Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System

Jean Stockard, Ph.D.
Director of Research
National Institute for Direct Instruction
Eugene, Oregon

September 2008

Technical Report 2008-03

PO Box 11248 • Eugene, OR 97440
1-877-485-1973

Table of Contents

Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System

Executive Summary

In the late 1990s, in response to very low achievement scores, the Baltimore City Public School System (BCPSS) implemented curricular reforms. Eleven schools in the district introduced Direct Instruction as part of a whole school reform under the guidance of the National Institute for Direct Instruction. They began by implementing the reading curriculum and, one to two years later, added mathematics Direct Instruction, using *Distar Arithmetic* and *Connecting Math Concepts*. The Comprehensive Test of Basic Skills (CTBS) was administered to all first graders and all fifth graders in BCPSS in the spring of 6 school years: 1997-1998 through 2002-2003. Data were obtained from over 40,000 first grade students on mathematics achievement in computations, concepts and applications, and a composite score. Of the students who were in first grade in 1997-98 and 1998-99, more than 4,500 remained in the system and in the same schools in 2001-02 and 2003-03.

This study compares students' mathematics achievement in schools that implemented Direct Instruction and (DI schools), and schools without Direct Instruction (the Control schools). Key findings are listed below.

- Combining data across all years (1998-2003) first grade students in schools that implemented Direct Instruction had significantly higher mathematics achievement than students in other schools, and these differences became even stronger when demographic characteristics of the schools were controlled.
- The impact of being in a DI school became greater over time as the curricular changes became institutionalized within the schools. At the beginning of the study period average scores in the DI schools were significantly lower than those in the Control Schools, but by the end of the period they were significantly higher.
- First grade students in both groups of schools had higher achievement scores in 2003 than in 1998, but the increase was significantly larger for students in NIFDI-supported schools than for students in the other schools. On average, first grade composite mathematics achievement scores in the DI schools increased by 135 percent from 1998 to 2003, but by only 54 percent in the Control Schools.
- The magnitude of the effect on first grade achievement from attending a school that implemented DI mathematics with NIFDI support was statistically significant and substantively large.
- Among the students who remained in the same school from first grade to fifth grade, those in the DI schools had significantly lower achievement in first grade than those in the other schools. By fifth grade, however, these differences had disappeared.

List of Tables

List of Figures

Improving Elementary Level Mathematics Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in Baltimore City Public Schools

Low achievement in mathematics has been a major concern of school districts throughout the country. The Baltimore City Public School System (BCPSS) 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. Eleven schools in the district introduced Direct Instruction as part of a whole school reform under the guidance of the National Institute for Direct Instruction (NIFDI).[1] The schools began with the reading curriculum and, after one to two years, introduced the mathematics curriculum. Two mathematics curricula were used, depending upon the students' level and needs: *Distar Arithmetic* and *Connecting Math Concepts*. Other schools in the district used a variety of curricular materials.

A large body of literature, including well designed meta-analyses, has demonstrated the effectiveness of Direct Instruction in promoting high achievement, although somewhat fewer studies address the mathematics curricula with data from an extended time period and from a relatively large number of schools. This paper addresses this area by examining data on the mathematics achievement of first and fifth graders in the BCPSS from 1998 through 2003. The first part of this report focuses on the mathematics achievement of first graders in the BCPSS from 1998 through 2003, comparing the achievement of students in schools that used the Direct Instruction to the achievement of other students in the district. The second part focuses on the subset of students who were in the same schools in the BCPSS in first and fifth grade and examines the extent to which experiencing the Direct Instruction curriculum in first grade is associated with higher achievement in fifth grade. In this report results are summarized in graphs and are also expressed as effect sizes. Appendices provide full statistical details.

A. First Grade Mathematics Achievement

The analysis of first grade mathematics achievement uses data from over 40,000 first graders in 119 schools in the Baltimore Public School System from 1997-98 through 2002-2003. Achievement measures were obtained each spring using the Comprehensive Test of Basic Skills (CTBS), a widely used, nationally normed standardized test. Subtest measures of "Computations" and "Concepts and Applications," as well as a Composite measure were available. Results in this part of the report are presented as percentiles. Percentiles can be simply interpreted as the percentage of students that would have scores lower than a given student. For instance, if a student has a score at the 60th percentile, 60 percent of all students had scores that were lower.

[1] One additional school was included in the original group that implemented Direct Instruction. However, it closed very soon after the implementation began and thus it is not included in this count or in any of the analyses. In addition, a few other schools in the system used Direct Instruction reading programs. They, however, did not use DI mathematics curricula.

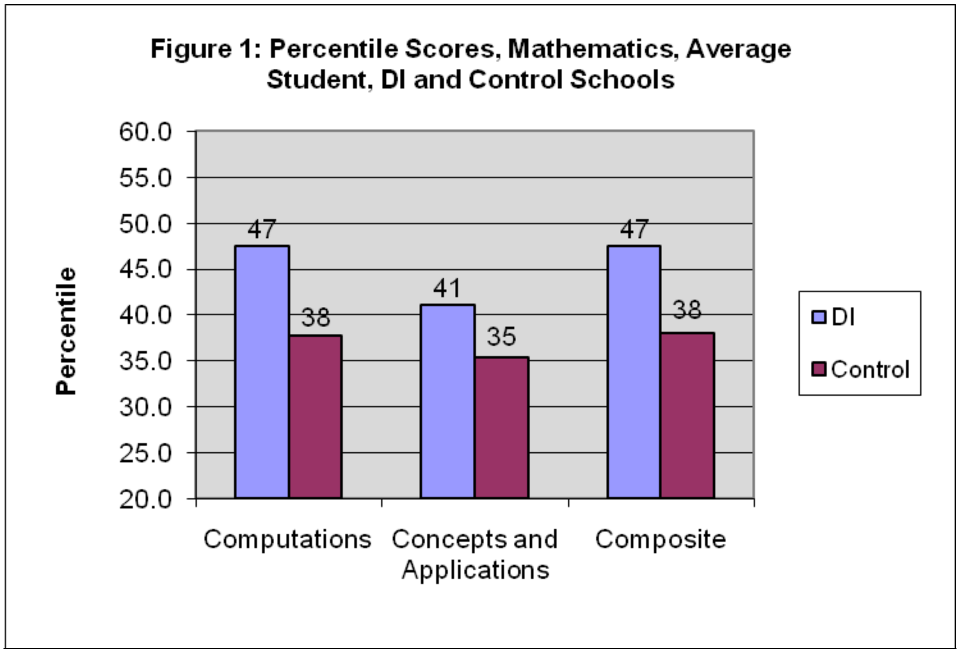
The average school in the BCPSS had large proportions of low income and minority students. On average, 75 percent of the students in a school were on free or reduced lunch. This varied however, from a low of 22 percent to a high of 93 percent. On average 84 percent of the students in average elementary school were African American, although the percentage varied from a minimum of 3 to almost 100 percent. Reflecting this high level of segregation, schools also varied in their representation of non-Hispanic white students, from having virtually no white students to being 94 percent white, with the average school having 14 percent of its students reported as being non-Hispanic white. Asian, Hispanic and Native American students were quite rare in the district, with an average representation of 1 percent in the student bodies. The DI Schools were similar to the Control Schools in racial-ethnic composition. However, the DI schools had proportionately more students who received free or reduced lunch: an average of 83 percent in the DI Schools compared to an average of 74 percent in the Control Schools.

