on the Yopp-Singer than they could segment in October, typically successfully segmenting three-phoneme words such as /d/ /o/ /g/ by the end of first grade. Most children were not, however, able to segment four-phoneme words at the end of first grade and, as a consequence, mean scores did not approach ceiling levels on the Yopp-Singer. In vocabulary, the area in which children showed the weakest achievement overall at the beginning of first grade, children moved from the nineteenth to the twenty-third percentile. Although gains in vocabulary were evident for children in all four of the literacy models, average vocabulary scores remained low at the end of first grade, still averaging about a year below age expectations. Reading comprehension, assessed for the first time in the spring of first grade with the Gates-MacGinitie, was weaker overall than children's word-level reading skills. The mean achievement for the whole sample was at the forty-second percentile, several months below grade-level expectations. Average reading comprehension showed variability across literacy models: the mean percentile scores at the end of first grade ranged from the thirty-fifth percentile for children receiving literacy instruction in SFA classrooms to the forty-eighth percentile for children in DLF classes.

The average writing score for the study children in the spring of first grade reflected an ability to write a two- to three-sentence description, reporting one or two major elements depicted in the picture prompt. A typical writing sample produced at the end of first grade included consistent use of upper- and lowercase letters and sentence punctuation, multiple conventionally spelled sight words such as "this" and "they," and conventionally spelled one-syllable words such as "hat" and "drum." Writing achievement at the end of first grade showed variability across literacy models, with children in LC classes receiving the highest average writing scores and children in BEL classes the lowest.

Comparisons across Models

To compare the relative performance of the models in improving student performance during first grade, we used multilevel regression techniques to control for important child characteristics. As Table 3 illustrates, there were somewhat different proportions of female students, students from different ethnic groups, children eli-

TABLE 5. Mean Test Scores in Spring of First Grade, by Literacy Model

	Literacy Model				
	Building Essential Literacy (BEL)	Developing Literacy First (DLF)	Literacy Collaborative (LC)	Success for All (SFA)	Total Sample
Letter-word identification:					
Raw score mean	25.1	27.4	27.4	25.5	26.3
Standard deviation	5.9	5.9	6.1	6.9	6.3
Standard score mean	100	107	107	101	104
Percentile rank	50	68	68	52	61
Word attack:					
Raw score mean	5.8	7.1	6.7	7.3	6.7
Standard deviation	4.8	6.0	6.0	6.1	5.8
Standard score mean	100	104	103	105	103
Percentile rank	50	60	58	62	58
Phonemic awareness:					
Raw score mean	15.1	14.7	15.4	15.6	15.2
Standard deviation	5.5	5.2	5.0	4.8	5.2
Vocabulary:					
Raw score mean	87.4	88.3	88.2	79.6	86.1
Standard deviation	16.3	16.7	17.4	15.0	16.7
Standard score mean	89.9	90.5	90.2	84.3	88.9
Percentile rank	23	26	25	14	23
Reading comprehension:					
Raw score mean	22.9	24.9	23.5	20.2	23.0
Standard deviation	8.1	8.7	9.4	8.7	8.9
Standard score mean	382	389	386	371	382
Percentile rank	42	48	45	35	42
Writing:					
Raw score	11.9	12.8	14.1	12.0	12.7
Standard deviation	4.0	4.0	4.0	4.2	4.2

NOTE.—Table entries are raw score means and standard deviations and, where available, the corresponding standard scores and percentile ranks based on test publisher information for nationwide norms. National norms are not available for the test of phonemic awareness and the writing sample. The standard scores for reading comprehension are Gates-MacGinitie extended scale scores.

gible for free lunch, and English-language learners within the four model groups. Additionally, as Table 4 indicates, there were some model-related differences in children's initial skill levels. In comparing the effects of the models, it was important to adjust for these differences in beginning literacy skills. We were also interested in investigating whether there might be differential effects of the models for children who differed in their initial levels of skill. We therefore used a two-level regression model to predict student literacy achievement, incorporating child information, such as gender and initial literacy skill, and school information about the literacy model implemented. For those outcomes (word reading, word attack, phonemic segmentation, and vocabulary) where it was possible to assess achievement at both the beginning and end of first grade, we looked at the effects of child and school characteristics on literacy *growth*, the difference between spring and fall test scores for individuals. For those outcomes assessed only at the end of first grade because of their greater demands (reading comprehension and writing), child and school characteristics were used to predict literacy *status* for individuals at the end of first grade.

