JEAN STOCKARD, National Institute for Direct Instruction and University of Oregon

# *Improving Elementary-Level Mathematics Achievement in a Large Urban District: The Effects of Direct Instruction*

**Abstract: This paper examines changes in the average mathematics achievement of students in the Baltimore City Public School System (BCPSS) from 1998 to 2003, comparing students in schools that implemented Direct Instruction with students in other schools. First-grade students who received Direct Instruction had significantly higher levels of achievement on the Comprehensive Test of Basic Skills (CTBS) subtests of mathematics computations and mathematics concepts and applications. Differences were greater when measures of schools' sociodemographic context were controlled and as Direct Instruction became fully implemented and incorporated within the schools. Among students who began first grade in the BCPSS and remained in the same schools five years later as fifth graders, those who had received Direct Instruction as first graders had significantly higher scores on the measure of mathematics concepts and applications than students in the other schools.**

International comparisons of mathematics achievement have consistently shown that students in the United States have lower scores than students in other countries. These differences appear in comparisons with countries of similar levels of economic development and, often, with countries of even lower levels (e.g., Mullis, Martin, & Foy, 2005; National Center for Education Statistics, 2008, Tables 389-390). Low levels of achievement are particularly problematic for students from families with fewer socioeconomic resources (e.g., National Center for Education Statistics, 2008, Table 125). Mathematical skills are often key requirements of employment in the growing and more highly paid, technologically sophisticated sectors of the economy (Spitz-Oener, 2006). Thus, improving mathematics achievement is an important element of enhancing social equity and well being for future generations and the country as a whole and a concern of school districts throughout the nation.

Two general approaches to mathematics instruction are often distinguished. One is the constructivist approach, which emphasizes learning as developing from students' understandings of the world and their interactions with each other and the world around them. It emphasizes student inquiry and discovery, inductive reasoning, and a spiral-based curriculum in which students revisit topics each year with growing depth. The second approach is direct or explicit instruction, which emphasizes clear objectives and direct teaching of key concepts and strategies. In contrast to the spiral-based curriculum, the explicit approach is organized around strands that use a relatively few "big ideas" or important topics as the curricular focus (Applefield, Huber, & Moallem, 2000/2001; Carnine, 1990; Davis, 1990; Przychodzin, Marchand-Martella,

---

*Journal of Direct Instruction, Vol. 10, pp. 1-16.* Address correspondence to Jean Stockard at jstockard@nifdi.org.

Martella, & Azim, 2004; Stein, Silbert, & Carnine, 1997; Woodward, 2004). Despite the philosophical disputes over the approaches, empirical data consistently demonstrate that children learn more rapidly and more efficiently when instruction is systemic, explicit, unambiguous, and well designed, and when it includes elements such as careful monitoring of progress and teaching with clear objectives (Engelmann & Carnine, 1982). These are key elements in the explicit or more direct approach (Darch, Carnine, & Gersten, 1984; Troff, 2004).

One of the most prominent explicit instructional approaches is Direct Instruction (DI) (distinguished from other "direct instruction" approaches by the use of capital letters), which was developed by Siegfried Engelmann and his colleagues (Engelmann, 2007; Engelmann & Carnine, 1982). Unlike many curricula, the DI programs are extensively field tested before dissemination with the hope that they will produce the greatest learning in the most efficient manner. Numerous studies have documented the superiority of DI in promoting mathematics achievement, both with the general population and with students receiving special education (e.g., Baker, Gersten, & Lee, 2002; Moore & Carnine, 1989; Przychodzin et al., 2004). In addition, DI has been hailed as an effective model of comprehensive school reform and one of the premier ways in which failing schools can be turned around (see Adams & Engelmann, 1996; Baker et al., 2002; Borman, Hewes, Overman, & Brown, 2003; Comprehensive School Reform Quality Center, 2006; Kidron & Darwin, 2007; Woodward, Baxter, & Robinson, 1999 for reviews).

While the literature base regarding DI appears to be larger than that dealing with other curricula (Borman et al., 2003), many individual studies have involved relatively few students and classrooms. Thus, studies that include more students, classrooms, and schools could provide an important replication of earlier, smaller studies of the DI mathematics curriculum. In addition, many of the earlier studies have involved data from a relatively short span of time. Examination of data from an extended time period is important for developing greater understandings of how the process of comprehensive school reform is related to changing achievement. The developers of DI have stressed that fully implementing and stabilizing the program within a school takes an extended period of time, "as much as five or six years" (Engelmann & Engelmann, 2004, p. 113). Developers of other programs also have documented and discussed the difficulties involved in large-scale reforms. See Glennan, Bodilly, Galegher, and Kerr (2004) for discussions of these issues in a wide variety of reform efforts including, among others, Cognitively Guided Instruction (CGI), the National Writing Project, and High Schools that Work.

This paper addresses these gaps. It analyzes data from a large number of students in one large urban public school system over six years, the time period that is suggested as needed for full implementation and stabilization of reforms (Engelmann & Engelmann, 2004). Some of the students in the sample were in schools that implemented DI as part of a model of whole-school reform. The other students were in schools that adopted the system's traditional programs. We first examine changes in the mathematics achievement of first graders, comparing the achievement of students in the DI schools with the achievement of first graders in other schools in the district, examining how these differences changed over time and the extent to which they remained when demographic characteristics of the schools were introduced. We then look at the subset of students who were in the same schools in both first and fifth grade and examine the extent to which experiencing DI in first grade is associated with higher mathematics achievement in fifth grade.

