JEAN STOCKARD, Director of Research, National Institute for Direct Instruction
and Professor Emerita, University of Oregon

Direct Instruction and First Grade Reading Achievement: The Role of Technical Support and Time of Implementation

Abstract: The purpose of this study was to examine the relationship of two implementation variables (source of technical support and time of implementation) to first grade reading achievement across three naturally-occurring conditions over a six-year time span. Two of the conditions involved implementation of Direct Instruction (DI) curriculum programs, one with technical support from a purveyor associated with the curriculum's developer and the other with a group not associated with the developer. The third condition involved implementation of another highly rated reading curriculum program: *Open Court*. Increases in the vocabulary and comprehension achievement scores of students across all three conditions were statistically significant. Increases in students' scores were significantly greater in the DI implementation conditions relative to the *Open Court* condition. Increases in students' scores were significantly larger in the DI implementation condition with technical support from the purveyor associated with the curriculum developer relative to the group not associated with the curriculum developer. Differences in achievement across the implementation conditions were both statistically significant and educationally meaningful.

Many have documented the strong relationship between students' academic achievement and their later academic and economic success and social well being (Farkas, 1996; Heckman, 2006; Jencks, Crouse, & Mueser, 1983). Early reading achievement is one of the most important predictors of later academic accomplishments. Students who are poor readers in first grade have substantially higher probabilities of later academic problems than students who achieve at grade level at that time (Juel, 1988). Problems of consistent low student achievement and poor performing schools have been most marked in urban settings (Stringfield & Land, 2002; Stringfield & Yakimowski-Srebnick, 2005).

These consistent and strong research findings have prompted extensive policy attention, especially in large urban school systems, to promoting first grade reading achievement and finding the most effective curricular-reform models. An extensive body of research has demonstrated the effectiveness of systematic and explicit instruction in promoting students' reading achievement (National Institute of Child Health and Human Development [NICHD], 2000; Juel & Minden-Cupp, 2000; Murphy, 2004; National Reading Panel, 2000). However, controversies continue in the "phonics–whole language debate," and the more highly structured curricula are still far from universally adopted (Foorman, 1995; Hempenstall, 1997, 1999). In addition, it is not clear whether all "systematic and explicit" curricula are equally effective (Engelmann, 2004).

While the literature on systematic and explicit instruction in promoting reading achievement is quite large, less attention has been given to

how reading programs are implemented. The general literature on implementation identifies two elements of special importance: a) the nature of the technical support provided to practitioners and b) providing sufficient time for a reform to be fully implemented and institutionalized. Although studies have examined one or the other of these elements, as well as more detailed elements of fidelity to a program, we have found no studies simultaneously examining the relationship of these two broad-based implementation variables to the development of higher reading achievement.

This paper addresses these gaps in the literature. Using data from a large urban school district we examined differences in the changes in first grade reading achievement of students in three groups of schools over six years. One group of schools implemented the highly structured Direct Instruction (DI) program with technical support from a purveyor associated with the developer of the program, another group implemented DI with an alternative source of technical support, and the third group used *Open Court* (OC). We examined the extent to which reading achievement increased in schools employing DI and OC, the association of variations in the source of technical support with changing achievement levels, and the extent to which the positive effects of systematic and explicit instruction become apparent as curricular programs are more fully institutionalized and stabilized within a school.

Background Literature

Effective Reading Curricula and Direct Instruction

Literature spanning several decades documents the importance of systematic and explicit instruction in promoting reading achievement (Adams, 1990; Anderson, Hiebert, Scott, & Wilkinson, 1985; Baker, Kameenui, Simmons, & Stahl, 1994; Bond & Dykstra, 1967; Chall, 1967; Foorman, 1995; Fukkink & deGlopper, 1998; Grossen, 1997; Juel & Minden-Cupp, 2000; Murphy, 2004; National Reading Panel, 2000; NICHD, 1996, 2000; Pflaum, Walberg, Karigianes, & Rasher, 1980; Smith et al., 2001; Snider, 1990; Snow, Burns, & Griffin, 1998; Stanovich, 1994). Meta-analyses that examine specific curricula also support this conclusion, showing that curricula that embody the specific and explicit elements consistently result in larger achievement gains (Adams & Engelmann, 1996; AFT, 1998; Beck & McCaslin, 1978; Borman, Hewes, Overman, & Brown, 2003; Hattie, 2009; Herman et al., 1999).

One explicit instructional approach is Direct Instruction (DI) (distinguished by its use of capital letters from other "direct instruction" approaches that embody only some of DI's characteristics), which was developed by Siegfried Engelmann and Wesley Becker (Engelmann & Carnine, 1982; Engelmann, 2007). DI curricula are specifically designed to accelerate students' learning by teaching more than traditional programs in the same amount of time. Unlike many curricula, DI programs are extensively field tested before publication to ensure they produce the greatest learning in the most efficient manner. The programs, commercially available through SRA/McGraw Hill, involve scripted lessons designed to provide teachers with the most effective wording to allow the presentation of tasks to students at a relatively high rate. The amount of new material introduced in each lesson is carefully controlled. Applications become increasingly complex and are designed for children to master the content presented by the end of each lesson. Mastery tests are included to help ensure students have made the expected progress (Collins & Carnine, 1988; Engelmann, 2007; Engelman & Carnine, 1982; Huitt, Monetti, & Hummel, 2009).

Numerous studies have documented the effectiveness of DI in promoting achievement, and several meta-analyses have summarized these results. For instance, Borman et al. (2003)

examined studies of 29 comprehensive school reform models. They found the most evidence was available for the DI model revealing "49 studies with 182 outcomes" compared to a median of four studies and 23 outcomes (p. 141). DI was found to produce the strongest effects (d = .21) of all models examined. It was one of three models meeting criteria of "strongest evidence of effectiveness," which involved replication of the outcomes "in a number of contexts, …statistically significant and positive achievement effects in studies using comparison groups or third-party comparison designs and…accumulated evidence from at least 5 third-party comparison studies" (p. 161). More recently, Hattie (2009) summarized the results of four meta-analyses that included DI, incorporating 304 studies, 597 effects and over 42,000 students. He found the average effect size associated with DI was .59 and noted the positive results were "similar for regular (d = .99) and special education and lower ability students (d = .86), … [and] similar for the more low-level word-attack (d = .64) and also for high-level comprehension (d = .54)" (pp. 206-207). (See Adams & Engelmann,1996; AFT, 1998; Beck & McCaslin, 1978; and Herman et al., 1999 for other meta-analyses incorporating DI.)