The analysis focused on comparing the achievement of students who received Direct Instruction with those who did not. Full details on the statistical techniques and results are given in Appendix II. Four research questions were addressed:

- Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?
- Did students in DI Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?
- Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?
- What is the magnitude of the effect of have Direct Instruction on mathematics achievement?

1. Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?

Figure 1 summarizes achievement data for students who received Direct Instruction and those who did not on the three measures of mathematics achievement. These results combine the data from all the available years: 1997-98 through 2002-03 and represent the percentile score of the average student in a school. Combining data from all the years in the study, students who received Direct Instruction had significantly higher achievement scores than students in the Control Schools. The average first grader in a DI school scored at the 47^{th} percentile, slightly below the national norm, in both the measure of computations and the composite score, but somewhat lower, at the 41^{st} percentile, on the measure of concepts and applications. The average student in the Control schools had scores that were substantially lower, between the 36^{th} and 38^{th} percentile. These differences were statistically significant.

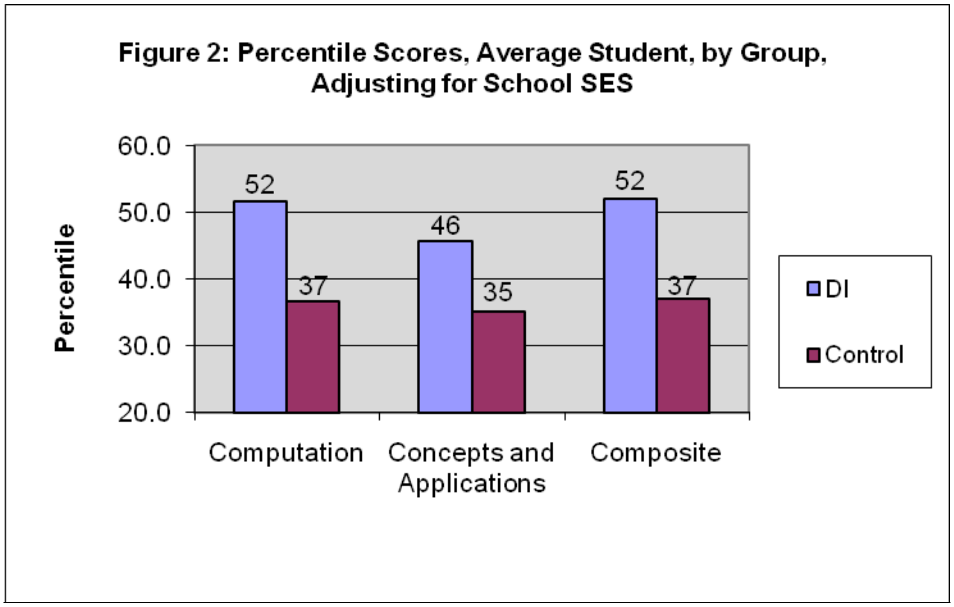


Note: Scores in this graph represent the percentile score of an average student in each type of school, averaging data over all years in the analysis: 1998-2003.

2. Did students in Direct Instruction Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?

Many studies have demonstrated that the socio-economic characteristics of a school have a strong influence on achievement, and thus it is important to control for this factor. As noted above, elementary schools in the Baltimore City Public School System varied substantially in their racial-ethnic composition and poverty rate, and, as in other studies students in BCPSS schools with fewer children from disadvantaged backgrounds had substantially higher average levels of achievement.

When the socio-economic characteristics of the schools were statistically controlled the differences in mathematics achievement between students in the DI schools and those in other schools became more marked. Figure 2 summarizes these results. The numbers in this graph represent the percentile score for an average student in each group if schools were equal in their racial-ethnic and poverty status. (Details regarding these calculations are in Appendix II.) After adjusting for the socioeconomic characteristics of the schools, the average student in a Direct Instruction School scored at the 52nd percentile on both the measures of computations and the composite score and at the 46th percentile on the measure of concepts and applications. In contrast, the average student at a Control School scored between the 35th and 37th percentile.



Note: Scores in this graph represent the percentile score of an average student in each type of school if the school contexts were equal in their socio-economic and demographic characteristics, averaging over all the years in the analysis: 1998-2003.

3. Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?

Changing school practices takes time and effort. It is reasonable to expect that the advantage of an effective instructional program would become more apparent as schools have more experience with a new curriculum, teachers gain more practice, and the procedures become part of the institutionalized and accepted practices within a school. During the time period under study the BCPSS paid concerted attention throughout the district to enhancing student achievement. The data indicate that these efforts were fruitful. Across the BCPSS first graders' mathematics achievement was significantly higher in 2003 than 1998. However, the changes over time were significantly larger for the students in the DI Schools than for those in the Control schools. Figures 3, 4, and 5 illustrate these changes. They show the percentile at which an average student in each group scored in each of the 6 years on each measure. Unlike the data in Figure 2, no adjustment was made for the school context of disadvantage.