As Table 6 illustrates, child-level characteristics showed few significant effects for specific domains of literacy achievement, such as word reading and vocabulary. Free-lunch eligible students, for example, who were by far the largest group in the sample,

TABLE 6. Linear Mixed Model Regression Results for Student Literacy Growth from Fall to Spring of First Grade

Effect	Letter-Word Identification	Word Attack	Phonemic Awareness	Vocabulary	Comprehension	Writing
Intercept	11.77	5.80	7.74	12.11	20.29	11.39
Fixed-effect estimates for student-level background characteristics:						
Free lunch	−1.82**	.36	.66	−.97	.54	−.24
Reduced-price lunch	−1.40˜	−.24	.79	1.60	.93	−.44
Non-English	−.86	.48	−.20	.07	.21	.97˜
Black	−.39	−.93˜	−.79	−.09	−.81	.59
Hispanic	.77	−.41	−.23	−.49	1.54	.69
Female	1.07**	.70˜	−.51	−.57	3.01**	1.98**
Initial skill level[a]	−.48*	1.13***	−1.77***	−3.72***	5.16**	1.54**
Adjusted means for literacy models:						
Building Essential Literacy	9.52	4.90	8.30	14.15	23.88	12.42
Developing Literacy First	10.47	5.46	7.10	12.93	25.01	13.08
Literacy Collaborative	9.72	4.93	7.86	12.48	22.70	13.97
Success for All	10.95	6.23	7.60	12.02	22.39	12.99
Model contrasts	N.S.	SFA > BEL*	N.S.	N.S.	DLF > LC˜ DLF > SFA˜	LC > BEL*
Teacher-level random effects:						
Variance estimate	.63	.33	1.56	1.56	2.04	1.48
Percentage of total	3.5	2.2	4.3	4.3	4.4	11.6
School-level random effects:						
Variance estimate	.39	.08	1.13	1.13	2.07	.69
Percentage of total	2.2	.5	3.1	3.1	4.5	5.4

NOTE.—Analyses based on raw scores from fall and spring of first grade. For letter-word identification, word attack, phonemic awareness, and vocabulary, analyses were carried out on changes in raw scores from fall to spring. For comprehension and writing, students were tested in the spring only, and analyses were carried out on the end-of-year raw scores.

[a]Initial skill level is a composite variable with mean = 0 and SD = 1, representing children's skills at the beginning of first grade. It is a composite based on fall letter-word identification and fall vocabulary scores. Tests of the two-way interaction of literacy model by initial skill level were not significant for any of the analyses.

˜$p < .10$.
*$p < .05$.
**$p < .01$.
***$p < .001$.

showed significantly less growth in letter-word identification than paid- and reduced-price-lunch students, but their performances in other literacy domains were not different from other students'. English-language learners (children whose home language was not English) did not show lower end-of-grade achievement or less growth than children who spoke English as a first language. African-American children in the study showed only slightly less growth in word attack during first grade than white and Asian-American students and showed similar growth patterns in other domains. Hispanic children did not show a significantly different achievement pattern than white and Asian-American students. Girls showed a statistically significant advantage on reading comprehension and writing and showed slightly greater growth than boys in letter-word identification and word-attack skills.

Children's initial skill levels, a composite of fall letter-word identification and fall vocabulary scores, showed significant relations with all of the outcomes, but the direction of the effect varied across literacy domains. For letter-word identification, phonemic awareness, and vocabulary, children with lower initial skill levels showed somewhat greater growth during the school year. In these areas, children with lower initial skill levels were able to narrow the gap between their performance and the perfor-

mance of other children. In word-attack skills, in contrast, children with stronger initial skills showed greater growth during first grade. In reading comprehension and writing, which were assessed only in the spring of grade 1, children with higher initial skill levels in word reading and vocabulary showed higher performance. This was most notable for reading comprehension, where children who differed by one standard deviation in initial skill level differed by over five points on end-of-year comprehension scores.

After controlling for child characteristics, there were few statistically significant contrasts in the literacy outcomes for students receiving instruction within the four reform models. Gains in letter-and-word reading varied across models by only one or two test items, as did gains in the Yopp-Singer test of phonemic segmentation and in vocabulary. Children in SFA classrooms showed significantly greater growth in word attack than children in BEL classrooms ($p = .05$), but the model means differed by just over one point. Children in DLF classrooms showed somewhat higher scores in reading comprehension than children in the other models. The contrast between DLF and LC ($p = .10$) and the contrast between DLF and SFA ($p = .06$) approached statistical significance.

Scores across the models were also generally similar on the writing composite, with children in LC classrooms showing slightly higher average scores, a difference that also appeared in the writing subscales that formed the composite writing score: sentence conventions, spelling, content, and genre features. Adjusting for beginning literacy skills, mean writing scores were significantly higher for LC children compared with children in BEL classrooms ($p = .04$).

There were no significant interactions of literacy model by initial skill level. That is, although children with higher initial skills usually differed in their rates of growth or level of performance when compared with children with lower initial skills, there was no evidence that the models had differential effects for students with different initial skills. No model stood out as more successful in helping students with weak initial skills, for example. In word-attack skills, reading comprehension, and writing, children with higher levels of initial skills tended to show better end-of-grade performance, and this pattern was evident across all four models.