Open Court, the other reading program examined in this study, is also a phonics-based, highly structured program and is often cited as an example of direct instruction (lower case) (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998). Various review bodies, including the Florida Center for Reading Research and the Oregon Reading First Center, have noted it contains all of the elements needed to promote learning to read (FCRR, 2004; ORFC, 2004). According to Dowling, and Schneck (2008), "there is a limited but growing body of research that has provided some support for the OCR [Open Court Reading] program" (p. 391)[1]. Using data from a multisite cluster randomized field trial of the program, Borman et al. (2008) found a small advantage for the *Open Court* program over other programs, with effect sizes of .16 for Reading Comprehension and .19 for Vocabulary.

Two studies have directly compared the effectiveness of *Reading Mastery*, the principal DI reading program, and *Open Court*. Both studies found students using *Reading Mastery* had significantly higher achievement than students using the other curricula (Crowe, Connor & Petscher, 2009; O'Brien and Ware, 2002). Our first research question addresses these differences, examining the extent to which reading achievement increases when schools employ highly systematic and explicit curricula and if schools employing DI and those using *Open Court* have different rates of change over time

Purveyors of Technical Assistance

In recent years, studies of effective implementation of social programs, primarily in the social services and health-related fields, have become more common. Noting "desirable outcomes are achieved only when effective programs are implemented well" (Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005, p. 12), researchers in the implementation research tradition examine how a program's effectiveness is influenced by the way it is implemented. One of the key elements identified is the role of "purveyors," a term used to refer to those who work with practitioners to ensure that a practice or program is implemented "with fidelity and good effect" (Fixsen et al., 2005, p. 14).

Within education, numerous authors highlight the key role of technical assistance in promoting teachers' skills and their fidelity of implementation. The literature increasingly recognizes teaching is a highly technical and

1 These include Edsource, (2006); McRae (2002); Skindrud & Gersten (2006), who reported positive results; and Westat (2001), who found no differences in achievement of OC students and those in a control group).

involved process, and training and support are crucial for developing and honing excellent teaching skills. Studies suggest this assistance should be ongoing and intensive, ideally involving on-site support (Berends, Bodilly, & Kirby, 2002; Blakeley, 2001; Bodilly, 1998; Bodilly, Glennan, Galegher, & Kerr, 2004). Such support may be especially important for programs such as DI that require more extensive changes in teacher behavior (Engelmann & Engelmann, 2004). As expected, studies have found the gap between students in DI programs and those in traditional programs is greater for students of teachers who implemented DI with higher fidelity (Gersten, Carnine, Zoref, & Cronin, 1986). Similarly, studies focused only on students receiving DI have found the highest achievers in classrooms or schools with higher levels of fidelity of implementation (Benner, Nelson, Stage, & Ralston, 2010; Carlson & Francis, 2002; Gersten, Carnine, & Williams, 1982; Ross et al., 2004).

Studies of program implementation within the field of health promotion and health care have indicated "a high level of involvement by program developers on a continuing basis is a feature of many successful implementation programs" (Fixsen et al., 2005, p. 21). Studies that have compared varying levels and sources of support indicate the most effective outcomes occur when the purveyors of this support are those most familiar and experienced with the program. We have not found similar studies regarding the relationship of variations in the purveyor of technical support to the effectiveness of curricular programs, but could ask if similar results would occur in education. Thus, our second research question is whether changes in reading achievement are greater in schools that receive support from those most knowledgeable and experienced with a curricular program.

Full Implementation and Achievement

Another relevant element within the implementation literature is the notion of *stages of implementation*. Stressing implementation of a program "is a process, not an event," reviewers stress full implementation only occurs when practices and procedures have been fully incorporated within the day-to-day operations and at all levels of an organization (Fixsen et al., 2005, p. 15). Analyses of whole-school reform efforts stress the amount of time needed to bring about extensive and meaningful change. For instance, in a review of studies of whole-school reform, Bodilly (1998) noted "it takes minimally five years, if it can be accomplished at all" (p. 19). An extensive meta-analysis of comprehensive school-reform efforts by Borman et al. (2003) supports this conclusion, finding effect sizes associated with a curriculum increased substantially after the fifth year of implementation[2].

The DI literature concurs with this assessment and adds a distinction between "implementation" of a model and "stabilization." Engelmann and Engelmann (2004) contend stabilization occurs when four factors are present: a) DI is fully implemented, b) the changes have been fully institutionalized within a school setting, c) the new system is fully familiar to both students and teachers, and d) the system has been present for enough years that a child who began at the school in kindergarten would have experienced the model throughout the elementary years. Based on decades of experience and data from numerous implementations, Engelmann and Engelmann (2004) suggest that while a site may implement DI in as little as two years, stabilization does not occur "until about the sixth year of implementation" (p. 117).

Engelmann and Engelmann (2004) also note the importance of assessing the efficacy of a

2 See also Bodilly et al. (2004) and Borman (2007).

school reform after it has been stabilized, or fully institutionalized. However, of all the studies regarding DI that we have reviewed, only a few examined data from more than one year (Addison & Yakimowski, 2003; Carlson & Francis, 2002; MacIver & Kemper, 2002; O'Brien & Ware, 2002; Vitale & Joseph, 2008). While the results of these studies suggest that achievement may increase as DI becomes more fully implemented and moves toward stabilization, only one had sufficient information to examine the impact of DI when the reforms were stabilized within the schools. Vitale and Joseph (2008) examined the percentage of elementary students who met state-defined proficiency levels in a very low income, North Carolina community. The data set began the year before implementation of DI and continued through six subsequent years. As expected, the percentage of students meeting proficiency continued to rise. The percentage of third through fifth graders ranked proficient moved from less than 25 percent before implementation to over 70 percent after six years, the point at which Engelmann and Engelmann (2004) would suggest that the program could be termed stabilized. Reviews of the general implementation literature suggest this lack of long-term research is relatively common. Most evaluations of program implementation occur within the initial implementation stage rather than after a program has become fully operational, stabilized, and part of an organization's ongoing structure (Fixsen et al., 2005).

Summary

The literature reviewed above suggests students have higher levels of reading achievement when provided systematic and explicit instruction, and the DI model may be more effective than the *Open Court* series. In addition, literature on program implementation suggests both the purveyor of technical support and the amount of time a program has been implemented influence program effectiveness. In our review of the literature we found no studies comparing achievement of students in schools receiving support from different purveyors of technical assistance and only one empirical study of DI tracking achievement over a time period equal to what is generally seen as necessary for full implementation and stabilization. Most important, none of the studies found explicitly examined how these two implementation variables jointly affect achievement.