# *Methods*

The data for this analysis come from the Baltimore City Public School System (BCPSS), which is similar to many other large city school districts that serve students with high levels of poverty and struggle with low levels of achievement. In the late 1990s, curricular reforms were implemented in the BCPSS elementary schools to address this low achievement. The sections below describe the procedures, sample, measures, and analysis techniques that were used.

## Procedures

Beginning in 1996, eleven BCPSS schools implemented DI as part of a whole-school reform effort (Berkeley, 2002; Bowden, Greene, & Campbell, n.d.). One additional school was part of the original DI intervention group but was closed shortly after the start of the study period. Because data are not available throughout the time span of the study, data from that school are not included. In addition, a few other schools in the system used DI reading programs. They, however, did not use DI mathematics curricula.

All schools that used DI began their implementation with the reading programs and later added DI instruction in mathematics. Three schools implemented Mathematics DI in 1997-98, three in 1998-99, four in 1999-2000, and one in 2000-01. In all but two schools the DI mathematics curriculum was added one year after beginning reading instruction.

Their implementations were guided throughout the study period by technical assistance from the National Institute for Direct Instruction (NIFDI), a not-for-profit corporation dedicated to providing school districts with training and implementation support. NIFDI provided preservice training as well as inservice training and in-class coaching and consultation throughout the study period. These frequent consultations with, and support to, teachers and other school staff helped ensure that the curriculum was implemented with high fidelity. Two DI mathematics curricula were used depending upon the students' level and needs: *DISTAR Arithmetic* and *Connecting Math Concepts*.

Beginning in 1999 the other schools in the district also began the systematic use of mathematics curricular material, adopting the McGraw-Hill *Math in My World* and Scott Foresman Addison Wesley *Investigations in Number, Data and Space* programs (Bowden et al., n.d.). The *Math in My World* program is best characterized as a standard basal textbook, while the *Investigations* program emphasizes numerous aspects of the constructivist approach. There was no way to differentiate the two alternative curricula that students could receive. However, both are substantially different from the explicit DI curricula.

The schools that implemented DI are referred to below as the DI schools, and the remaining schools are termed the control schools. In the analysis, students were designated as belonging to the DI condition only during the years in which their school was receiving the DI mathematics implementation.

## Sample

The sample for the analysis of first-grade achievement included approximately 45,000 first-grade students from 1997-98 through 2002-2003. Enrollment in the BCPSS declined over this time period, with data available for approximately 9,000 first graders for the first year and 6,400 for the last year. On average, data were available for approximately 380 students per school. Only four schools, all in the control group, had fewer than 100 students in the sample.

Because there was substantial student turnover within the BCPSS, the sample used to examine the research questions regarding the impact of DI math instruction in first grade on fifth-grade math achievement was much smaller. The potential pool included only students who were in the BCPSS in 1997-98 or 1998-99 and thus could still be in the system in fifth grade in 2001-02 or 2002-03, the last years for which data were available (students who were held back were not included because of a lack of data). Only 27% of the students who were in first grade in the system in 1997-98 and 1998-99 were still in the same school in fifth grade in 2001-02 or 2002-03 (164 who had received DI in first grade and 4,607 who had not).

The students in the DI group had significantly lower first-grade achievement scores than the students in the control group and thus two sets of comparisons were conducted. First, the achievement of the 164 DI students was contrasted with that of students in all the control schools (n = 4,607). Second, schools within the control group were rank ordered by their composite mathematics achievement score and a subsample of schools was selected (beginning with the lowest) that produced an average as close as possible to that of students in the DI schools. The resulting alternative control group included 1,820 students in 50 schools. The minimum number of cases from a school was 16 and the maximum was 101.

## Instrumentation

The Comprehensive Test of Basic Skills (CTBS), a widely used standardized achievement test, was administered to all students in the spring of each year, from 1997-98 through 2002-03 as part of a system-wide testing program, and the data for first- and fifth-grade students were provided to the author. The 4th edition was administered in the spring of 1998 and 1999, and the 5th edition was administered in the spring of 2000 through 2003. Two subtest scores, Mathematics Computations and Mathematics Concepts and Applications, were analyzed. Normal Curve Equivalent (NCE) scores were used for all statistical analyses to help ensure comparability from one year to another and one grade 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.

The demographic characteristics of the schools were measured by two composite (factor) scores based on a factor analysis of 6 school characteristics: the proportion of students receiving free and reduced-price lunch and the proportion of students who were African American, Asian American, Hispanic, Native American, and non-Hispanic whites. The first factor accounted for 44% of the total variance between schools, had a strong positive loading of the proportion of non-Hispanic white students, a slightly smaller positive loading for the proportion of Asian students, and strong negative loadings for the proportion of African American students and the proportion receiving free and reduced-price lunch. Thus, schools with positive scores on this factor had proportionately more non-Hispanic white and Asian students and many fewer African American and poor students. The second factor, which accounted for 29% of the total variance, had positive loadings on the proportion of Hispanics and the proportion of Native Americans in a school, and a negative loading, of a somewhat smaller magnitude, of the proportion of African Americans. Thus, schools with a positive score on this factor had proportionately more Hispanic and Native American students and somewhat fewer African American students. There were no significant differences between the groups of schools on either factor score.

## Analysis

Changes in first-grade mathematics achievement were examined in two ways. First, the average scores of the students in the DI and the control groups were compared over all years combined as well as in each year. Two-way analysis of variance (ANOVA) (year and group as factors), t-tests for comparisons within each year, and effect sizes (Cohen's d) were computed. McLean, O'Neal, and Barnette (2000) caution that effect sizes calculated with NCE scores are inherently smaller than those with other metrics. Thus, the effect sizes presented might be a conservative estimate.