Achievement in Different Domains of Literacy for Each Model

None of the literacy models showed a consistent pattern of superior achievement across all the domains of literacy that were assessed: word reading, word attack, phonemic awareness, vocabulary, reading comprehension, and writing. In fact, all of the models were comparatively effective in promoting student achievement in one or two aspects of early literacy while producing less strong results in other areas. SFA students, for example, made slightly larger average gains in word reading and word attack than students receiving instruction in the other three models, although SFA students showed lower achievement gains in vocabulary and reading comprehension. BEL students showed the weakest gains of all four models in word reading and word attack while showing the strongest gains in phonemic segmentation and vocabulary. Although these patterns were not dramatic, they point to possible trade-offs in early literacy instruction that may result from different time allocations to various first-grade literacy activities, use of different types of beginning reading materials, and from the models' distinctive pedagogical emphases.

Pedagogical Focus

Table 7 shows mean time allocation for the most common instructional focuses observed in the groups of classrooms implementing each model, averaged across the fall and spring observation sessions. Analysis of variance was used to test for possible

TABLE 7. Average Number of Minutes Spent on Literacy Activities during the Daily Literacy Block

Activity	Means				ANOVA Results		
	Building Essential Literacy (BEL)	Developing Literacy First (DLF)	Literacy Collaborative (LC)	Success for All (SFA)	F	p	Contrasts*
Independent reading	5.26	2.46	4.46	1.33	3.62	.0188	B, L > S
Teacher-supported reading	6.02	2.46	5.37	2.33	2.45	.0740	
Teacher read aloud	4.26	2.46	4.86	7.06	1.75	.1684	
Phonics activities	6.78	7.38	8.59	12.44	1.87	.1466	
Writing words, letters, sentences	13.23	11.27	11.58	4.01	4.93	.0043	B, D, L > S
Writing stories or reports	7.75	3.35	4.88	.88	3.63	.0185	B > S
Teacher modeling of writing	5.80	2.19	1.02	.21	14.84	.0001	B > D, L, S D > S
Worksheet activities	.78	.99	1.57	4.34	1.27	.2951	
Comprehension discussion	5.46	9.53	7.87	12.60	2.98	.0396	S > B
Vocabulary work	2.78	1.11	1.64	6.38	2.68	.0560	
Listening to directions and transitions	9.79	14.08	13.46	7.06	2.98	.0396	D, L > S

*$p < .05$.

model differences in average time allocation to particular literacy activities.

SFA classrooms showed several areas of contrasting time allocation compared with classrooms implementing the other three literacy models. In the lessons observed, SFA classrooms spent significantly less time on independent reading and all forms of writing than the classrooms using the other models. In contrast, SFA classes spent more time on phonics, worksheet activities, and vocabulary and comprehension discussion (much of it occurring during scripted teacher read-alouds). SFA classes, on average, showed significantly less student time spent listening to directions compared with DLF and LC classes, a likely consequence of the highly routine and scripted nature of SFA instruction.

More time was allocated to teacher-supported reading, including work with small guided-reading groups and one-on-one work with individual children reading aloud, in BEL and LC classrooms where guided reading was emphasized as the central literacy activity. BEL classrooms, on average, also spent the most time on all forms of writing: teacher modeling, child composing of stories or reports, and child composing of sentences and words. BEL classrooms, however, allocated the least time to comprehension discussion of the four models examined.

Table 8 shows mean amount of time spent using common types of written texts (e.g., leveled readers, trade books, worksheets) in the groups of classrooms implementing each model, averaged across the fall and spring observation sessions. Analysis of variance was used to test for possible model differences in use of particular types of texts during literacy blocks.

Consistent with the design of SFA's Roots program, there was significantly greater use of phonetically regular readers in the SFA classrooms, compared with classrooms implementing the other three models. DLF classrooms made significantly greater use of leveled readers than BEL classrooms, and LC classrooms made significantly more use of this type of literacy material than SFA classrooms. Despite philosophical differences in the models, average use of trade books, largely in teacher read-alouds, was similar across models. "Big books," oversized books with enlarged text for group reading, were used much more frequently in BEL classrooms than in classrooms implementing the other literacy models. Finally, use of worksheets or work-

TABLE 8. Average Number of Minutes Using Different Literacy Materials during the Daily Literacy Block

Materials	Means				ANOVA Results		
	Building Essential Literacy (BEL)	Developing Literacy First (DLF)	Literacy Collaborative (LC)	Success for All (SFA)	*F*	*p*	Contrasts*
Phonetically regular readers	.4	1.6	4.1	23.3	19.08	.0001	S > B, D, L
Leveled readers	10.7	22.1	13.5	1.0	20.50	.0001	D > B, L > S
Basals	0	2.4	1.3	.9	1.04	.3824	
Trade books	8.0	10.4	13.5	12.4	.73	.5377	
Big books	8.7	1.9	.2	.7	17.81	.0001	B > D, L, S
Worksheets	2.5	8.3	5.7	11.1	2.32	.0855	

* $p < .05$.

books was greater in SFA than in classrooms using other models, although the contrast only reached a .10 significance level.