This paper addresses these gaps. We compare first grade reading achievement of students receiving DI with students receiving *Open Court*, another highly regarded curriculum, in a large urban school system over a six year time period. In addition, we examine the extent to which any trends in first grade achievement are related to the purveyor of technical assistance and the amount of time a reform has been implemented within a school. We examined three guiding questions:

1. To what extent does reading achievement increase when schools employ highly systematic and explicit curricula and are there differences in the increases between schools employing DI and those using *Open Court*?

2. To what extent is receiving technical support from a purveyor associated with the development of DI associated with changes in achievement over time?

3. To what extent do the positive effects of systematic and explicit instruction become apparent as the curricular programs are more fully institutionalized and stabilized within a school?

Two previous studies (Addison & Yakimowski, 2003; MacIver & Kemper, 2002) examined data from some of the schools in our analysis. Our work differs from their analyses in three ways. For instance, these studies examined data from only a three to four year time period, and we examined data over six years, allowing

for the possibility that the reforms would have become fully implemented. Neither of these studies controlled for differences in the purveyors of the reform, which is a key variable in our analysis. Finally, MacIver and Kemper (2002) only looked at data from 12 schools (6 DI and 6 control schools), and we examined data from a much larger group. Notably, however, when our analyses can be compared, the results were similar.

Methodology

The data for this analysis were provided by the Baltimore City Public School System (BCPSS) and include the reading achievement of first graders for six years: 1997-98 through 2002-03. The sections below describe the procedures, , participants, measures, and analysis techniques that were used. The author was not involved in the data gathering process or in any implementations or instruction in the BCPSS. The work reported here is secondary analysis of data provided by the school system.

Procedures

The BCPSS is similar to many other large city school districts serving students with high levels of poverty and struggling with low achievement. In the late 1990s, curricular reforms were implemented in the BCPSS elementary schools to address this low achievement. Sixteen schools, many of which were among the lowest performing in the system, chose to use DI for reading instruction[3]. The curriculum was introduced as part of a whole-school reform effort called the Baltimore Curriculum Project sponsored by the Abell Foundation. The other schools followed the BCPSS curricular guidelines (Addison & Yakimowski, 2003; Berkeley, 2002; MacIver and Kemper, 2002; Stringfield & Yakimowski-Srebnick, 2005).

Direct Instruction. The major DI reading curriculum used was *Reading Mastery Classic*, a scripted, mastery-based core reading program that focuses on decoding and comprehension. It utilizes a specialized orthography to help students discriminate between confusing letters and letter combinations. Because the implementation was part of a whole-school reform effort, the students in the DI schools also received language instruction for 30 minutes per day. The curricula used for this instruction were *Language for Learning, Language for Thinking,* and *Reasoning and Writing*. These are general-knowledge programs that focus on oral language development. The specific program employed for each child depended upon his/her skill levels and placement.

DI is designed to elicit frequent oral student responses, which increase engagement and create a high rate of active responding. Student skill is continuously monitored through observations, in-program tests, records of lessons completed, and at least weekly check-ins. Students are placed in homogeneous groups according to skill level. Ideally, the teacher ensures all members of the group achieve mastery on all material the program introduces. Students who master content substantially faster or slower than others in their group are placed into other groups in which students have skill profiles similar to those of the incoming student. Informal regrouping occurs throughout the school year, and schoolwide regrouping can occur as often as four times a year.

At the beginning of implementation, the Baltimore Curriculum Project contracted with the National Institute for Direct Instruction (NIFDI) to provide implementation support through preservice and in-service training, coaching, and technical assistance for all the DI

3 One additional school was part of the original DI intervention group but was closed shortly after the start of the study period. Because data were not available throughout the time span of the study, data from that school were not included.

schools. The developer of DI is affiliated with NIFDI, and the organization prides itself on strict fidelity to the DI model as it was originally validated through extensive field testing (see Engelmann, n.d.). When a school implements the NIFDI model, instructional programs are phased in over several years. In the first year, language and reading programs are introduced in kindergarten and first grade. In subsequent years mathematics is added and the curriculum is expanded to the higher grades.

The NIFDI model encompasses the elements found in the implementation literature to be especially effective and important in technical support (Fixsen et al., 2005). A NIFDI Implementation Manager (IM) trains teachers, assistants, and coaches. The IM is typically on-site about 35 days per year, working in classrooms with the teachers and presenting in-service sessions that address problems teachers are experiencing. There also are weekly conference calls to address any problems a classroom is experiencing in meeting projected performance gains. All teachers receive preservice training and coaching until the IM determines they can teach each program to a minimum adequate level of fidelity. Teachers continue to receive in-service coaching to improve implementation fidelity. During the second year, teachers who perform well are identified as coaches and are deployed to work with other teachers in the school. Beginning in the third year NIFDI support is gradually phased out so schools can become more self-sufficient. However, the Baltimore Curriculum Project, operating as charter schools, continued to fund some coaching and consultation to the DI schools.

Five of the sixteen schools that introduced DI ceased their association with NIFDI soon after beginning implementation. Although they continued to implement the DI curriculum, they worked with another purveyor of technical support. Both NIFDI and the other purveyor supplied implementation training and consultations to teachers and other school staff. However, because the NIFDI staff includes the original developers of the DI curriculum, it could be expected, based on the implementation research in other areas (e.g. health and social services), the NIFDI-sponsored implementations would be more likely to conform to the full tenets of DI. Some support for this assumption comes from MacIver and Kemper's (2002) description of some of the other DI schools, indicating they "were not willing to implement all dimensions of the whole-school reform program" and even though they continued use of *Reading Mastery* and the other DI programs, this implementation was "not according to the exact specifications of the original developer" (p. 200, footnote 4). In the discussion below we differentiate the NIFDI-supported (NSDI) schools from the Other DI (ODI) schools.

Open Court schools. Stringfield and Yakimowski-Srebnick (2005) provide an extensive description of developments in the BCPSS from the late 1990s through the early part of the new century. According to their account, the other schools in the BCPSS were free to use any curriculum program they desired before 1998, and there was no district-wide structured reading program. Yet, system schools had substantially lower achievement than other schools in the state, and there was extensive pressure for change. In response to these pressures the BCPSS embarked on a concentrated and extensive reform process.

As part of these changes the system adopted new citywide reading curricula, using *Open Court Reading* in kindergarten through second grade. *Open Court* is a phonics-based, highly structured program that has been favorably reviewed by the Florida Center for Reading Research and the Oregon Reading First Center as a core instructional program for Reading First (FCRR, 2004; ORFC, 2004). The BCPSS provided extensive professional development support in 1998-99 for teachers as they began to use this new curriculum. At the same time, the system introduced other reforms through-

out the system, such as lowering student/teacher ratios, expanding kindergarten programs to full day, and expanding before-school, after-school, and summer school programs for students and schools considered most at risk (Berkeley, 2002; Stringfield & Yakimowski-Srebnick, 2005). These reforms remained in place throughout the duration of the study.