Second, mixed model regressions were conducted including group (DI versus Control), the year in which data were gathered, and the factor scores of school characteristics as predictors. To the extent that DI is more effective at promoting student achievement, it would be expected that students in the DI schools would have higher achievement scores than those in the control schools, even when school characteristics were controlled. In addition, to the extent that the positive impact of DI becomes more apparent as reforms become stabilized within a school, it would be expected that the differences between the scores in the DI and control schools would be larger in later years.

Mixed models are particularly appropriate for analyzing multi-level data, such as data regarding students and the schools that 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 equivalent to having a separate intercept in the regression equation for each school. The coefficients associated with the various individual and school-related 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 and the extent to which various independent variables can account for this between-school variance (Raudenbush & Bryk, 2002; Singer, 1998).

The impact on the achievement of fifth graders who received DI in first grade also was examined in two ways. First, descriptive statistics and effect sizes were computed, as well as repeated measures ANOVA (first- and fifth-grade scores as repeated measures and group as a factor). Because the issue of interest was the change in scores from one year to another, the standard method of calculating effect sizes (dividing the difference between two means by the common standard deviation) is not appropriate. Instead, the correction suggested by Dunlap, Cortina, Vaslow, and Burke (1996), which includes an adjustment for the correlation between the two scores, was used. Results may be interpreted in the same manner as Cohen's d. Second, mixed model regressions were used, having schools as a random variable and regressing students' fifth-grade scores on first-grade scores, the measures of the school context of advantage/disadvantage, and school type (DI or Control). For all analyses, two sets of comparisons were used: one with students in all the control schools and one with students in the schools with more similar levels of achievement in first grade, as described above. The latter group is referred to as the "reduced sample" of control schools.

# *Results*

## First-Grade Achievement

Descriptive results regarding the first-grade mathematics achievement of students in the DI and control schools are summarized in Table 1. NCE scores are in Panel A and the corresponding percentiles are in Panel B. Also included are the results of a two-way ANOVA (year and group as factors), t-tests for results for each year, and effect sizes comparing the two groups.

Results are similar for both measures. The ANOVA results indicated a highly significant interaction between year and group for both computations and concepts and applications scores. That is, the nature of the differences between the achievement scores of students in the two groups varied from one year to another.

The average scores for each group and the t-values that compare these scores illustrate the pattern of change. In the first one to two years of implementation, the average mathematics achievement scores of students in the DI schools were significantly lower than those of the students in the control schools, reflecting the very low achievement that characterized the DI schools (see Berkeley, 2002). By the last two to three years, the situation had changed markedly, with the DI students scoring significantly higher.

The results in Panel B of Table 1 illustrate the magnitude of these changes. The percentiles correspond to the average normal curve equivalent scores in Panel A and represent the score that an average student in each type of school would have in each year. In 1998, an average first grader in a DI school scored at the 12th percentile in computations and the 16th percentile in concepts and applications. By 2003, however, an average first grader in a DI school scored at the 66th percentile in computations and the 60th percentile in concepts and applications. In contrast, an average student in a control school scored at the 21st percentile in computations and the 29th percentile in concepts and applications in 1998 and at the 56th percentile in computations and the 46th percentile in concepts and applications in 2003.

Another way to examine the change is to look at the percentage increase in average NCE scores from 1998 to 2003. The increases for the DI schools were much larger than those for the control schools: 134% versus 60% for the measure of computations, and 87% versus 26% for the measure of concepts and applications.

Finally, because the sample size is large, it is important to examine effect sizes in addition to tests of statistical significance. Table 1 includes the effect sizes, measured by Cohen's d, for each year and each measure. An often-used criterion is the value of 0.25 to designate an effect that can be seen as "educationally significant" (Fashola & Slavin, 1997). When data for all years are combined, the impact of being in a DI school, while positive, does not meet the criterion of educational significance. However, when data from the last years of implementation are examined, the effect sizes meet this level, especially with the measure of concepts and applications.

Table 2 provides the results of the mixed model analyses. Three models were tested. Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This model tests the null hypothesis that the schools are equal in average mathematics achievement, and the associated statistics are reported in the footnote to Table 2. The correlation ratio is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that from 8% to 9% of the variance in achievement is between schools rather than simply between students. The estimates, z-values, and probabilities associated with the random effects test the null hypothesis that the variation between schools equals zero once variables in a 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 and school differences are controlled. These null hypotheses can be easily rejected with all the models. There is significant variation between schools and also between students in all the models. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those available in this analysis.

The second model that was tested (and the first with coefficients in Table 2) adds the year in which data were collected, the two factor scores, and the grouping variable (a dummy variable for DI versus control schools) to the

**Table 1**

*Mathematics Achievement by Year and Group, First Graders, BCPSS, 1998-2003*

*Panel A: Means (NCE Scores), t-test Results, and Cohen's d*

| | DI | Control | *t* value | Df | prob. | Cohen's *d* |
|---|---|---|---|---|---|---|
| **Computations** | | | | | | |
| All Years Combined | 49 | 43 | 9.73 | 2716 | <0.001 | 0.21 |
| 1998 | 25 | 33 | -4.76 | 176 | <0.001 | -0.33 |
| 1999 | 30 | 32 | -1.27 | 343 | 0.21 | -0.07 |
| 2000 | 45 | 46 | -0.56 | 7700 | 0.58 | -0.03 |
| 2001 | 53 | 50 | 2.26 | 608 | 0.02 | 0.11 |
| 2002 | 57 | 52 | 4.57 | 570 | <0.001 | 0.23 |
| 2003 | 59 | 53 | 5.01 | 530 | <0.001 | 0.25 |
| Percent Increase (1998-2003) | 134 | 60 | | | | |