Figure 3: Computation Achievement, Average Student, BCPSS, 1998-2003 by Group

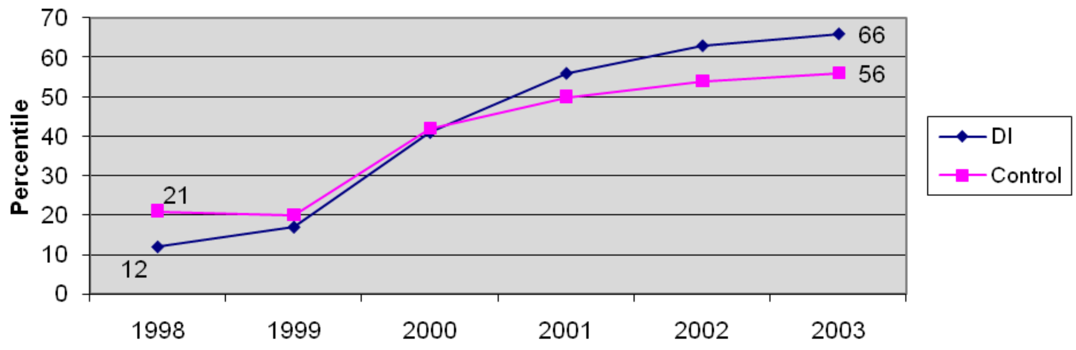


Figure 4: Concepts and Applications Achievement, Average Student, 1998-2003 by Group

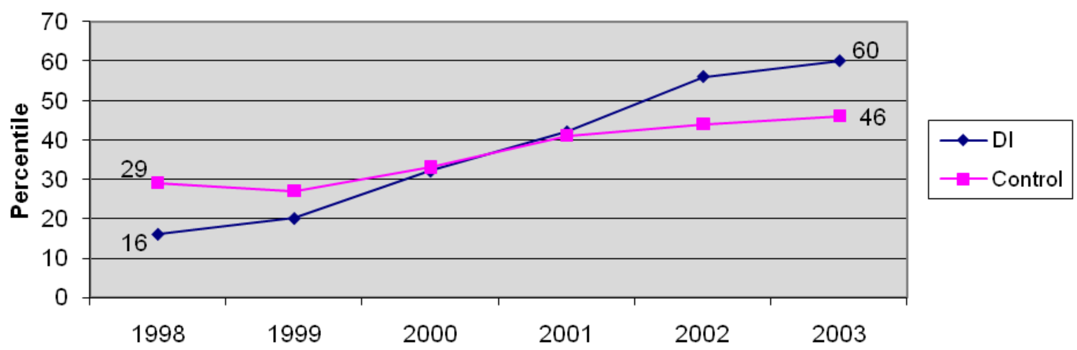
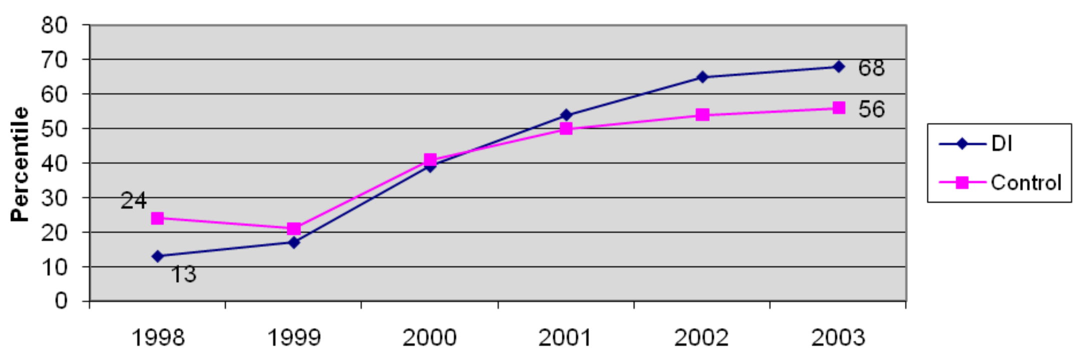


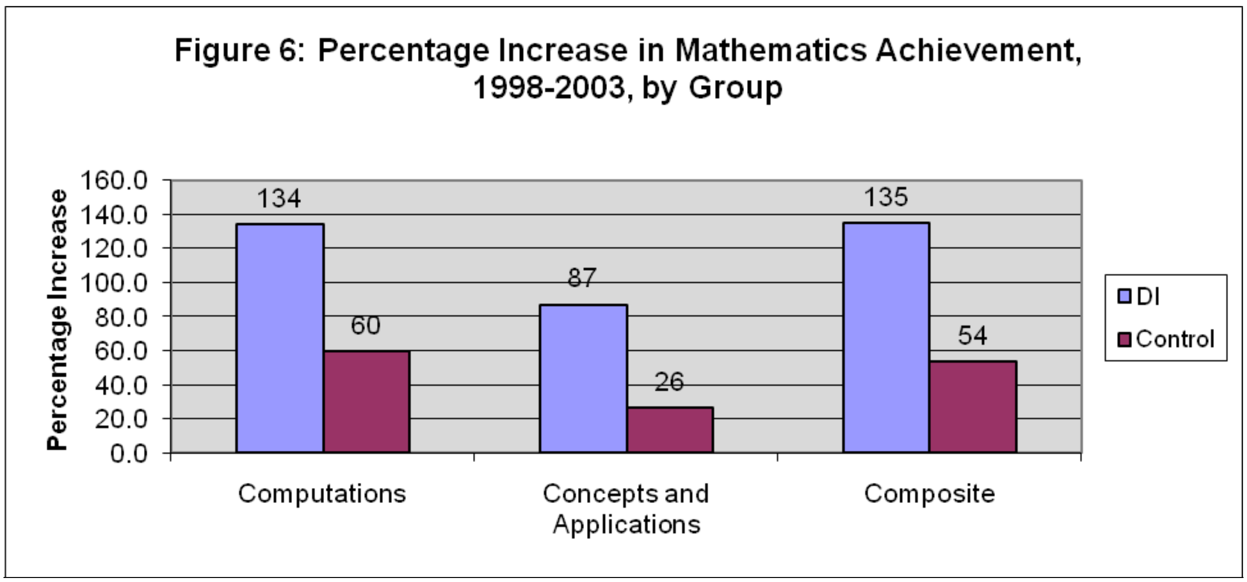
Figure 5: Composite Achievement, Average Student, BCPSS, 1998-2003 by Group



Note: Scores in Figures 3, 4, and 5 represent the percentile score of an average first grade student in each type of school for each year in the analysis.

At the beginning of implementation, in 1998, students in the DI Schools had significantly lower scores, on average, than students in the Control Schools. An average student in a DI school scored from the 12th (the computation and composite scores) to the 16th percentile (the concepts and applications score), while an average student in the Control Schools had scores ranging from the 21st to the 29th percentile. By 2003 the situation had changed markedly, with students in the DI schools having significantly higher scores than students in the Control Schools on all measures. By 2003, an average student in the DI Schools had scores were well above the national average, ranging from the 60th percentile on the measure of concepts and applications to the 68th percentile on the composite measure. In contrast, an average student in the Control Schools had scores ranging from the 46th percentile (on the measure of concepts and applications) to the 56th percentile.