Participants. The participants for this analysis included approximately 45,000 first grade students enrolled in the BCPSS from 1997-98 through 2002-03.[4] On average, data were available for about 380 students per school. Only four schools, all in the *Open Court* group, had fewer than 100 students in the sample. Table 1 reports the number of students and number of schools in each group for each year of implementation. The total number of students was about 3,000 in the NSDI schools, 1,800 in the ODI schools, and 40,000 in the *Open Court* (OC) schools.

Enrollment in the system declined over the time period used in the analysis, with data available for approximately 9,000 first graders in 1998 and 6,400 in 2003. Reflecting the declining enrollment, the number of schools also changed over time (i.e. some were consolidated and closed). When the sample was restricted to include only students in schools that were open at both the beginning and the end of the data period the results were identical to those reported here.

Table 1 also reports the average racial-ethnic composition and average percentage of students receiving free or reduced lunches within the three groups of schools. Like many large urban districts, most schools had large percentages of African American students, only a minority of non-Hispanic white students, and a majority of students receiving free or reduced lunch. The NSDI schools averaged the lowest percentage of African American students, but the highest percentage receiving free or reduced lunch. The ODI schools had the highest percentage of African American students, but the lowest percentage of students receiving free or reduced lunch.

Measures. The Comprehensive Test of Basic Skills/TerraNova (CTB/ McGraw-Hill, 2001), a widely used standardized achievement test, was administered to all students in the spring of each year from 1998 through 2003 as part of a system-wide testing program . The CTB is reported to have strong content validity and reliability: "in the .90s for the complete battery and the .80s for individual tests" (Cizek, 1998, p. 22). Testing was conducted by district staff independently of the various implementers and purveyors of technical support. The fourth edition was administered in the spring of 1998 and 1999, and the fifth edition in each subsequent spring. Two subtest scores, Reading Comprehension and Reading Vocabulary, were reported. Normal Curve Equivalent (NCE) scores were used for all analyses to help ensure comparability from one year to another and to allow the use of statistical calculations. Because the meaning of NCE scores is not intuitively obvious, the scores also have been converted, when appropriate for descriptive purposes, into percentiles using a standard conversion table.

Receiving DI and the source of technical support were measured by two dummy variables, one indicating students were in schools that taught DI and had support from NIFDI and the other indicating they implemented DI

4 There were two schools in the control group that, unfortunately, had the same name. The number assigned to the schools was not available in the data set for all years, and alternative spellings of the schools' name across years made it impossible to clearly differentiate them. Thus students in these two schools were eliminated from the analysis. Data were available for one home-schooled student and that was also omitted. Results did not alter if these students were included.

Table 1

Number of Students and Schools, by Years of Implementation and Group, and School Characteristics by Group

Number of Students

Years of Implementation	Group NSDI	ODI	OC	Total
0	341	0	8220	8561
1	417	90	7464	7971
2	588	369	7146	8103
3	562	318	6607	7487
4	517	322	6093	6932
5	467	283	5758	6508
6	295	279	0	574
7	193	191	0	384
Total	3380	1852	41288	46520

Number of Schools

Years of Implementation	Group NSDI	ODI	OC	Total
0	5	0	101	106
1	7	1	102	110
2	11	5	102	118
3	11	5	98	114
4	11	5	95	111
5	11	5	92	108
6	6	5	0	11
7	4	4	0	8

School Characteristics

	Group NSDI	ODI	OC	Total
Average % African American	75	93	85	84
Average % Non-Hispanic White	17	6	14	14
Average % Free or Reduced Lunch	83	72	74	75

Note: The school characteristics were calculated with schools as the unit of analysis. The values are similar when students are used as the unit of analysis. Four of the NSDI schools and four of the ODI schools began DI reading instruction in 1996-97, two more NSDI schools and the remaining ODI school began DI in 1997-98 (the first year of data available for analysis), and the final five NSDI schools began DI in 1998-99. All of the Open Court schools began implementation of OC in 1998-99, the second year for which data were available.

with support from another purveyor. The omitted category was the OC schools.

To examine changes over time, we calculated the number of years that the reforms had been implemented. The number of students and the corresponding number of schools for each value of this variable are shown in the first two panels of Table 1. For the OC schools, the measure of implementation time varied from 0, corresponding to data from the spring of 1998, to 5, corresponding to the spring of 2003, the final year of data collection and five years after implementation of the *Open Court* curriculum. There was more variability in this measure for students in schools that implemented DI. Four NSDI schools and four ODI schools began DI reading instruction in 1996-97, two more NSDI schools and the remaining ODI school began instruction in 1997-98, and the final five NSDI schools began DI in 1998-99. As a result, values for the years of implementation for NSDI schools ranged from 0 (for the five schools in 1997-98 that had not yet begun DI) to 7 (for students in 2002-03 in the four schools that began implementation in 1996-97). As shown in Table 1, all schools in the ODI group had at least one year of implementation at the start of data collection. Thus, values of the measure of years of implementation for this group ranged from 1 to 7.[5]

Finally, we used, as a control variable, the proportion of students within a school receiving free or reduced lunch. This control variable is included to adjust for the impact of socioeconomic status on student achievement. Because we did not have access to individual-level demographic data, a school-level measure was used as a proxy. In preliminary analyses we also used a factor score incorporating measures of race-ethnicity and school poverty as a control variable. Results, available from the author, were identical to those reported here.

Analysis. The relationship of student achievement to receiving DI, the source of technical support, and time of implementation was examined with mixed model regression. Mixed models are particularly appropriate for analyzing multilevel data, such as those regarding students and the schools they attend. In these models a "random variable" is used to control for differences between schools (often termed the Level 2 entity) while calculating regression coefficients regarding the impact of variables from both students and schools on achievement. The random variable is roughly equivalent to having a separate intercept in the regression equation for each school. The coefficients associated with the independent variables are then calculated while this between-school variance is controlled. The analysis also allows one to calculate the amount of variance in the dependent variable that occurs between schools (Raudenbush & Bryk, 2002; Singer, 1998).