Analysis of Variance Results: F (year) = 295.3, df = 5, 45275, p <0.001; F (group) = 0.73, df = 1,45275, p = 0.39; F (interaction) = 12.61, df = 5,45287, p <0.001

| | DI | Control | *t* value | Df | prob. | Cohen's *d* |
|---|---|---|---|---|---|---|
| **Concepts and Applications** | | | | | | |
| All Years Combined | 45 | 42 | 6.40 | 45533 | <0.001 | 0.13 |
| 1998 | 29 | 38 | -5.64 | 176 | <0.001 | -0.38 |
| 1999 | 32 | 37 | -4.42 | 342 | <0.001 | -0.22 |
| 2000 | 40 | 41 | -0.1 | 7851 | 0.92 | 0.00 |
| 2001 | 46 | 45 | 0.91 | 7352 | 0.36 | 0.04 |
| 2002 | 53 | 47 | 5.97 | 6805 | <0.001 | 0.28 |
| 2003 | 55 | 48 | 5.98 | 518 | <0.001 | 0.32 |
| Percent Increase (1998-2003) | 87 | 26 | | | | |

Analysis of Variance Results: F (year) = 153.9, df = 5, 45523, p <0.001; F (group) = 0.023, df = 1, 45523, p = 0.88; F (interaction) = 22.34, df = 5,45523, p <0.001

*Panel B: Percentile Equivalents of Norm Equivalent Scores, by Year*

| | Computations | | Concepts and Applications | |
|---|---|---|---|---|
| | DI | Control | DI | Control |
| 1998 | 12 | 21 | 16 | 29 |
| 1999 | 17 | 20 | 20 | 27 |
| 2000 | 41 | 42 | 32 | 33 |
| 2001 | 56 | 50 | 42 | 41 |
| 2002 | 63 | 54 | 56 | 44 |
| 2003 | 66 | 56 | 60 | 46 |

*Note.* In Panel A, the smaller degrees of freedom for some years reflect cases where F tests revealed that the variances of the two groups were unequal and the alternative formula for calculating t and degrees of freedom was used.

## Table 2

*Mixed Model Regressions of First-Grade Mathematics Achievement on Year, School Context, and Group, BCPSS, 1998-2003*

| | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|
| *Panel A: Computations* | *b* | *SE* | *p* | *b* | *SE* | *p* |
| Intercept | -9569 | 129 | <0.0001 | -9442 | 132 | <0.0001 |
| Year | 4.8 | 0.1 | <0.0001 | 4.7 | 0.1 | <0.0001 |
| Factor 1 | 1.6 | 0.6 | 0.005 | 1.7 | 0.6 | 0.004 |
| Factor 2 | -1.7 | 0.6 | 0.004 | -1.5 | 0.6 | 0.009 |
| DI | 7.9 | 1.0 | <0.0001 | -3091 | 653 | <0.0001 |
| DI* Year | — | — | — | 1.5 | 0.3 | <0.0001 |
| | *estimate* | *SD* | *Sig.* | *estimate* | *SD* | *Sig.* |
| Random Effects Estimate | 38.0 | 5.2 | <0.0001 | 37.9 | 5.2 | <0.0001 |
| Residual Estimate | 514.5 | 3.4 | <0.0001 | 514.3 | 3.4 | <0.0001 |
| BIC | 411679 | | | 411661 | | |
| - 2 Log Likelihood | 411646 | | | 411623 | | |
| | *value* | *df* | *Sig.* | *value* | *df* | *Sig.* |
| Change in Log Likelihood | 5660 | 1 | <0.0001 | 23 | 1 | <0.0001 |
| *Panel B: Concepts and Applications* | *b* | *SE* | *p* | *b* | *SE* | *p* |
| Intercept | -4769 | 118 | <0.0001 | -4533 | 121 | <0.0001 |
| Year | 2.4 | 0.1 | <0.0001 | 2.3 | 0.1 | <0.0001 |
| Factor 1 | 2.7 | 0.6 | <0.0001 | 2.7 | 0.6 | <0.0001 |
| Factor 2 | -1.7 | 0.6 | 0.003 | -1.4 | 0.6 | 0.01 |
| DI | 5.7 | 0.9 | <0.0001 | -5824 | 600 | <0.0001 |
| DI * Year | — | — | — | 2.9 | 0.3 | <0.0001 |
| | *estimate* | *SD* | *Sig.* | *estimate* | *SD* | *Sig.* |
| Random Effects Estimate | 34.5 | 4.8 | <0.0001 | 34.6 | 4.8 | <0.0001 |
| Residual Estimate | 434.1 | 2.9 | <0.0001 | 433.2 | 2.9 | <0.0001 |
| BIC | 406203 | | | 406113 | | |
| - 2 Log Likelihood | 406169 | | | 406075 | | |
| | *value* | *df* | *Sig.* | *value* | *df* | *Sig.* |
| Change in Log Likelihood | 1831 | 1 | <0.0001 | 94 | 1 | <0.0001 |

*Note.* Results for the Random Effects Model were, for computations: random effects estimate = 47.5, *SE* = 6.5, *Sig.* < 0.0001; residual estimate = 582.9, *SE* = 3.9, *Sig.* <0.0001; BIC = 417320, -2 Log Likelihood = 417306, correlation ratio = 0.08. Results for concepts and applications were random effects estimate = 46.4, *SE* = 6.4, *Sig.* <0.0001; residual estimate = 451.6, *SE* = 3.0, *Sig.* <0.0001, BIC = 408014, -2 Log Likelihood = 408000; correlation ratio = 0.09. In all cases t-ratios can be computed by dividing the coefficient (random or residual estimate or the b values) by the standard error. Change in log likelihood is calculated by subtracting the value in the more complex value from the value in the less complex value.