Figure 6 illustrates the magnitude of these changes. Values in Figure 6 represent the difference between the 1998 and 2003 achievement scores as a percentage of the 1998 score, thus showing the percentage increase over time. For each measure, the changes over time were substantially larger for students in the DI schools than for students in the Control Schools. For instance, for the composite scores, first graders in DI schools had composite mathematics achievement scores in 2003 that were 135 percent higher than the scores of first graders in the same schools in 1998. In contrast, the scores of first graders in the Control Schools improved by only 54 percent. Similar differences appear with the other measures.



Note: The figures in this graph represent the percentage change in achievement from 1998 to 2003 for each of the groups of schools and each measure. These change figures were calculated using normal curve equivalent (NCE) scores rather than percentiles. As explained more fully in Appendix I, NCE scores are preferable for statistical procedures and were used for all calculations in the analysis. NCE scores were translated to percentiles for Figures 1-5.

4. What is the Magnitude of the Effect of NIFDI Supported Direct Instruction?

Unless specifically noted, all of the results presented in this report are statistically significant. That is, various statistical tests indicate that the results do not appear by chance. However, it is possible that statistical significance can simply reflect sample size. When data from large samples are subjected to statistical tests relatively small substantive differences can produce significant results. To combat this problem, researchers use measures of effect size. These measures provide descriptions of the magnitude of the results. One of the most commonly used measures is Cohen's *d*. It is simply calculated as the difference between two average values divided by the common standard deviation. Thus, it reports the magnitude of a difference between two groups as a proportion of the standard deviation. A d-value, or effect size, of .25 or larger has traditionally been considered educationally significant, indicating that an intervention has an important impact on students' achievement.

Figures 7 and 8 summarize the magnitude of the various effects reported earlier. Figure 7 summarizes the effects without any adjustment for school characteristics, using the data summarized in Figure 1. When data from all the years are examined together, the effect sizes comparing achievement of students in the DI Schools and those in the Control Schools are slightly lower than the .25 criterion (.13 for the measure of concepts and applications to .21 for the computations and composite scores). However, by 2003 the effect sizes were substantially larger, meeting or exceeding the .25 criterion: .25 for the measure of computations and.31 and .32 for the composite and concepts and applications measures.

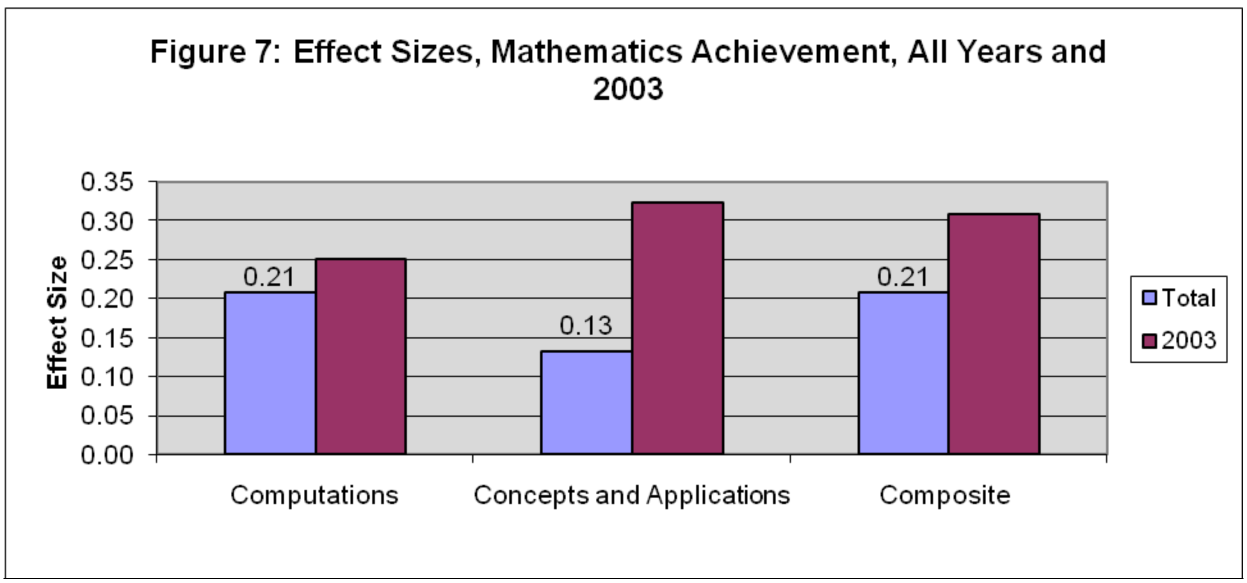
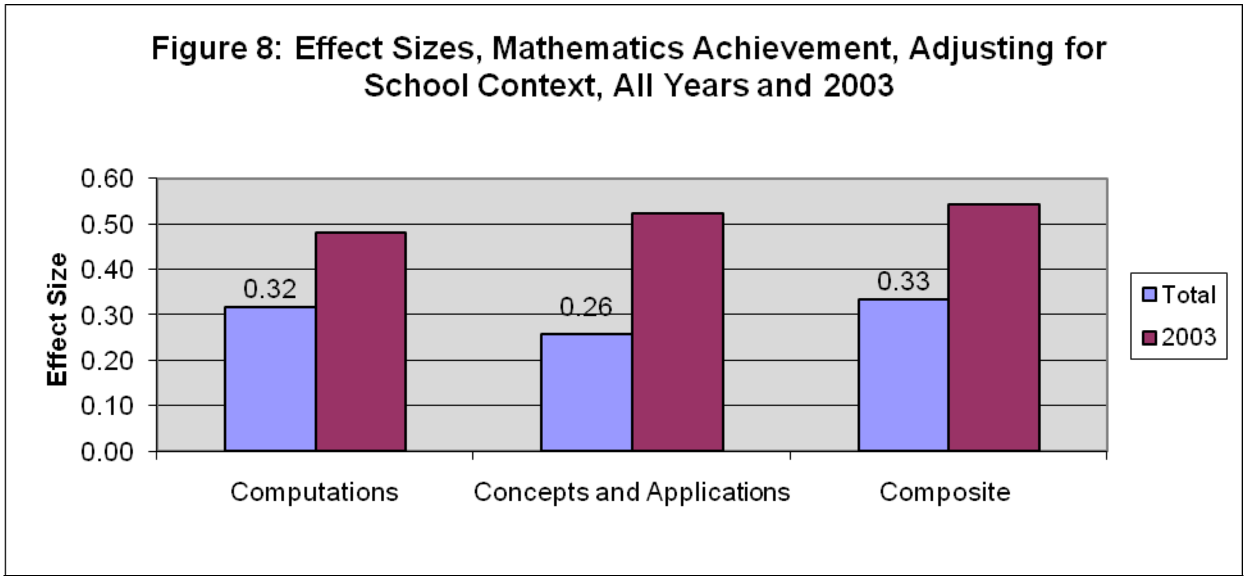


Figure 8 gives the effect sizes for the measure for all years and 2003 after the impact of school socio-economic status has been controlled (the data shown in Figure 2). Averaged across all years, the effect of being in a DI School relative to a Control School ranged from .26 for the measure of concepts and applications to 32 and .33 for computations and the composite score. At the end of the study period, the effect sizes were substantially larger: ranging from .48 to .54.