We report results for a model that includes school grouping, years of program implementation, the measure of school poverty and the interaction of school group and years of implementation. To illustrate the results we use the coefficients from the mixed models to calculate expected NCE achievement scores for each group of schools for each year of implementation, translate these into the correspon-

5 As noted above, the BCPSS only introduced funding for full-day kindergarten in the 1998-99 school year. The Direct Instruction model is designed for a full-day kindergarten, seeing this early academic focus as a key contributor to accelerating learning. Because the full day program was not universal until 1998-99, first graders in the DI schools would not have had the full DI program until 1999-2000, the third year for which data were available. In addition, the Baltimore schools had high rates of student mobility and, thus, some of the first graders in the DI schools may not have had DI in kindergarten. We had no information on where the students attended kindergarten. However, any impact this factor could have on the results would be to minimize differences (making the advantage of DI smaller), thus providing a conservative influence.

ding percentiles, and also calculate effect sizes (Cohen's *d*) of differences between the groups, using the predicted NCE scores, for each year of implementation.

Results

Table 2 gives the results of the mixed model analyses. As a preliminary step, baseline "intercept only" or "random effects" models, which only included schools as a random variable, were run. (This is equivalent to a simple one-way analysis of variance.) The resulting correlation ratio, or proportion of variance in the dependent variable that was between schools as opposed to between students, was approximately .10 for both dependent variables. The estimates, z-values, and probabilities associated with the random effects for the models in Table 2 (in the bottom rows of the table) test the null hypothesis that the variation between schools equals zero once vari-

Table 2

Mixed Model Regressions of First Grade Achievement on Year, School Poverty, and Years of Implementation

	Vocabulary			Comprehension		
Fixed Effects	*b*	*SE*	*p*	*b*	*SE*	*p*
Intercept	56.78	3.08	<.0001	49.24	2.99	<.0001
Poverty (% FRL)	-0.22	0.04	<.0001	-0.18	0.04	<.0001
Years of Implementation	2.20	0.06	<.0001	4.20	0.06	<.0001
ODI School	-5.35	3.19	0.09	-7.44	3.11	0.02
NSDI School	-9.98	2.20	<.0001	-7.70	2.15	0.0003
Years * ODI	0.62	0.31	0.04	0.69	0.30	0.02
Years * NSDI	3.37	0.23	<.0001	2.67	0.23	<.0001
Random Effects Est.		40.00			37.86	
SE		5.43			5.13	
p		<.0001			<.0001	
Residual Estimate		477.98			466.14	
SE		3.17			3.08	
p		<.0001			<.0001	

Note: The fixed effects parameters can be interpreted as unstandardized regression coefficients, indicating the expected change in the normal curve equivalent vocabulary and comprehension scores with a unit change in a given independent variable. The t-values associated with each parameter can be obtained by dividing the coefficient (the *b*) by the standard error (*SE*). The probabilities associated with these t-values are given in the columns and rows labeled *p*.

ables in the model are controlled. These values associated with the residual test the null hypothesis that variation between individuals equals zero once the variables in the model are controlled. These null hypotheses can be easily rejected for both dependent variables. There was significant unexplained variation between schools and also between students. This would be expected because there are many factors that can influence student achievement in addition to those available for this analysis.

The coefficients reported in Table 2 can be interpreted as regression coefficients. As expected, the negative coefficients associated with the measure of school poverty indicate students attending schools with larger populations of students receiving free or reduced lunches had significantly lower vocabulary and comprehension scores. At the same time, and independently of school context, the positive coefficients associated with years of implementation indicate average first grade achievement scores were significantly higher when the reforms had been implemented for longer periods of time. However, the significant interaction effects indicate the rate of improvement from one year to the next varied significantly among the three groups. The strongest gains were in the NSDI schools and the smallest gains in the OC schools. In the OC schools, controlling for poverty composition, average first grade vocabulary achievement scores increased by about 2.2 NCE points each year the curriculum was implemented, and comprehension scores increased by 4.2 NCE points. (These values equal the coefficients associated with years of implementation because OC schools were the reference, or zero category, in the equations.) In the ODI schools, the results indicate a predicted yearly increase of 2.8 (2.2 + .6) NCE points in average vocabulary scores and 4.9 (4.2 + .7) NCE points in average comprehension scores. In the NSDI schools the yearly expected increase, net of poverty status, was 5.6 (2.2 + 3.4) NCE points on the vocabulary measure and 6.9 (4.2 + 2.7) NCE points on the comprehension measure.

Table 3 reports the achievement scores that would be predicted, given the results of the mixed models, for each of the three groups in each year of implementation. A level of school poverty equal to the system's average (75% free or reduced lunch was used for the calculations to equalize the schools on this variable).[6] The first panel of Table 3 reports the calculated NCE scores and the second panel translates these scores into the corresponding percentiles. The percentiles may be interpreted as the score of an average student in a given group of schools in each year of implementation, if the schools had similar percentages of students receiving free or reduced lunch. Because there were no OC schools with more than five years of implementation and no ODI schools with data from zero years of implementation, the results for those cells are extrapolations from the trends presented by the equations.

The results in Table 3 illustrate how first grade achievement changed over the years of implementation in each group of schools. Before implementation started, the average student in each group had achievement scores that were well below the national average, ranging from the 17th to 32nd percentile in vocabulary and the 15th to 25th percentile in comprehension. At that point the highest scores were in the OC schools and the lowest in the DI schools. After seven years of implementation, the predicted scores were well above the national norm. The average stu-

6 The pattern of results is identical when unadjusted raw scores are compared although, as would be expected given the differences between the groups in the percentage of students receiving free or reduced lunch, the actual numbers differ.

dent in each group of schools, once school poverty was equalized, had vocabulary scores ranging from the 58th to the 82nd percentile and comprehension scores ranging from the 73rd to the 89th percentile. At this point, scores of students in the OC and ODI schools were similar, with the average student in the ODI schools having achievement equal to that of those in the OC schools. The average score of students in the NSDI schools

Table 3

Predicted NCE Achievement Scores by Group and Years of Implementation, Corresponding Percentiles and Effect Sizes of Differences

Predicted NCE Scores

Years of Implementation	Vocabulary			Comprehension		
	NSDI	ODI	OC	NSDI	ODI	OC
0	30	(35)	40	28	(28)	36
1	36	37	42	35	33	40
2	41	40	44	42	38	44
3	47	43	47	49	43	48
4	52	46	49	56	48	53
5	58	49	51	63	53	57
6	63	52	(53)	69	58	(61)
7	69	54	(55)	76	63	(65)
% Change	130	55	37	171	124	81

Corresponding Percentiles

Years of Implementation	Vocabulary			Comprehension		
	NSDI	ODI	OC	NSDI	ODI	OC
0	17	(24)	32	15	(15)	25
1	25	28	36	24	21	32
2	34	32	40	35	29	39
3	44	37	44	48	37	47
4	54	42	48	61	46	55
5	65	48	52	72	56	63
6	74	53	(56)	82	64	(70)
7	82	58	(60)	89	73	(74)

continued on next page

was markedly higher than the other two schools, reflecting the much greater year to year gain within this group.