baseline model. Model 3 adds the interaction of being in a DI school and year, testing the hypothesis that the impact of being in a DI school increases over time. The -2 log likelihood measures and the BIC values can be used to examine the relative fit of the data to the models. Lower values indicate a better fit. Differences between the log likelihood measures have a chi-square distribution, and the comparisons between these values are in the final rows of each part of Table 2. The BIC values provide a descriptive summary of the fit of the models, with lower values indicating a better fit. Both the BIC values and the changes in the -2 log likelihoods indicate that Model 3, which includes the interaction of year and group, provides the best fit to the data, reinforcing the conclusion obtained with the more descriptive analysis in Table 1.

The coefficients associated with the control variables indicate that scores increased significantly over time (positive coefficient associated with year), that students in schools with fewer African American and low-income students had significantly higher scores (positive coefficient associated with Factor 1), and that students in schools with more Hispanic and Native American students had significantly lower scores (negative coefficient associated with Factor 2). Yet, even when these variables were controlled, the impact of being in a DI school was highly significant.

The differences in coefficients associated with DI from Model 2 to Model 3 reflect the strong changes over the years as differences between the two groups became stronger. Table 3 summarizes these differences by giving the expected values of the two dependent variables for each group calculated from the equations in Model 3 for 1998, the first year, and for 2003, the last year of data. Panel A gives the NCE scores, which were the basis of the mixed model analyses and the calculations. Panel B translates these scores into the corresponding percentiles. The percentiles may be interpreted as the score that an average student in a DI or control school would receive, in either 1998 or 2003, if the schools had

## Table 3

*Simulated Expected Values, First Grade, Mathematics Achievement, Controlling for Factor Scores, 1998 and 2003*

| | 1998 | | 2003 | |
|---|---|---|---|---|
| *Panel A: NCE scores* | DI | Control | DI | Control |
| Computations | 35.0 | 32.0 | 66.5 | 55.7 |
| Concepts and Applications | 33.2 | 36.8 | 59.2 | 48.2 |
| *Panel B: Percentiles* | | | | |
| Computations | 23.8 | 19.6 | 78.3 | 60.7 |
| Concepts and Applications | 21.3 | 26.5 | 66.8 | 46.6 |

*Note.* The following equations were used to calculate these values: For computations, Direct Instruction Achievement = -12,533.22 + 6.2904*year; Control Schools Achievement = -9442.34 + 4.7419*year, where year = 1998 or 2003. For concepts and applications, Direct Instruction Achievement = -10356.79 + 5.2002*year; Control Schools Achievement = -4532.670 + 2.287*year. Factor scores are assumed to equal zero (the average when schools are the unit of analysis).

equivalent social-demographic characteristics as measured by the two factor scores. It can be seen that in 1998, the average student in a DI school would have a computations score that was slightly higher than those of average students in the control schools (3 NCE points) and a concepts and applications score that was slightly lower (4 NCE points). However, by 2003, the average DI student had scores on both measures that were substantially higher than those of students in the control school, a difference of 10 to 11 NCE points.

## Fifth-Grade Achievement

Tables 4 and 5 summarize the analysis of fifth-grade data with both the full set of control schools and the reduced set of schools, which included only those with first-grade scores closer to those of students in the DI schools. Table 4 gives the descriptive statistics (Panel A) and results of repeated measures ANOVA (Panel B), with mathematics scores as the repeated measure and group as the factor. If the changes in students' achievement from first to fifth grade were greater for the students in the DI schools than for those in the control schools, the interaction effect in the ANOVA would be significant. Even though the results are in the expected direction with the measure of computations, the interaction effects miss traditional levels of significance ($p = 0.06$ for the comparison with the full set of control schools and $p = 0.11$ for the comparison with the reduced set of control schools). In contrast, the interaction effect is significant with the analysis of concepts and applications ($p < 0.001$) for comparison with the full set of control schools and for the comparison with the reduced set ($p = 0.002$). On average, the concepts and applications scores of students in the DI schools increased by 30% from first to fifth grade, but those of students in the full set of control schools increased by only 3%, and those of students in the reduced set of control schools increased by 16%. In other words, the average increase from first grade to fifth grade on the measure of concepts and applications was at least twice as high for students in the DI schools as for students in the control schools.

Panel C of Table 4 gives the percentile values that correspond to the average NCE scores. In first grade the average student in a DI school scored at the 26th percentile on the measure of concepts and applications, substantially below the scores of the average student in the control schools (42nd percentile for the full group and 30th percentile for the reduced set). By fifth grade the average DI student (by that time at the 45th percentile) had caught up with the average student in the full set of control schools (46th percentile) and outpaced the average student in the reduced set of control schools (41st percentile).

Table 5 provides the results of the mixed model analysis with fifth-grade scores as the dependent variable, schools as a random effect, and first-grade scores and the two measures of school context as independent variables. This provides additional controls for the analysis by comparing students' achievement after equalization on the measures of school context and first-grade achievement. The results confirm the findings summarized in Table 4. Having DI in first grade does not have a significant impact on students' fifth-grade computation scores, once first-grade scores and school context are equalized. However, having DI in first grade has a significant impact on the fifth-grade measure of concepts and applications. In comparisons with both the full set and reduced set of control schools, students in the DI schools have higher concepts and applications scores, even when school context and first-grade achievement are equalized. The magnitude of the coefficient (about 7 NCE points) is similar for both comparisons.