5. Summary

This analysis of the mathematics achievement of first grade students in the Baltimore City Public School System indicates that students in schools that implemented Direct Instruction had significantly higher achievement scores on all three indicators (computations, concepts and applications, and a composite score) than students in the Control Schools. These differences became even larger when the socio-demographic context of the schools was controlled and after several years of implementation of Direct Instruction. The effects were not just statistically significant, but substantively large and educationally important.

B. The Relationship of First Grade Direct Instruction and Fifth Grade Mathematics Achievement

Data were available for first and fifth graders in 1997-98 through 2002-03. While the first part of this report examined changes in first grade achievement over this time period, this section examines changes in achievement of individual students who began in the BCPSS in 1997-98 or 1998-99 and were still in the same schools five years later, in 2001-2 or 2002-3. As in the first part of the report, comparisons are made between students in the DI Schools and the Control Schools. (See Appendix III for full details on the sample used in the analysis.)

Two factors limited the size of the sample for this analysis: a high rate of turnover of students in the BCPSS and the fact that only six schools within the DI group had implemented Direct Instruction in mathematics by 1998-99. To provide additional insight into the results, two different control groups were used in the analysis: One includes all students in the Control Schools who started first grade in 1997-98 or 1998-99 and were in fifth grade in the same schools in 2001-02 or 2002-03. The second includes only students from schools with the lowest levels of achievement in the BCPSS in first grade in 1997-98 and 1998-99.[2] The DI group included 164 students, the total Control Group included 4,607 students, and the reduced Control Group included 1,820 students.

[2] Students who were retained, or held back a grade, were excluded from the analysis.

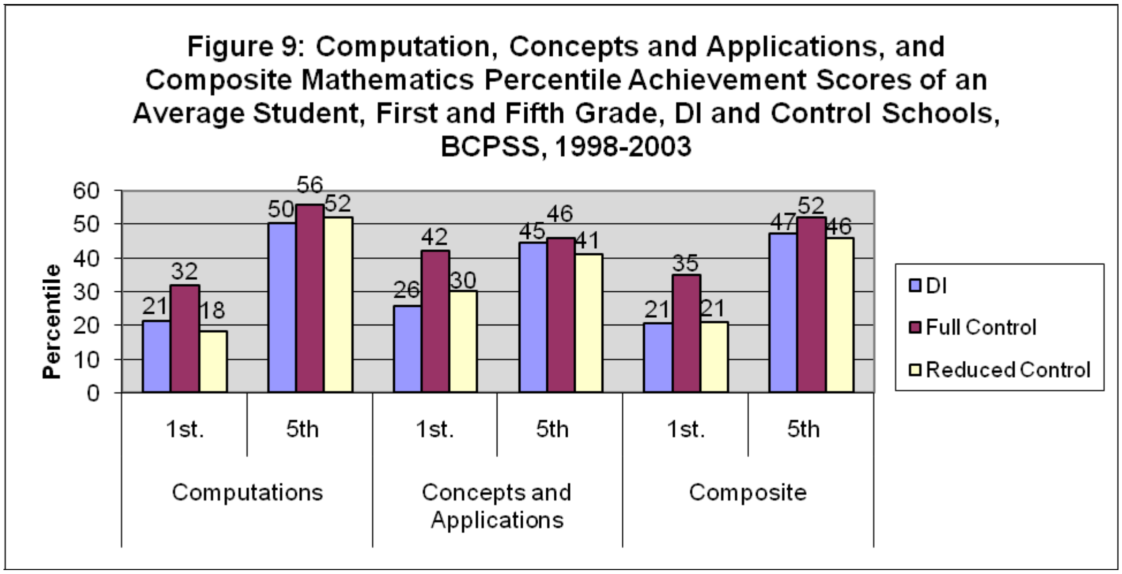
Three research questions were addressed:

- Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools?
- Did students who received Direct Instruction in first grade have higher mathematics achievement in fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?
- What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

1. Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Mathematics Achievement from First to Fifth Grade than Students in Other Schools?

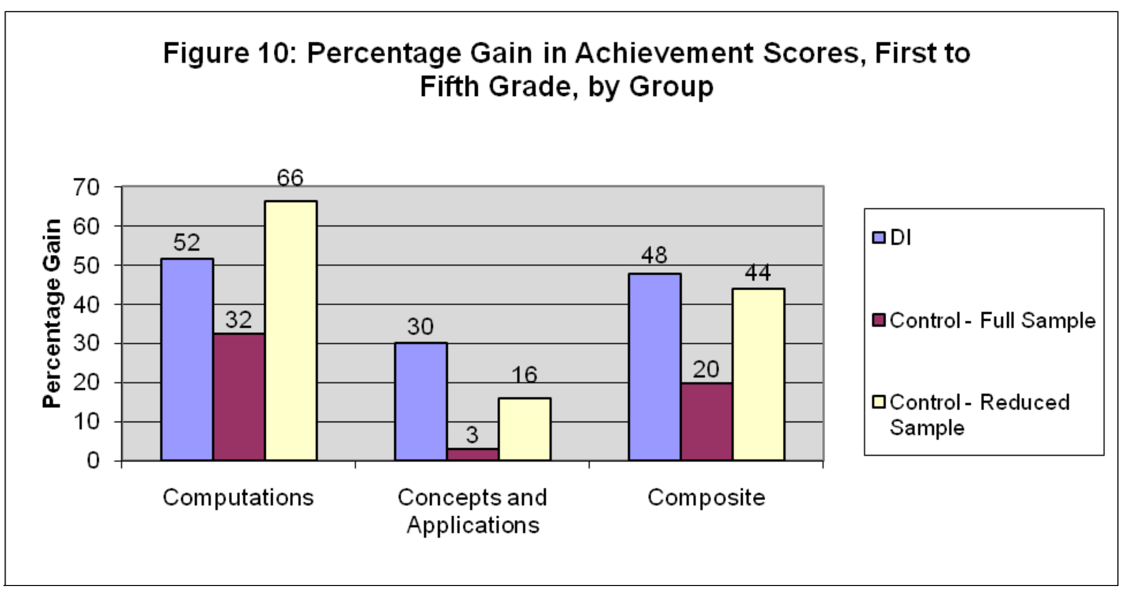
As explained fully in Appendix IV, several different statistical methods were used to examine differences in the changes in achievement from first to fifth grade of students who received DI and those who did not. All the results, whether with the full or reduced set of Control Schools, pointed to the same substantive conclusions: Students in the Direct Instruction schools had significantly greater gains than students in either of the set of Control schools on the measure of concepts and applications. Increases on the measure of computations and the composite measure did not differ significantly between the three groups. Figures 9 and 10 summarize these results.