Discussion

Children in high-poverty study schools showed a mixed profile of early literacy attainment at the beginning of first grade. As the letter-and-word identification and word-attack subtests of the WDRB indicate, low-income study children knew upper- and lowercase letters and had a small stock of sight words in the fall of first grade. Average-performing children in the sample had word-analysis skills for reading simple three-letter words (such as "rat") at the beginning of first grade. They could segment simple two-phoneme words (with adult prompting), and they had, on average, the vocabulary knowledge of typical 5-year-olds.

The four literacy reform models appeared to be able to move low-income children with these initial skills to higher achievement levels in word reading and phonemic segmentation by the end of first grade, a considerable accomplishment. Although vocabulary skills improved over the year as well, average vocabulary and reading comprehension abilities, those literacy skills that are related to the construction of meaning, still lagged behind grade-level expectations at the end of first grade.

Despite the enormous debate about the efficacy of different approaches to early literacy instruction, all four models appeared to do an equally effective job of promoting growth in first-grade word reading, word attack, and phonemic awareness. Children within all four reform models showed strong achievement in these basic reading skills by the end of the first-grade year, despite philosophical and practical differences in the pedagogy they had received. Although past research points to a strong advantage for more direct and systematic approaches to teaching word recognition (Ehri, Nunes, Stahl, & Willows, 2001; Foorman et al., 1998), children receiving instruction in models that taught phonics more indirectly showed gains in word reading and decoding skill that were similar to those for children in SFA classes.

The four models appeared to be similarly effective in promoting modest vocabulary growth in low-income children, but not in bringing children close to age-level expectations in this domain. The models showed only slight differences in promoting reading comprehension skill and in fostering writing achievement in first grade.

DLF, the model that placed the greatest emphasis on training teachers in conducting effective guided reading groups, and which was least prescriptive about the types of early reading materials to be employed, showed the greatest success in bringing study children close to grade-level expectations in reading comprehension at

the end of first grade. Because DLF teachers were free to use an eclectic range of texts for read-alouds and small-group reading, they often used more challenging texts than were evident in classrooms using the other models (particularly in comparison to BEL and SFA classrooms). Greater exposure to more complex texts and more time spent discussing these texts appeared to give DLF study children a slight advantage when they were asked to read and make sense of moderately difficult passages on the GMRT.

SFA instruction improved word-reading and word-attack skills for first-grade study children but may have done so at the expense of promoting other literate competencies. Consistent with much previous research, explicit phonics instruction was effective in promoting accurate word reading, but teacher-directed and scripted lessons in comprehension and word meaning, also a feature of first-grade SFA instruction, appeared less successful in supporting achievement in these aspects of literacy. The use of the phonetically regular Roots readers, a central feature of first-grade teaching in SFA classrooms, promoted students' reading fluency, but the simplified character of these readers provided scant exposure to words that were outside students' listening vocabularies and little motivation to explore text meaning. Although SFA Roots classes also included teacher read-alouds of children's picture books and structured discussion of these works, the scripted and closely timed nature of these read-aloud activities limited opportunities for student participation in talk about text. In addition, weak initial vocabulary abilities, a characteristic of many children in SFA classes, posed particular challenges to their development of adequate comprehension skills across the first-grade year.

The modestly better performance of LC students in first-grade writing may have been a consequence of greater emphasis on writing pedagogy in LC professional development. Teachers learned techniques not only for modeling their composing strategies with children (a feature of BEL training, as well) but also for helping children edit and revise the texts they produced when attempting to write independently. Frequent teacher help in editing and revising, evident during classroom observations, appeared to give LC first graders a stronger understanding of sentence structure and writing conventions than was evident for students who only observed teachers employ these conventions in modeled writing. LC's emphasis on learning to read and spell a core group of high-frequency words also appeared to give LC study children a ready stock of words to use when composing independently and thus the confidence to compose longer and more sophisticated texts.

Although we noted some differences in the time spent on different types of literacy activities and with different types of literacy materials, the main finding of this study is that average achievement appears similar for children in one district's urban, high-poverty classrooms implementing four nationally disseminated reform models. The substantially similar first-grade achievement patterns across literacy models may reflect, at least in part, the powerful effects of district policies that accompanied the implementation of the literacy reforms: lowered primary-grade class size, provision of substantial classroom libraries in all primary classrooms, and district-designed professional development that supplemented training provided by the specific models.

The school system also instituted district-wide accountability mechanisms focused on literacy progress, a key practice in other cities that have raised reading achievement (Snipes et al., 2002). Finally, all of the study schools implemented some form of safety-net tutoring for first-grade children who were not making adequate progress in early reading. Because of the district-wide emphasis on improving early literacy achievement, some administrators also shifted their most effective teachers to

first grade in order to give students the best possible start in literacy.