Because NCE scores have a common standard deviation (21.06), the values in the first panel of Table 3 can be used to calculate effect sizes that illustrate the substantive strength of the differences among the groups. The effect sizes for comparisons between each group for each year of implementation are shown in the third panel of Table 3. Employing the often used criterion of .25 as an educationally meaningful result (Wolf, 1986), it can be seen that the advantage of the OC schools over both the ODI and NSDI schools reached this level before implementation began and, to some extent, in the early years of implementation.

By the later years, the effects for comparisons with the ODI schools were no longer educationally meaningful. However, the comparisons with the NSDI schools changed in sign with strong, educationally meaningful effects in favor of the NSDI schools from Years 5 through 7 (effect sizes ranging from .33 to .67 for the measure of vocabulary and .27 to .54 for the measure of comprehension). Similarly, the effect sizes comparing the NSDI and ODI schools show educationally meaningful effects in favor of the NSDI schools by the later years. Effect sizes range from .30 to .70 for Years 4 to 7 for the vocabulary scores and from .27 to .65 for Years 3 to 7 for comprehension. The effect sizes were larger for the latest years of implementation.

Table 3, continued

Effect Sizes

Years of Implementation	OC v. ODI	OC v. NSDI	ODI v. NSDI	OC v. ODI	OC v. NSDI	ODI v. NSDI
0	(.24)	0.48	(.24)	(.37)	0.37	(-0.01)
1	0.23	0.32	0.09	0.32	0.24	-0.08
2	0.2	0.16	-0.04	0.29	0.11	-0.18
3	0.17	-0.01	-0.17	0.26	-0.01	-0.27
4	0.14	-0.17	-0.3	0.22	-0.14	-0.36
5	0.11	-0.33	-0.44	0.19	-0.27	-0.46
6	(0.07)	(-0.50)	-0.57	(0.15)	(-0.40)	-0.55
7	(0.03)	(-0.67)	-0.70	(0.11)	(-0.54)	-0.65

Note. The NCE values were calculated using the coefficients in Table 2, with a free or reduced lunch percentage of 75. The following equations were used with NCE = NCE score and IY = implementation year: For vocabulary, for OC Schools, NCE = 40.3 + 2.2 (IY); for ODI Schools, NCE = 34.9 + 2.8 (IY); and for NSDI Schools, NCE = 30.3 + 5.6 (IY). For comprehension, for OC Schools, NCE = 35.7 + 4.2 (IY); for ODI Schools, NCE = 28.3 + 4.9 (IY); and for NSDI Schools, NCE = 28.0 + 6.9 (IY). Effect sizes were calculated for each pair of schools and each implementation year using the predicted NCE scores and the standard formula for Cohen's d of $(M_1-M_2)/SD$, where SD = 21.06. Percentiles were calculated with a standard conversion table and represent the percentile at which an average student would score in a given year and group. Parenthesized values are those that fall beyond the range of actual data for years of implementation for a group: Years 6 and 7 for the OC schools and Year 0 for the ODI schools.

Summary and Discussion

The sections below summarize the results of the analysis, discuss limitations of the study and directions for further research, and examine possible implications for practitioners and policy makers.

Summary of the Findings

Our first research question asked if students would have higher reading achievement when using highly systematic and explicit curricula and whether those receiving DI would have higher achievement than those using *Open Court*. First grade reading achievement rose in all schools throughout the BCPSS as a more explicit and systematic curriculum was implemented throughout the system. At the same time, these increases were significantly stronger in schools that implemented DI, replicating other studies that have compared these curricula (Crowe et al., 2009; O'Brien & Ware, 2002). Thus, these results confirm previous results that have found even within the body of reading curricula that employ explicit, systematic instruction, some appear to be significantly more effective than others.

Our second research question grew out of the implementation literature and asked if changes in reading achievement would be greater in schools that received support from purveyors associated with the developer and, presumably, most knowledgeable and experienced with the program. Again, as suggested by previous literature, the increases in achievement were significantly larger in the NSDI schools (i.e., those supported by a purveyor with ties to the DI developer), than in the ODI schools (i.e., those with other sources of support). The differences were both statistically significant and substantively strong and resemble studies that have found significantly greater achievement in schools and classrooms that implemented DI with the greatest fidelity (Benner, et al., 2010; Carlson & Francis, 2002; Gersten et al., 1986).

Our third research question also grew out of the implementation literature, as well as writings in the school-reform literature regarding time needed for full implementation and stabilization of organizational change. Based on this literature we examined the extent to which the positive effects of DI changed after several years of implementation when the program was more fully institutionalized and stabilized within a school. Again, the results supported the previous literature. While the average first grader had scores well below the national norm before implementation, scores were at or above the national average after full implementation of the new curricula. The highest scores for all groups appeared in the last years of implementation. This finding parallels the results of the meta-analyses conducted by Borman et al. (2003), which found substantially stronger results for programs that had been implemented over longer periods of time.[7]

Taken together, the results indicate the greatest changes in first graders' reading achievement occurred within schools that had three characteristics: a) the use of DI curriculum, b) implemented with technical support from purveyors who were highly familiar with the program, and c) implemented long enough that the reform was presumably stabilized and institutionalized within the school culture.

7 Borman and associates' meta-analysis examined 49 separate studies of Direct Instruction, with 182 observations, and found an average effect size of .21 (95 percent confidence interval of .17 to .25), the strongest of all models included in the analysis. However, the average length of implementation of DI programs in their analysis was less than 3 years. Our results suggest that their estimate of the effect size was, in fact, lower than what would be achieved if the DI programs had been fully stabilized.

Limitations of the Study and Future Research

This study had several limitations, and there are several ways future research could improve upon this work and/or examine further questions based on our results. First, given the limitations of the secondary data set, we had no information on individual sociodemographic characteristics of the students and relied upon proxy, school-level measures. We also had no information on teacher qualities, such as their training or experience, or other school and classroom variables that could be related to achievement. Having data on measures such as these could provide more precise estimates of effects.

Second, we had only limited information on the content of the curricula used in the OC schools. As noted above, the DI schools, as part of a whole-school reform model, used both reading and language curricula. We do not know the extent to which OC schools incorporated language instruction or the extent to which they used either a curricular reform or a whole-school reform model in developing change. Future research should include such data. Future research could also focus specifically on aspects of the development and content of the OC and DI curricula to see if key elements that explain differential results can be identified. Finally, future research should examine the extent to which comprehensive school reform measures, as occurred with the DI schools, impact achievement apart from more focused changes in curricula, as was apparently more typical in the OC schools.