Figure 9 reports the percentile score of an average student in each group of schools on each measure in both first grade and fifth grade. On average, all students, in both the DI and the Control Groups, had higher average achievement scores in fifth grade than in first grade. In first grade, the average student scored well below the national average, but by fifth grade, the average student, in all three groups, had scores that were much closer to, and in some cases above, the national average.



Note: Data in the figure represent the percentile at which an average student in a Control School or a DI school scored in first grade and in fifth grade. As with the analysis in Part A, all statistical calculations were conducted with Norm Equivalent Scores, which can be compared from one year to another, and the results were translated to percentiles for graphic display.

Figure 10 gives the percentage change in scores from first grade to fifth grade for students in each group on each measure. Students in the DI group had the highest average gain on both the measure of concepts and applications and the composite measure but, as noted above, these differences were only significant with the measure of concepts and applications. On average, students in the DI schools had concepts and applications scores in fifth grade that were 30 percent higher than their first grade scores, compared to an average increase of 3 percent for students in the full set of Control schools and 20 percent for students in the reduced set of Control schools.



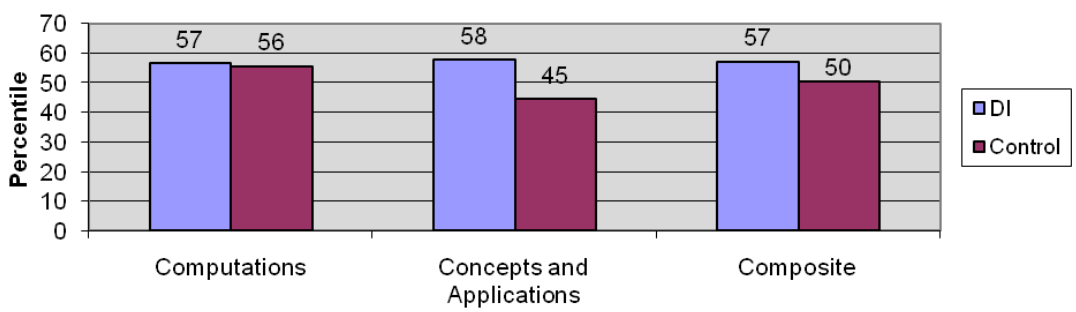
Note: The values in this table were calculated by dividing the difference in the fifth and first grade norm equivalent scores by the first grade score. Details on these calculations are in Appendix IV.

2. Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?

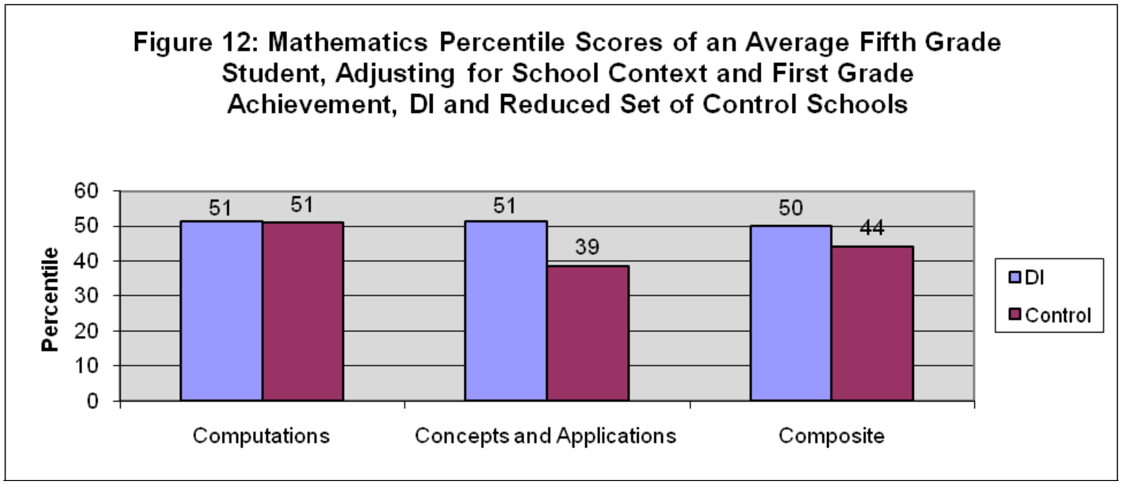
As described in Appendix IV, multivariate statistical techniques were used to address the second research question. This analysis compared the differences between students in the DI schools and those in the two sets of Control schools after statistically equalizing students on both their level of first grade achievement and the socio-demographic characteristics of their schools. The results were similar with analyses with both the full set of Control schools and the reduced set and parallel the results with Research Question 1: Students in Direct Instruction schools had significantly higher scores on the measure of mathematics concepts and applications than students in other schools, even when school context and first grade achievement were controlled. While the DI students also had higher scores on the measure of computations and the composite score, these differences were not statistically significant.

Figures 11 and 12 summarize the results of this analysis. The data in the figures are the percentile score that an average student in each group of schools would be expected to achieve if the schools had equal socio-demographic characteristics and if the students all had equal levels of achievement in first grade. Figure 11 compares the DI schools with the full set of Control schools, while Figure 12 compares the DI schools with the reduced sample of Control schools.[3] In both comparisons, the students in the DI Schools have higher levels of achievement, but the differences are substantially greater with the measure of mathematics concepts and applications.

Figure 11: Mathematics Percentile Scores of an Average Fifth Grade Student, Adjusting for School Context and First Grade Achievement, DI and Full Set of Control Schools



[3] Different multivariate analyses were conducted with these two separate samples. The lower average values in Figure 12 reflect the fact that the reduced sample of Control Schools had lower levels of first grade achievement as well as a socio-demographic school context that was more high-risk.