Table 4 provides effect sizes that summarize the magnitude of the effects of changes from first grade to fifth grade. The results provide further confirmation of the importance of DI

**Table 4**

*Descriptive Statistics, NCE Scores and Associated Percentiles, Computations and Concepts and Applications Scores, Effect Sizes, and Repeated Measures of Analysis of Variance, First- and Fifth-Grade Panel Group, DI and Control Schools, Full and Reduced Samples, BCPSS, 1998-2003*

*Panel A: Descriptive Statistics (Norm Equivalent Scores)*

| | Grade | DI Schools Full Sample | | Control Schools Full Sample | | Control Schools Reduced Sample | |
|---|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Computations | 1st. | 33.1 | 22.1 | 40.3 | 25.1 | 30.8 | 21.7 |
| | 5th | 50.2 | 19.5 | 53.4 | 22.1 | 51.2 | 22.0 |
| | % ch. | 0.52 | | 0.32 | | 0.66 | |
| Effect Size of Change | | 0.82 | | 0.55 | | 0.93 | |
| Concepts and Applications | 1st. | 36.2 | 19.0 | 46.2 | 22.2 | 38.7 | 20.3 |
| | 5th | 47.1 | 16.2 | 47.5 | 17.4 | 44.8 | 16.9 |
| | % ch. | 0.30 | | 0.03 | | 0.16 | |
| Effect Size of Change | | 0.61 | | 0.07 | | 0.32 | |

*Panel B: Repeated Measures Analysis of Variance Results*

| | Mathematics | | Group | | Interaction | |
|---|---|---|---|---|---|---|
| DI Schools vs. Full Set of Control Schools | *F* | *Sig.* | *F* | *Sig.* | *F* | *Sig.* |
| Computations | 203.16 | <0.001 | 10.76 | 0.001 | 3.59 | 0.06 |
| Concepts and Applications | 60.02 | <0.001 | 13.94 | <0.001 | 36.56 | <0.001 |
| DI Schools vs. Reduced Set of Control Schools | *F* | *Sig.* | *F* | *Sig.* | *F* | *Sig.* |
| Computations | 329.69 | <0.001 | 0.20 | 0.66 | 2.59 | 0.11 |
| Concepts and Applications | 126.76 | <0.001 | 0.004 | 0.95 | 10.03 | 0.002 |

*Panel C: Percentile Score of Average Student by School Type, First and Fifth Grade*

| | Grade | DI Schools Full Sample | Control Schools Full Sample | Control Schools Reduced Sample |
|---|---|---|---|---|
| Computations | 1st | 21 | 32 | 18 |
| | 5th | 50 | 56 | 52 |
| Concepts and Applications | 1st | 26 | 42 | 30 |
| | 5th | 45 | 46 | 41 |

*Note.* In the analysis of variance, degrees of freedom for the comparisons of DI schools and the full set of control schools are 1, 4652 for the analysis of computations and 1, 4655 for the analysis of concepts and applications. Degrees of freedom for the comparisons of DI schools and the reduced set of control schools are 1, 1912 for the analysis of computations and 1, 1913 for the analysis of concepts and applications.

## Table 5

*Mixed Model Results, Fifth-Grade Mathematics Achievement Regressed on First-Grade Achievement, School Advantage/Disadvantage, and Attending a DI School*

| | Comparing DI Schools and Total Group of Control Schools | | | Comparing DI Schools and Reduced Group of Control Schools | | |
|---|---|---|---|---|---|---|
| *Panel A: Computations* | *b* | *SE* | *p* | *b* | *SE* | *p* |
| Intercept | 38.5 | 0.9 | <0.0001 | 39.4 | 1.4 | <0.0001 |
| First Grade Score | 0.4 | 0.01 | <0.0001 | 0.4 | 0.02 | <0.0001 |
| Factor 1 | 1.4 | 0.7 | 0.06 | -0.04 | 1.3 | 0.97 |
| Factor 2 | -0.3 | 0.8 | 0.76 | -0.2 | 0.9 | 0.80 |
| Had DI in First Grade | 0.5 | 3.2 | 0.87 | 0.1 | 3.3 | 0.97 |
| | *estimate* | *SD* | *Sig.* | *estimate* | *SD* | *Sig.* |
| Random Effects Estimate | 51.4 | 8.4 | <0.0001 | 57.4 | 13.5 | <0.0001 |
| Residual Estimate | 363.2 | 7.6 | <0.0001 | 361.6 | 11.8 | <0.0001 |
| BIC | 40877 | | | 16829 | | |
| - 2 Log Likelihood | 40845 | | | 16801 | | |
| | *value* | *df* | *Sig.* | *value* | *df* | *Sig.* |
| Change in Log Likelihood | 1593 | 4 | <0.0001 | 713.9 | 4 | <0.0001 |
| *Panel B: Concepts and Applications* | *b* | *SE* | *p* | *b* | *SE* | *p* |
| Intercept | 26.9 | 0.8 | <0.0001 | 26.1 | 1.1 | <0.0001 |
| First Grade Score | 0.44 | 0.01 | <0.0001 | 0.47 | 0.02 | <0.0001 |
| Factor 1 | 1.7 | 0.6 | 0.004 | 0.9 | 1.0 | 0.35 |
| Factor 2 | -0.7 | 0.6 | 0.25 | -0.6 | 0.7 | 0.43 |
| Had DI in First Grade | 7.0 | 2.3 | 0.003 | 6.8 | 2.3 | 0.004 |
| | *estimate* | *SD* | *Sig.* | *estimate* | *SD* | *Sig.* |
| Random Effects Estimate | 34.7 | 5.5 | <0.0001 | 38.3 | 8.7 | <0.0001 |
| Residual Estimate | 181.0 | 3.8 | <0.0001 | 167.9 | 5.5 | <0.0001 |
| BIC | 37686 | | | 15384 | | |
| - 2 Log Likelihood | 37653 | | | 15357 | | |
| | *value* | *df* | *Sig.* | *value* | *df* | *Sig.* |
| Change in Log Likelihood | 2643.3 | 4 | <0.0001 | 1335.8 | 4 | <0.0001 |