Third, our measures of the "purveyor" of support and time of implementation can best be seen as proxies for more precise indicators of implementation fidelity; and the lack of more detailed indicators regarding the implementations is a clear limitation. Researchers who study implementation have described several criteria that can be used to measure fidelity of implementation, such as whether or not components of a program have been fully delivered, the amount of exposure or time devoted to a program, the extent to which practitioners use prescribed methods and techniques, and practitioners' engagement and involvement in the content of the program (Benner et al., 2010; Dane & Schneider, 1998; Mowbray, Holter, Teague, & Bybee, 2003). Our use of secondary data precluded obtaining any of this information and we had no data on variations in fidelity of implementation within schools or between schools in any of the groups. Examining the ways in which variations in these elements are related to improvements in reading achievement would be an important avenue for future research.

While our results illustrate the relationship of variations in the purveyor of technical assistance and time of implementation to changing achievement, our data tell us little about why these differences appeared. Future research should examine the ways in which different purveyors of support produce different results within schools and, especially, how the most effective techniques might be more widely used. Future research should also examine the ways in which implementations change as they become institutionalized within schools and how these changes make them more effective. An important element of such research should be examining the extent to which these variations are related to the elements of program fidelity such as those mentioned above.

Finally, our results highlight the importance of longitudinal research. Even though there are dozens of studies on the effectiveness of DI, we found only one covered the six year time span the developers have suggested is necessary for full stabilization of reform (Vitale & Joseph, 2008). If such studies are missing within the DI literature, which has been described as the most researched of the various curricula (Borman et al., 2003), it is unlikely such studies are readily available for other programs. This suggests a potentially serious gap in the research literature. Our

results, as well the meta-analysis by Borman et al. (2003), indicate the impact of a curriculum can appear very different when viewed shortly after implementation and when viewed after fully stabilized. It would appear very important for researchers to examine data over as long a time period as possible. When assessing results, those engaging in meta-analyses and other literature reviews should consider the extent to which programs have been fully implemented and stabilized

Implications for Policy and Practice

Our results have possible implications for practitioners and policy makers concerned with school reform. They illustrate the impact that explicit and systematic curricula can have on student achievement. Children in very high-poverty urban environments can have their reading achievement far surpass the national norms when they are exposed to well-designed, explicit, and systematic instruction. However, our results also suggest even when curricula are determined by examination of their content to be "explicit and systematic," they may not produce equivalent results. This finding replicates other comparisons of the two curricula used in this study (e.g., Crowe et al., 2009; O'Brien & Ware, 2002).

At the same time, however, our results illustrate increased achievement with DI is more likely to appear when the program is implemented with support that requires firm adherence to the program's design. This conforms to implementation research in other areas and suggests school systems wishing to produce change would be well served if they procured technical assistance from those who promote firm allegiance to the full range of program requirements (see also Benner et al., 2010).

In addition, the results indicate time and patience are needed before developing final assessments of the utility of reforms. The highest levels of achievement were only observable in the BCPSS after reforms had become stabilized within the schools—at least five years after they were first implemented. If school systems are using achievement data to guide their curricular choices, our results would suggest more accurate decisions would likely be made if given sufficient time to achieve a stabilized intervention.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.

Adams, G. L., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR*. Seattle: Educational Achievement Systems.

Addison, K. L., &Yakimowski, M. E. (2003). *An evaluation of the Direct Instruction program*. Baltimore City Public School System: Division of Research, Evaluation, Assessment, and Accountability.

American Federation of Teachers (AFT). (1998). *Building on the best, learning from what works: Seven promising reading and language arts programs*. Washington, D.C.: Author.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A.G. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Washington, D.C.: National Institute of Education.

Baker, S. K., Kameenui, E. J., Simmons, D. C., & Stahl, S. A. (1994). Beginning reading: Educational tools for diverse learners. *School Psychology Review, 23*, 372-391.

Beck, I. L., & McCaslin, E. S. (1978). *An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs*. LRDC Report No. 1978/6 Pittsburgh, PA.

Benner, G. J., Nelson, J. R., Stage, S. A., & Ralston, N. C. (2010). The influence of fidelity of implementation on the reading outcomes of middle school students experiencing reading difficulties. *Remedial and Special Education*, *32*, 79-88. doi: 10.1177/0741932510361265

Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). *Facing the challenges of whole-school reform: New American schools after a decade*. Santa Monica, CA: Rand.

Berkeley, M. (2002). The importance and difficulty of disciplined adherence to the educational reform model. *Journal of Education for Students Placed at Risk, 7*, 221-239.

Blakeley, M. R. (2001). A survey of levels of supervisory support and maintenance of effects reported

by educators involved in Direct Instruction implementations. *Journal of Direct Instruction*, *1*, 3-83.

Bodilly, S. J., Glennan, T. K., Jr., Galegher, J. R., & Kerr, K. A. (2004). Introduction: Framing the problem. In T. K. Glennan, Jr., S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 1-35). Santa Monica, CA: Rand.

Bodilly, S. J. (1998). *Lessons from new American schools' scale-up phase: Prospects for bringing designs to multiple schools*. Santa Monica, CA: Rand.

Bond, G., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2*, 5-142.

Borman, G. D., Dowling, N. M., & Schneck, C. (2008). A multisite cluster randomized field trial of Open Court Reading. *Educational Evaluation and Policy Analysis, 30*, 389-407.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research 73*, 125-230.

Borman, G. D. (2007). National efforts to bring reform to scale in high-poverty schools: Outcomes and implications. In B. Schneider & S-K. McDonald (Eds.) *Scale-up in education, Volume II: Issues in practice* (pp. 41-67). Lanham: Rowman & Littlefield.

Carlson, C. D., & Francis, D. J. (2002). Increasing the reading achievement of at-risk children through Direct Instruction: Evaluation of the Rodeo Institute for Teacher Excellence (RITE). *Journal of Education for Students Placed at Risk*, 7, 141-166.

Center for the Future of Teaching and Learning. (CFTL). (1996). Thirty years of NICHD research: What we now know about how children learn to read. *Effective School Practices, 15*, 33-46.

Chall, J. S. (1967). *Learning to read: The great debate*. New York, NY: McGraw Hill.

Cizek, G. J. (1998). *Filling in the blanks: Putting standardized tests to the test*. Washington, D.C.: Thomas B. Fordham Foundation.

Collins, M., & Carnine, D. (1988). Evaluating the field test revision process by comparing two versions of a reasoning skills CAI program. *Journal of Learning Disabilities, 21*, 375-379.