All of these initiatives collectively constituted a kind of district-level reform strategy, with model differences mainly evident in the kinds of texts used in guided reading, the degree of emphasis on early writing instruction, and in the teaching practices used to develop students' word analysis and comprehension skills. Interestingly, these pedagogical differences did not appear to result in different achievement profiles for children receiving literacy instruction in the four models. The similar profiles for average achievement at the end of first grade likely also reflect the homogeneity of the study schools, all of them high poverty, all of them primarily serving African-American and Hispanic students.

Despite differences in pedagogical focus and even core beliefs (e.g., SFA's emphasis on direct teaching of phonics principles; LC's emphasis on embedded phonics teaching), all four models instituted a substantial literacy block of 90 minutes to 2 hours a day. All of the models provided a mixture of whole-class and small-group instruction. All models emphasized the development of reading fluency as the core of first-grade literacy learning and included some emphasis on text meaning, largely through teacher questioning after passages were read. All four models provided some regular form of phonics practice and gave children simplified texts for practicing beginning reading, texts that grew more challenging during the first-grade year. Virtually all study classrooms, regardless of the model being implemented, showed at least adequate student engagement and at least moderately effective management of instruction. These are among the classroom characteristics that have been identified as important for early literacy progress (Wharton-McDonald, Pressley, & Hampston, 1998). Finally, all study schools were in the mature phase of implementation and had been nominated as good implementers of their chosen model. Thus, weak or uncommitted implementation of reform components was largely eliminated as a factor that might contribute to school or model achievement differences.

The largest source of variability in first-grade outcomes, apart from differences in child ability, appeared to be substantial differences in the pedagogical skills and orientations of individual teachers. In the comparatively challenging area of reading comprehension, some study teachers brought 80% of their class to grade-level expectations on the GMRT-4 at the end of first grade, whereas other study teachers brought less than 20% to the grade-level benchmark. Individual teachers varied widely, for example, in the strategies they used for guiding children in how to decode and comprehend text, and in their skill at orchestrating extended talk about text, practices that have been identified as important for early literacy progress (Taylor & Pearson, 2002; Wharton-McDonald et al., 1998). Classrooms also displayed considerable variability in the time children spent reading connected text with teacher support. An irony of literacy reform is that, although the adoption of structured models may provide some leveling of teacher knowledge and resources, there remain significant differences in capacity across teachers, even in a district with substantial supports for improved instruction.

Implementing any of the four models in the context of a district-wide focus on early literacy achievement, class-size reductions, and professional development and coaching for teachers resulted in most low-income study children reaching grade level in basic word reading and decoding. The outcomes of the early literacy reforms appeared not to be as positive for reading comprehension, however, and appeared to have especially limited effects on vocabulary knowledge. Because multiple longitudinal studies have identified early vocabulary scores as among the best predictors of later reading success (Dickinson & Tabors, 2001; Mason et al., 1992), the reform models adopted in this district and elsewhere may

fall short of what is needed to reverse patterns of underachievement in high-poverty schools The unsolved problem for urban schools is how best to simultaneously develop children's early word reading skills, an area of relative strength for the study participants, while also fostering attention to text meaning and accelerating the development of literacy-related language skills.

Appendix

Characteristics of the Four Literacy Models

Building Essential Literacy (BEL). BEL was first developed by Peter Hill and Carmen Crévola at the Center for Applied Educational Research at the University of Melbourne and was formerly called Balanced Early Literacy. BEL professional development is offered through the educational publishing company, Mondo, based in New York City, which has been one of the major sites for the dissemination of BEL, along with Boston, Palm Beach, and Philadelphia. BEL teachers in this study participated in an initial 4-day workshop series emphasizing guided reading and writing and the use of the Clay Observation Survey, followed by additional teacher workshops during the first 3 years of implementation and several site visits by BEL training staff. A primary-grade teacher or reading specialist in each school was released at least half-time (full-time in the larger schools) to serve as an ongoing coach for her peers.

During the reading hour of the 2-hour literacy block, BEL first-grade teachers read aloud simple, often highly predictable big books, often followed by whole-class choral reading of the enlarged text and discussion of text meaning. Students were divided into small homogeneous reading groups that worked with leveled readers, led by the teacher, while the rest of the class rotated through assigned literacy activities. First-grade leveled readers provided support for early reading through text features such as picture cues, repetition of sentence patterns, and repetition of high-frequency words.

During guided reading groups in BEL study classrooms, teachers previewed the text with students, asked about background knowledge, and then supported students with prompts while they orally read the text aloud. Most teachers asked simple questions about text meaning after oral reading. Literacy center activities in first-grade BEL study classrooms included building words with magnetic letters, finding words with a designated spelling pattern in big books or charts, and partner and independent reading of big books, leveled readers, and trade books.