*Note.* For the comparisons with the full set of control schools, results for the Random Effects Model were, for computations: random effects estimate = 54.8, *SE* = 9.0, *Sig.* <0.0001; residual estimate = 434.6, *SE* = 9.0, *Sig.* <0.0001; BIC = 42452, -2 Log Likelihood = 42438, correlation ratio = 0.11. Results for concepts and applications were random effects estimate = 35.1, *SE* = 5.8, *Sig.* <0.0001; residual estimate = 267.7, *SE* = 5.6, *Sig.* <0.0001, BIC = 40303, -2 Log Likelihood = 40297; correlation ratio = 0.12. For the comparisons with the reduced set of control schools, results for the Random Effects model were, for computations: random effects estimate = 53.2, *SE* = 12.8, *Sig.* <0.0001; residual estimate = 422.0, *SE* = 13.6, *Sig.* <0.0001, BIC = 17527, -2 Log Likelihood = 17515, correlation ratio = 0.11. For concepts and applications the values were random effects estimates = 26.4, *SE* = 6.7, *Sig.* <0.0001; residual estimate = 262.5, *SE* = 8.5, *Sig.* <0.0001, BIC = 16704, -2 Log Likelihood = 16692, correlation ratio = 0.09.

In all cases t-ratios can be computed by dividing the coefficient (random or residual estimate or the b values) by the standard error. The -2 log likelihood figure is a standard measure of model fit with a chi-square distribution. The change in log likelihood is calculated by subtracting the value in the more complex value from the value in the less complex value.

in improving concepts and applications scores. In each group of schools, the students in the sample had gains in computational achievement from first to fifth grade that would be considered educationally significant, well beyond the 0.25 criterion. With the measure of concepts and applications, gains for students in both the reduced set of control schools and the DI schools were educationally significant, but the effect size for students in the DI schools was twice as large (0.61 compared to 0.32).

## *Summary and Discussion*

This paper had two aims. First, we examined changes in the average achievement of first graders in the BCPSS from 1998 to 2003, comparing students in schools that implemented DI with students in other schools in the district. First-grade students who received DI had significantly higher levels of mathematics achievement than students who did not receive DI, and these differences became stronger when measures of the school context were statistically controlled. Even more important, the differences became stronger over time as school reforms were fully implemented and incorporated within the schools. The students in the DI schools had achievement scores significantly lower than those in the control schools at the beginning of the implementation (1998) but significantly higher by the end of the study period (2003). Although first graders throughout the system had higher achievement scores in 2003 than in 1998, the percentage change over this time period was more than twice as great for students in the DI schools than for students in the control schools. These differences in achievement were statistically significant and also met established criteria for educationally meaningful results.

We then looked at a smaller subset of students (those who began first grade in the BCPSS in 1997-98 or 1998-99 and remained in the same schools as fifth graders in 2001-02 or 2002-03) and examined changes in the students' achievement from first to fifth grade. All students in the sample had higher levels of mathematics achievement in fifth grade than in first grade. However, students in the DI schools had significantly higher increases in their mathematics concepts and applications scores than other students. This result appeared when comparisons were made to students in all control schools and to students in a more limited set of low-achieving schools. They also continued to appear when strong controls for first-grade achievement and school context were applied.

Several possible implications of these results should be noted. First, they provide additional support to the body of studies that has documented the superiority of DI in promoting mathematics achievement. Students in schools with DI had significantly higher mathematics achievement than students in schools with other curricula. These results appeared in first grade on both the measure of computations and the measure of concepts and applications and in fifth grade on the measure of mathematics concepts and applications. The differences were statistically significant and usually surpassed the standard criterion of educational importance.

Second, the results show the necessity of examining a longer period of time to assess the full implications and results of school-reform efforts. The changes in first-grade achievement only occurred after several years of implementation within a school. Berkeley (2002) provided a fascinating description of the process of change within one of the DI schools in this sample, beginning with the spring of 1996, before implementation began. At that point the school was:

> ...a place of failure for both students and teachers. Many children could not read, and test scores were abysmal. ...
> Children listened to teachers only when

> they felt like it, roamed the halls, and left the building. The faculty, spinning like tops, reacted to one crisis after another. When DI was implemented the following fall, the primary focus was behavior management, as the faculty had to establish order before they could teach children to read and write. (Berkeley, 2002, p. 222)

This description shows that part of successfully implementing a curriculum, especially in a failing school, is simply establishing a context in which children can learn and teachers can teach. As noted above, Engelmann and Engelmann (2004) suggest that it takes approximately six years for reforms to become fully institutionalized within a school. The results in this paper appear to confirm that estimate.