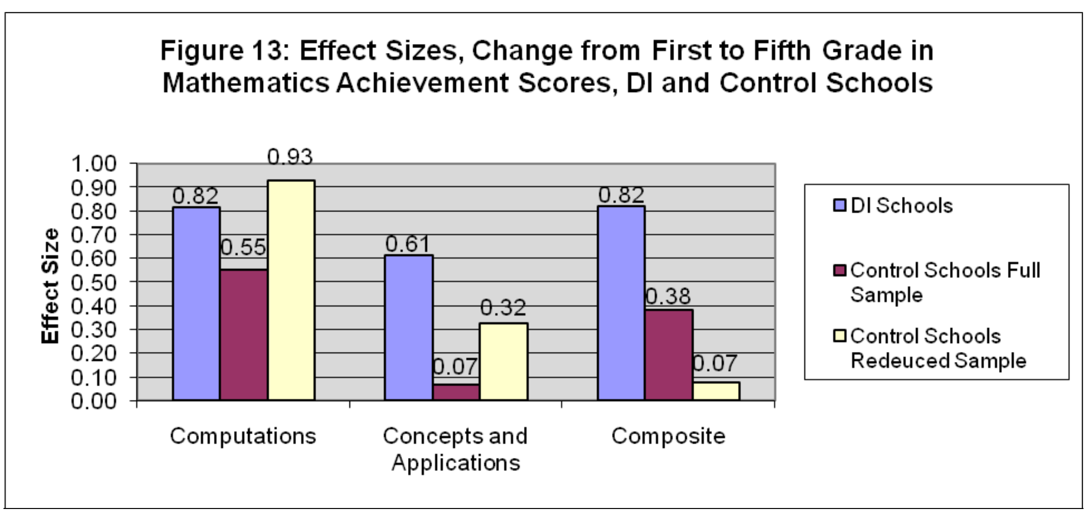


Note: Data in these figures represent the percentile score of an average student in each type of school in fifth grade if they had the same achievement scores in first grade and were in schools with equal socio-demographic characteristics.

3. What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

Standard tests of statistical significance are influenced by the size of a sample in the calculations. To avoid this bias researchers sometimes use measures of effect sizes. With data involving change over time these calculations consider both the size of the scores at the two time points as well as the correlation between these two measures. (Details are in Appendix IV.) Traditionally, measures of effect size of .25 or larger are considered educationally important.

Figure 13 gives the effect sizes associated with the changes from first to fifth grade for each of the measures of achievement for students in the DI schools and for those in the two sets of Control schools. All of the effect sizes associated with the DI schools far surpass the .25 criterion. For the full set of Control schools, effects surpass the .25 criterion for both the measure of computations and the composite score. For the reduced set of Control schools, effects surpass the .25 criterion for the measure of computations and the composite score.



C. Summary and Discussion

The first section of the report examined changes in the average achievement of first graders in the BCPSS from 1998 to 2003, comparing students in schools that implemented Direct Instruction and students in other schools in the district. Results indicated that, averaging over all years, first grade students who received Direct Instruction had significantly higher levels of mathematics achievement than students who did not receive Direct Instruction. These differences became stronger when measures of the school context were statistically controlled and also became stronger when trends were examined over time. The students in the DI schools had achievement scores that were significantly lower than those in the Control schools at the beginning of the implementation (1998) but scores that were significantly higher by the end (2003). The percentage change in average first grade achievement was more than twice as high for students in the DI schools than in the Control schools. These differences in achievement were statistically significant and also met established criteria for educationally meaningful results.

The second part of the report 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. It 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 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 large body of studies that has documented the superiority of Direct Instruction in promoting mathematics achievement. Students in schools with Direct Instruction had significantly higher achievement than students in schools with other curricula. These results appeared in first grade on all measures and in fifth grade on the measure of mathematics concepts and applications. The differences met and, usually, surpassed the criteria of statistical significance and educational importance.

Second, the results show the necessity of examining a relatively long 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. In addition, it should be noted that all of the DI schools had strong technical support from the National Institute of Direct Instruction. Other work on reading achievement has shown that such strong technical support is important for helping to ensure that strong achievement gains occur. This result also may address a common claim that Direct Instruction 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 suggests that learning to teach Direct Instruction involves a great deal of skill and practice. Those that are most successful have more practice and more skilled, technical guidance.

Finally, it is important to note the impact of Direct Instruction on the measure of concepts and applications, typically assumed to be tapping more of a "higher order" skill than the measure of computations, the other element of the composite score. Some have suggested that Direct Instruction 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 long-term impact of Direct Instruction was most apparent.

Appendix I
Sample and Measures – First Grade Analysis

This appendix provides details on the sample and the measures used in the analysis of first grade mathematics achievement.

Sample

The analysis uses data from over 40,000 first graders in 119 schools in the Baltimore Public School System from 1997-98 through 2002-2003. Eleven schools implemented Direct Instruction with technical assistance provided by NIFDI and are referred to below as the DI schools.[4] The remaining 103 schools are termed the Control schools. Schools in the intervention group are listed in Table A-1[5]

Table A-1 also lists the date at which programs were implemented in each school. All schools that used Direct Instruction began their implementation with the reading programs. They added DI instruction in mathematics in a subsequent year. In all but two schools (City Springs and Collington Square) the Direct Instruction mathematics curriculum was added one year after beginning reading instruction.

In the analysis schools were designated as belonging to a given condition only during the years in which their school was receiving the mathematics implementation. For instance, Collington Square had no mathematics DI instruction in either the first or second year of data collection (1997-98 and 1998-99), but began implementing Direct Instruction mathematics programs in 2000-2001. Thus, first grade students in Collington Square were included in the mathematics intervention group in 2000-2001 and subsequent years, but in the control condition in the two earlier years.

The analysis also considers the amount of time that a school had implemented the program. For instance, Collington Square was considered to be in its first year of implementation of Direct Instruction in 2000-2001, its second in 2001-2002, etc.

Data were available for a total of approximately 45,000 students (n = 44,766 for the composite measure, n = 45,287 for the measure of computation, and n = 45,535 for the concept and applications measure). Enrollment in the BCPSS declined from 1997-98 to 2000-2003, with data available for approximately 9000 students for the first year and 6400 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 one hundred students in the sample.

---

[4] One additional school, Charles Carroll, was part of the original 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.

[5] There were two additional schools in the set of control schools, but they, unfortunately, had the same name: Highland Town. The number assigned to the schools was not available for all years, and alternative spellings of the schools' name across years made it impossible to clearly differentiate them. Thus these two schools were eliminated from the analysis. Data were available for one home-schooled student and that was also omitted.