During the hour devoted to writing instruction, BEL teachers modeled and discussed their own writing strategies while composing simple texts such as an account of a weekend happening on a whiteboard. Teachers emphasized developing a topic, using sentence punctuation, and using word walls and other supports for spelling. Students then worked independently on compositions on the same topic with occasional teacher support for individuals as they wrote.

Developing Literacy First (DLF). DLF (formerly the Early Literacy In-Service Course, ELIC) was developed as an in-service course by the educational publishing company Rigby to complement the use of Rigby leveled readers. Primary-grade teachers took an initial 12-week course in reading and writing development followed by on-site coaching from a Rigby consultant during the first year of implementation. The topics for the professional development sequence included reading "cueing systems," principles of shared and guided reading, organizing opportunities for independent reading, and early writing. Guided reading using leveled readers was emphasized as a core literacy practice, along with teacher read-alouds of more challenging texts. DLF study classrooms used Rigby readers, often in combination with basal readers and/or trade books for guided reading groups. DLF schools designated an "in-house facilitator" but did not necessarily release the facilitator from teaching responsibilities.

The 90-minute daily literacy block in DLF study schools included teacher read-alouds of children's fiction, word study activities, and teacher-led reading groups. Word study activities were highly variable across study classrooms, including direct, whole-class instruction of phonics generalizations, word building with manipulatives, worksheets, and sight word drills. Teacher-led reading groups in DLF classrooms emphasized round-robin oral reading, with teacher prompting for decoding strategies, followed by comprehension discussion. Although most DLF classrooms used Rigby and other publishers' leveled readers in reading groups, children also worked with basal reader passages and simple trade books. Writing instruction was a feature of some but not all DLF literacy blocks.

Literacy Collaborative (LC). LC (formerly the Early Literacy Learning Initiative, ELLI) was developed by Irene Fountas at Lesley University and Gay Su Pinnell at Ohio State University to extend Reading Recovery principles to whole-

class instruction. Distinctive features of LC included an emphasis on reading and writing as important focuses for early literacy, and an approach of gradually reducing teacher supports for children's independent reading and composing over time. LC used a "train the trainers" model in which schools released a teacher to receive training in LC theory and pedagogy at a regional site, usually a university. The teacher trainer then reproduced the training sequence at her own school for her peers in a yearlong professional development course, accompanied by on-site demonstration lessons. The teacher-trainer was supposed to continue to conduct a model classroom whose practices other teachers could observe, and ideally she was also supposed to offer ongoing coaching. LC schools were required to provide Reading Recovery tutoring for first graders at risk for reading failure.

In first-grade study classrooms, the reading portion of the LC literacy block began with teacher read-alouds of children's fiction or nonfiction, followed by group discussion of text meaning. Teachers reviewed a weekly list of high-frequency words in a daily, whole-class word study lesson. Children then worked in small groups on independent writing assignments (often linked to a content area topic such as a science observation), independent or partner reading, and word study practice while the teacher led small reading groups. First-grade LC reading groups used leveled readers or simple trade books. Teachers generally engaged students first in previewing and prediction activities, followed by silent or soft oral reading. Teachers posed comprehension questions during or, more frequently, after students had read a passage.

Like BEL classes, teachers in LC classes often modeled composing practices during the writing portion of the literacy block, emphasizing topic development and sentence and spelling conventions. Children worked independently on teacher-assigned writing topics, and in many LC study classrooms, teachers offered one-on-one editing and revising support.

More information about Literacy Collaborative can be found at http://www.lcosu.org/.

Success for All (SFA). Success for All was developed by Nancy Madden and Robert Slavin at the Center for the Study of Schools as Social Organizations at Johns Hopkins University. Teachers in SFA schools receive an initial 3 days of training, and a teacher or administrator is released by the school to serve as on-site supervisor. SFA, like LC, requires that participating schools provide tutoring for first graders at risk for reading failure.

Specifically designed for high-poverty schools, SFA regrouped students by reading levels. In relatively large (8–24) homogeneous Roots reading groups, first-grade-level students received highly scripted instruction, emphasizing direct teaching of letter names and sounds, sight word and phonics pattern drills, and repeated reading of Roots readers, often chorally or in pairs. SFA classes also included teacher read-alouds of trade books selected by the program developers with scripted discussion of story meaning. First-grade-level classes used phonetically regular, leveled Roots readers, but in second-grade-level classes and beyond, designated as "Wings," students worked with school-chosen materials (in our SFA study schools, Houghton Mifflin basal readers) and accompanying comprehension workbooks developed by SFA. Writing received little time allocation in SFA Roots classes but began to be emphasized more at the Wings level. More information about Success for All can be found at http://www.successforall.net/.

Note

This work was supported by grants from the U.S. Department of Education Planning and Evaluation Service, the Noyce Foundation, and Wheelock College.

References

Armbruster, B., Lehr, F., & Osborn, J. (2001). *Put reading first: The research building blocks for teaching children to read*. Washington, DC: National Institute for Literacy.