Third, the results reinforce the notion that learning to teach DI well takes time and practice as well as skilled technical support. It should be recalled that all of the DI schools had strong technical support from the National Institute for Direct Instruction. Work on reading achievement has shown that such strong technical support is important for helping to ensure that strong achievement gains occur (Stockard, 2008). Berkeley's (2002) description of one of the DI schools in the sample also stresses the extent to which learning to teach well takes time, practice, and extensive support.

As a corollary, it could be noted that these results address a common claim that DI involves a relatively mindless application of a script with little intellectual engagement or skill required by the teacher. The fact that strong changes in achievement only appeared over an extended period of time shows that learning to teach DI involves a great deal of skill and practice. Those who are most successful have more practice and more skilled, technical guidance.

Finally, it is important to note the impact of DI on the measure of concepts and applications, typically assumed to be tapping more of a "higher-order" skill than the measure of computations. Some have suggested that DI is useful for teaching rote, elementary skills, such as computations, but has more limited utility in teaching higher-level skills, such as those tapped by the mathematics concepts and applications measure. The results in this paper would contradict that view, for it was in the area of concepts and applications that the impact of DI was most apparent, especially in the analysis of effects on achievement in fifth grade.

The results presented in this paper add to the literature that has documented the efficacy of DI as a whole-school reform model. In contrast to many studies that have focused only on reading, this study examined changing achievement patterns in mathematics, showing impacts on average levels of achievement in the first grade as well as continuing effects related to the more advanced concepts and applications area in fifth grade. The implementation was strongly supported with extensive preservice, inservice, and consultation services, and the strongest effects were found as DI became fully implemented and incorporated within the schools.

## *References*

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

Applefield, J., Huber, R., & Moallem, M. (2000/2001). Constructivism in theory and practice: Toward a better understanding. *The High School Journal, 84,* 35-53.

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103,* 51-69.

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.

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.

Bowden, A. R., Greene, M., & Campbell, P.F. (n.d.) *Baltimore's K-5 Mathematics Program: How Synergistic Partnerships Result in Success.* Retrieved October 13, 2008, from http://www.project2061.org/events/meetings/textbook/policy/presentations/bowden.pdf

Carnine, D. (1990). Reforming mathematics instruction: The role of curriculum materials. *Direct Instruction News, 10,* 5-16.

Comprehensive School Reform Quality Center. (2006). *CSRQ center report on elementary school comprehensive school reform models.* Washington, DC: American Institutes for Research.

Darch, C., Carnine, D., & Gersten, R. (1984). Explicit instruction in mathematics problem solving. *Journal of Educational Research,* 77, 351-359.

Davis, R. B. (Ed.) (1990). Constructivist views on the teaching and learning of mathematics. *Journal for Research in Mathematics Education: Monograph No. 4.* Reston, VA: National Council of Teachers of Mathematics.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170-177.

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

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

Engelmann, S. E., & Engelmann, K. E. (2004). Impediments to scaling up effective comprehensive school reform models. In T. K. Glennan, S. J. Bodilly, J. R. Galegher, & K. S. 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 Corporation.

Fashola, O. S., & Slavin, R. E. (1997). Promising programs for elementary and middle schools: Evidence of effectiveness and replicability. *Journal of Education for Students Placed at Risk, 2,* 251-307.

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

Kidron, Y., & Darwin, M. J. (2007). A systematic review of whole school improvement models. *Journal of Education for Students Placed at Risk, 12,* 9-35.

Moore, L. J., & Carnine, D. W. (1989). A comparison of two approaches to teaching ratio and proportions to remedial and learning disabled students: Active teaching with either basal or empirically validated curriculum design material. *Remedial and Special Education, 10,* 28-37.

McLean, J. E., O'Neal, M. R., & Barnette, J. J. (2000, November). *Are all effect sizes created equal?* Paper presented at the annual meeting of the Mid-South Educational Research Association ED 448 188, Bowling Green, KY.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2005). *IEA's TIMSS 2003 international report on achievement in the mathematics cognitive domains: Findings from a developmental project.* Boston: TIMSS & PIRLS International Study Center.

National Center for Education Statistics (2008). *Digest of Education Statistics: 2007.* Retrieved January 3, 2010, from http://nces.ed.gov/programs/digest/d07/

Przychodzin, A. M., Marchand-Martella, N. E., Martella, R. C., & Azim, D. (2004). Direct Instruction mathematics programs: An overview and research summary. *Journal of Direct Instruction, 4,* 53-84.

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

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.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics, 24,* 235-270.

Stein, M., Silbert, J., & Carnine, D. (1997). *Designing effective mathematics instruction: A Direct Instruction approach.* Upper Saddle River, NJ: Prentice-Hall.

Stockard, J. (2008). Improving first grade reaching achievement in a large urban district: The effects of NIFDI-supported implementation of Direct Instruction in the Baltimore City Public School System (Technical Report # 2008-1). Eugene, OR: National Institute for Direct Instruction.

Troff, D. (2004). *An explicit instruction design approach for teaching students with learning disabilities to solve mathematical problems involving proportions.* Unpublished doctoral dissertation, Utah State University.

Woodward, J. (2004). Mathematics education in the United States: Past to present. *Journal of Learning Disabilities, 37,* 16-31.

Woodward, J., Baxter, J., & Robinson, R. (1999). Rules and reasons: Decimal instruction for academically low achieving students. *Learning Disabilities Research and Practice, 14,* 15-24.

# Author Notes

The author is grateful to the following people for careful reading and advice on earlier versions of this paper: Douglas Carnine, Cristy Coughlin, Gary Davis, Kurt Engelmann, Siegfried Engelmann, Amy Johnston, Daniel Johnston, and Jerry Silbert. Any errors or problems that remain with the manuscript are solely the responsibility of the author.