Crowe, E. C., Connor, C. M., & Petsher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. *Journal of School Psychology 47*, 187-214.

CTB/McGraw-Hill. (2001). *TerraNova technical manual report*. Monterey, CA: Author.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45.

Edsource. (2006). *Elementary school curriculum program and API*. Mountain View, CA: Author.

Engelmann, S. (2004). The dalmatian and its spots. *Education Week 23*, (20), 34-35, 48.

Engelmann, S. (2007). *Teaching needy kids in our backward system: 42 Years of trying*. Eugene,OR: ADI Press.

Engelmann, S. (n.d.). *Developer's guidelines: Achieving a full-school, full-immersion implementation of Direct Instruction.* Eugene, OR: National Institute for Direct Instruction. Retrieved from NIFDI website: http://www.nifdi.org/pdfs/Dev_Guide.pdf

Engelmann, S. E., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. New York, NY: Irvington Publishers.

Engelmann, S. E., & Engelmann, K. E. (2004). Impediments to scaling up effective comprehensive school reform models. In T. K. Glennan, Jr., S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 107-133). Santa Monica, CA: Rand.

Farkas, G. (1996). *Human capital or cultural capital: Ethnicity and poverty groups in an urban school district*. New York, NY: Aldine.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Florida Center for Reading Research (FCRR). (2004). *Open Court Reading Pre-K.* Florida Center for Reading Research. Retrieved from FCRR website: http://www.fcrr.org/FCRRReports/PDF/OpenCourtReadingFinal2.pdf Foorman, B. R. (1995). Research on "the great debate": Code-oriented versus whole language approaches to reading instruction. *School Psychology Review, 24*, 376-392.

Foorman, B. R., Francis, D. J., Fletcher, J. M., 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.

Fukkink, R. G., & deGlopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of Educational Research, 68*, 450-469.

Gersten, R. M., Carnine, D. W., & Williams, P. B. (1982). Measuring implementation of a structured educational model in an urban school district: An observational approach. *Educational Evaluation and Policy Analysis 4*, 67-79.

Gersten, R., Carnine, D., Zoref, L., & Cronin, D. (1986). A multifaceted study of change in seven inner-city schools. *The Elementary School Journal 86*, 257-276.

Grossen, B. (1997). *A synthesis of research on reading from the National Institute of Child Health and Human Development*. Eugene,OR: University of Oregon.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*, 1900-1902.

Hempenstall, K. (1999). The role of phonics in learning to read: What does recent research say? *Fine Print 22* (1), 7-12.

Hempenstall, K. (1997). The whole language-phonics controversy: An historical perspective. *Educational Psychology: An International Journal of Experimental Educational Psychology, 17*, 399-418.

Herman, R., Aladjam, D., McMahon, P., Masem, E., Mulligan, I., Smith, O., . . . Woodruff, D. (1999). *An educator's guide to schoolwide reform*. Washington, D.C.: American Institutes for Research.

Huitt, W. G., Monetti, D. M., & Hummel, J. H. (2009). Direct approach to instruction. In C. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models: Volume III, Building a common knowledge base* (pp. 73-98). Mahwah, NJ: Lawrence Erlbaum.

Jencks, C., Crouse, J., & Mueser, P. (1983). The Wisconsin model of status attainment: A national replication with improved measures of ability and aspiration. *Sociology of Education 56*, 3-19.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly, 35*, 458-492.

MacIver, M. A., & Kemper, E. (2002). The impact of direct instruction on elementary students' reading achievement in an urban school district. *Journal of Education for Students Placed at Risk, 7*, 197-220.

McRae, D. J. (2002). Test score gains for Open Court schools in California: Results from three cohorts of schools. Retrieved from http://www.imagineitreading.com/TX/documents/Docs/McRae%20Study.pdf.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and evaluation. *American Journal of Evaluation, 24*, 315-340.

Murphy, J. (2004). *Leadership for literacy: Research-based practice, preK-3*. Thousand Oaks, CA: Corwin Press.

National Institute of Child Health and Human Development (NICHD). (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* (NIH Publication No. 00-4769). Washington, D. C.: U. S. Government Printing Office.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, D.C.: National Institute of Child Health and Human Development.

O'Brien, D. M., & Ware, A. M. (2002). Implementing research-based reading programs in the Fort Worth Independent school district. *Journal of Education for Students Placed at Risk, 7*, 167-195.

Oregon Reading First Center (ORFC). (2004). *Review of Open Court*. Eugene,OR: Oregon Reading First. Retrieved from: http://oregonreadingfirst.uoregon.edu/downloads/instruction/curriculum_review/core_review/oc2002_core_review.pdf

Pflaum, S. W., Walberg, H. J., Karigianes, M. L., & Rasher, S. P. (1980). Reading instruction: A quantitative analysis. *Educational Researcher, 9*, 12-18.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.

Ross, S. M., Nummery, J. A., Goldfeder, E., McDonald, A., Racho, R., Hornbeck, M., & Fleishman, S. (2004). Using school reform models to improve reading achievement: A longitudinal study of Direct Instruction and Success for All in an urban district. *Journal of Education for Students Placed at Risk, 9*, 357-388.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics 23*, 323-355.

Skindrud, K., & Gersten, R. (2006). An evaluation of two contrasting approaches for improving reading achievement in a large urban district. *The Elementary School Journal, 106*, 389-408.

Smith, S., Simmons, D., Gleason, M., Kameenui, E., Baker, S., Sprick, M.… Thomas, C. (2001). An analysis of phonological awareness instruction in four kindergarten basal reading programs. *Reading and Writing Quarterly, 17*, 25-51.

Snider, V. E. (1990). Direct Instruction reading with average first graders. *Reading Improvement, 27*, 143-148.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, D. C.: National Academy Press.

Stanovich, K. E. (1994). Romance and reality. *The Reading Teacher, 47*, 280-291.

Stringfield, S. C., & Land, D. (Eds.). (2002). *Educating at-risk students*. Chicago, IL: National Society for the Study of Education.

Stringfield, S. C., & Yakimowski-Srebnick. M. E. (2005). Promise, progress, problems, and paradoxes of three phases of accountability: A longitudinal case study of the Baltimore city public schools. *American Educational Research Journal 42*, 43-75.

Vitale, M. R., & Joseph, B. L. (2008). Broadening the institutional value of Direct Instruction implemented in a low-SES elementary school: Implications for scale-up and reform. *Journal of Direct Instruction, 8*, 1-18.

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.* Rockville, MD: Author.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Newbury Park, CA: Sage.