Barone, D. (2002). Literacy teaching and learning in two kindergarten classrooms in a school labeled at-risk. *Elementary School Journal*, **102,** 415–443.

Borman, G., & Hewes, G. (2002). The long-term effects and cost-effectiveness of Success for All. *Educational Evaluation and Policy Analysis,* **24,** 243–266.

Britto, P. R., & Brooks-Gunn, J. (2001). Beyond shared book reading: Dimensions of home literacy and low-income African-American

preschoolers' skills. In P. R. Britto & J. Brooks-Gunn (Eds.), *New directions for child development: Vol.* **92**. *The role of family literacy environments in promoting young children's emerging literacy skills* (pp. 73–89). San Francisco: Jossey-Bass.

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology,* **40,** 415–436.

Carlisle, J. F., Beeman, M., Davis, L. H., & Spharo, G. (1999). Relationship of metalinguistic capabilities and reading achievement for children who are becoming bilingual. *Applied Psycholinguistics,* **20,** 459–478.

Cooper, R. (1998). *Socio-cultural and within-school factors that affect the quality of implementation of school-wide programs* (Technical Rep. No. 28). Baltimore: Johns Hopkins University, Center for the Social Organization of Schools.

Crévola, C. A., & Hill, P. W. (1998). Evaluation of a whole-school approach to prevention and intervention in early literacy. *Journal of Education for Students Placed at Risk,* **3,** 133–157.

Dahl, K. L., Scharer, P. L., Lawson, L. L., & Grogan, P. R. (1999). Phonics instruction and student achievement in whole language first-grade classrooms. *Reading Research Quarterly,* **34,** 312–341.

Datnow, A., & Castellano, M. (2000a). *An inside look at Success for All: A qualitative study of implementation and learning and teaching* (Report No. 45). Baltimore: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Datnow, A., & Castellano, M. (2000b). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal,* **37,** 775–799.

Dickinson, D. K., & Tabors, P. O. (Eds.). (2001). *Beginning literacy with language: Young children learning at home and school.* Baltimore: Paul H. Brookes.

Donahue, P. L., Finnegan, R. J., Lutkus, A. D., Allen, N. L., & Campbell, J. R. (2001). *The nation's report card: Fourth grade reading 2000* (NCES 2001-499). Washington, DC: National Center for Education Statistics.

Duffy-Hester, A. M. (1999). Teaching struggling readers in elementary school classrooms: A review of classroom reading programs and principles for instruction. *Reading Teacher,* **52**(5), 480–495.

Duke, N. (2000). For the rich it's richer: Print experiences and environments offered to children in very low and very high-socioeconomic status first-grade classrooms. *American Educational Research Journal,* **37,** 441–478.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test* (3d ed.) (PPVT-3). Circle Pines, MN: American Guidance Service.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read. *Review of Educational Research,* **71,** 393–447.

Foorman, B., Francis, D., Fletcher, J., Shatschneider, 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.

Ford, M. P., & Opitz, M. W. (2002). Using centers to engage children during guided reading time. *Reading Teacher,* **55,** 710–717.

Freppon, P. A., & McIntyre, E. (1999). A comparison of young children learning to read in different instructional settings. *Journal of Educational Research,* **92,** 206–218.

Grigg, W. S., Daane, M. C., Jin, Y., & Campbell, J. R. (2003). *The nation's report card: Reading 2002.* Washington, DC: National Center for Education Statistics.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H. Brookes.

Haycock, K. (2000). No more settling for less. *Thinking K–16,* **4,** 3–16.

Hill, P. W., & Crévola, C.A. (1999). The role of standards in educational reform for the twenty-first century. *Yearbook of the Association for Supervision and Curriculum Development,* **1999**, 117–142.

Hill, P. W., & Jaggar, A. M. (2003). *An evaluation report on the impact of Mondo Publishing's Building Essential Literacy (BEL) design and Bookshop Reading Program: 1998–1999 and 1999–2000 school years.* New York: Mondo.

Hurley, E. A., Chamberlain, A., Slavin, R. E., & Madden, N. A. (2001). Effects of Success for All on TAAS Reading: A Texas statewide evaluation. *Phi Delta Kappan,* **82,** 750–756.

Johnson, E. M. (2001). *Developing Literacy First: External evaluation* (Technical Rep. No. 686). Chicago: OER Associates.

Lutkus, A. D., & Weiner, A. W. (2003). *The nation's report card: Trial urban district assessment, Reading highlights 2003.* Washington, DC: National Center for Education Statistics.

MacGinitie, W., & MacGinitie, R. (2000). *Gates-MacGinite Reading Tests* (4th ed.) (GMRT-4). Itasca, IL: Riverside.

Madden, N. A., Slavin, R. E., Karweit, N. L., Do-

lan, L. J., & Wasik, B. A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal,* **30,** 123–148.

Mason, J., Stewart, J., Peterman, C., & Dunning, D. (1992). *Toward an integrated model of early reading development* (Technical Report No. 566). Champaign: University of Illinois, Center for the Study of Reading.

Mosenthal, J., Lipson, M., Torncello, S., Russ, B., & Mekkelsen, J. (2004). Contexts and practices of six schools successful in obtaining reading achievement. *Elementary School Journal,* **104**(5), 343–367.

Nunnery, J., Slavin, R. E., Madden, N. A., Ross, S., Smith, L., Hunter, P., et al. (1997). *Effects of full and partial implementations of Success for All on student reading achievement in English and Spanish.* Baltimore: Johns Hopkins University, Center for Social Organization of Schools. (ERIC Document Reproduction Service No. ED 408 558)

Payne, C. M. (1998). *So much reform, so little change: Building-level obstacles to urban school reform* (Working Paper 98-26). Evanston, IL: Northwestern University, Institute for Policy Research.

Pressley, M. (2003). A few things reading educators should know about instructional experiments. *Reading Teacher,* **51**(3), 64–71.

Puma, M. J., Jones, C. C., Rock, D., & Fernández, R. (1993). *Prospects: The congressionally mandated study of educational growth and opportunity, Interim report.* Bethesda, MD: Abt Associates

Ross, S. M., Smith, L. J., & Casey, J. P. (1997). Preventing early school failure: Impacts of Success for All on standardized test outcomes, minority group performance, and school effectiveness. *Journal of Education for Students Placed at Risk,* **2,** 29–53.

Scharer, P., Williams, E. J., & Pinnell, G. S. (2001). *Literacy Collaborative 2001: Research report.* Columbus: Ohio State University

Smith, L. J., Ross, S. V., & Casey, J. (1996). Multisite comparison of the effects of Success for All on reading achievement. *Journal of Literacy Research,* **28,** 329–353.

Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement.* Washington, DC: Council of the Great City Schools.

Snow, C. E. (1983). Literacy and language: Relationships during the preschool years. *Harvard Educational Review,* **53,** 165–189.

St. John, E. P., Loescher, S. A., & Bardzell, J. S. (2003). *Improving reading and literacy in grades 1-5: A resource guide to research-based programs.* Thousand Oaks, CA: Corwin.

St. John, E. P., Manset, G., Chung, C.-G., Simmons, A. B., & Musoba, G. D. (2000). *Research-based reading interventions: The impact of Indiana's Early Literacy Grant Program.* Bloomington: Indiana Education Policy Center.

St. John, E. P., Manset, G., Chung, C.-G., Simmons, A. B., Musoba, G. D., Manoil, K., et al. (2003). Research-based reading reforms: The impact of state-funded interventions on educational outcomes in urban elementary schools. In L. Mirón & E. P. St. John (Eds.), *Reinterpreting urban school reform: Have urban schools failed, or has the reform movement failed urban schools?* (pp. 129–153). Albany, NY: SUNY Press.

Sutton, A., & Soderstrom. I. (1999). Predicting elementary and secondary school achievement with school-related and demographic factors. *Journal of Educational Research,* **92,** 330–338.

Taylor, B. M., & Pearson, P. D. (Eds.). (2002). *Teaching reading: Effective schools, accomplished teachers.* Mahwah, NJ: Erlbaum.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *Elementary School Journal,* **100**(1), 3–28.

U.S. Department of Education. (2001). *The longitudinal evaluation of school change and performance in Title I schools, Final report* (Vol. **1,** Executive summary). Washington, DC: Planning and Evaluation Service.

Useem, E. L., Christman, J. B., Gold, E., & Simon, E. (1997). Reforming alone: Barriers to organizational learning in urban school change initiatives. *Journal of Education for Students Placed at Risk,* **2,** 55–78.

Venezky, R. L. (1998). An alternative perspective on Success for All. In K. K. Wong (Ed.), *Advances in educational policy* (Vol. **4,** pp. 145–165). Stamford, CT: JAI.

Washington, J. A. (2001). Early literacy skills in African-American children: Research considerations. *Learning Disabilities: Research and Practice,* **16,** 213–221.

Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal,* **99,** 101–128.

Williams, E. J., Scharer, P. L., & Pinnell, G. S. (2000). *Literacy Collaborative: 2000 research report.* Columbus: Ohio State University.

Wong, K. K., & Meyer, S. J. (2001). Title I schoolwide programs as an alternative to categorical practices: An organizational analysis of surveys from the Prospects Study. In G. D. Borman, S. C. Stringfield, & R. E. Slavin (Eds.), *Title I: Compensatory education at the crossroads* (pp. 195–234). Mahwah, NJ: Erlbaum.

Woodcock, R. W. (1997). *Woodcock Diagnostic Reading Battery.* Itasca, IL: Riverside.

Yopp, H. (1995). Yopp-Singer Test of Phoneme Segmentation. *Reading Teacher,* **49**(1), 20–29.