Table A-1: Schools in Treatment Group and Start Dates of Reading and Mathematics Direct Instruction

	Reading Start Year	Math Start Year
Arundel	96 - 97	97 – 98
CC Barrister	97 - 98	98 – 99
City Springs	96 - 97	98 – 99
Collington Square	98 - 99	00 – 01
Dickey Hill	98 - 99	99 – 00
Federal Hill	97 - 98	98 – 99
General Wolfe	96 - 97	97 – 98
Hampstead Hill	96 - 97	97 – 98
Langston Hughes	98 - 99	99 – 00
Margaret Brent	98 - 99	99 – 00
Dr. Rayner Browne	98 - 99	99 – 00

Measuring Achievement

The Comprehensive Test of Basic Skills (CTBS), a widely used standardized achievement test, was administered to all first graders in the spring of each year, from 1997-1998 through 2002-2003. 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 Concepts and Applications and Mathematics Computations, and a Composite Mathematics Achievement score were analyzed.

Normal Curve Equivalent (NCE) scores were used for all statistical analyses. NCE scores are calculated by translating the percentile scores to a distribution that is normally distributed. Like percentile scores, NCE scores range from 1 to 99 with a mean of 50. However, while percentile scores are evenly distributed (the graph of percentile scores would look like a rectangle), NCE scores comprise a normal distribution. The transformation results in scores that can be meaningfully added and subtracted, so that the difference of an NCE score of 50 and 55 (=5) is the same as the difference between 30 and 35. This interval scale allows computations of most common statistics. Percentiles and raw scores do not meet the technical requirements of interval scales, and thus statisticians urge researchers to analyze data based on NCE scores rather than percentiles or raw scores. All statistical calculations for this report were done using NCE scores.

The meaning of NCE scores is, however, not intuitively obvious. For this reason, results in the body of the text and the executive summary have been translated into percentile scores. After completing statistical calculations with NCE scores the resulting descriptive information was converted into percentiles for display in the graphs in the body of the report. This conversion was done using a standard conversion table.[6]

Measuring School Context

Given the strong influence of school context on student achievement, it was important to develop an efficient, yet strong, measure of the demographic context of the schools in the sample. Preliminary analysis indicated that the demographic characteristics of schools were very highly correlated from one year to another. Thus, one summary measure was developed for each school that would be valid for all the years included in the data set.

There were a few cases in which data were not available for a school for all years. A regression-based method was used to predict values of missing cases from other years (e.g. predicting 1998 levels of proportion African-American from levels in 1999 through 2003).[7] Values for each variable were averaged across the years to produce aggregate measures of the demographic characteristics of the school.

Table A-2 gives means and standard deviations on each of the demographic characteristics for the DI and the Control schools. Results of t-tests indicated no significant differences between the groups of schools in their racial-ethnic composition. However, the DI schools had significantly more students who received free or reduced lunch: 83 percent in the DI schools versus 74 percent in the Control schools.

A factor analysis was conducted to develop summary measures of school context. Six variables were included in this analysis: the average proportion of Asian, Hispanic, African American, Native American, and non-Hispanic White students as well as the average proportion of students receiving free and reduced lunch. A principle components extraction method with varimax rotation was used. Two significant factors (with eigenvalues greater than 1.0) were found.[8] Factor scores were computed for each school and saved.

Results for the factor analysis are given in Table A-3. The commonalities indicate that, except for the measure of proportion of Asian students, there is a great deal of shared variance among the indicators, ranging from .66 for the proportion of students receiving free and reduced lunch to .94 for the proportion of African American students. Eigenvalues are also relatively high, with 73.5 percent of the total variance between schools explained by the two factors.[9]

[6] The conversion can also be accomplished manually through using a normal curve table. The NCE scores can be converted to z-scores (z = ((nce-50)/21.06))). The percentile that corresponds to the z score can then be found in the normal curve table.
[7] There were two cases where a predicted value fell outside of the theoretical range. Both of these involved the predictions for proportion white, where the predicted values were less than zero. For these cases (Malcolm X and Mildred Monroe schools), an average of the other years in the data file was used as the predictor. The average values for both cases were .01 or less.
[8] The factor analyses were also conducted with data for each year separately (i.e. the non-aggregated data), and the results were virtually identical to those obtained with the aggregated data.
[9] N = 120 for these analyses, for Carroll School was included.

Table A-2: Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels By Group

	DI		Control			
	Mean	S.D.	Mean	S.D.	t	prob.
Asian	0.01	0.03	0.01	0.01	-0.88	0.40
Black	0.75	0.34	0.85	0.26	1.12	0.27
Hispanic	0.05	0.12	0.01	0.01	-1.10	0.30
Native American	0.01	0.02	0.002	0.004	-1.31	0.22
Non-Hispanic White	0.17	0.25	0.13	0.25	-0.49	0.62
Free and Reduced Lunch	0.83	0.06	0.74	0.15	-4.02	<.001
Factor 1	-0.12	0.96	0.02	1.01	0.44	0.66
Factor 2	1.18	2.82	-0.12	0.46	-1.53	0.16
N	11		108			

The first factor, which accounts for 44% of the total variance, has 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 lunch. Thus, schools with positive scores on this factor would have proportionately more non-Hispanic White and Asian students and many fewer African-American and poor students. As can be seen in Table A-2, there are no significant differences between the two groups on this factor.

Table A-3: Factor Analysis of School Characteristics, Elementary Schools, BCPSS,1998-2003

	Rotated Factor Loadings		Communalities
	Factor 1	Factor 2	
Proportion Asian	0.64	0.09	0.42
Proportion African-American	-0.88	-0.40	0.94
Proportion Hispanic	0.13	0.85	0.74
Proportion Native American	0.08	0.87	0.77
Proportion Non-Hispanic White	0.90	0.27	0.89
Proportion Free/Reduced Lunch Eligible	-0.80	0.14	0.66
Eigenvalues	2.66	1.75	
% of total variance	44.4	29.1	

N = 120 schools

The second factor, which accounts for 29% of the total variance, has 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 would have proportionately more Hispanic and Native American students and somewhat fewer African American students. The results in Table A-2 again indicate no significant differences between the groups on this factor.

Appendix II: Detailed Statistical Results – First Grade Analysis

This appendix expands upon the material presented in the text including an extended discussion of the analysis techniques employed and the results that address each research question.

Methodology

As noted in the text, the data analysis focused on comparing the mathematics achievement of first grade students who received Direct Instruction with those in the Control Schools. Four research questions were addressed.

The first research question was, "Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?" Simple t-tests were used to answer this question, comparing the average achievement scores of students in the DI schools with scores of students in the Control schools, combining data from all years of implementation.

The second research question asked, "Did students who received Direct Instruction have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools were equalized?" Mixed model regression, including measures of school advantage and disadvantage from the factor analysis as predictors, was used to answer this question. To the extent that Direct Instruction 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.

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.

The third research question asked, "Did the impact of being in a DI school increase over time as reforms became institutionalized within a school?" It was expected that the impact of the DI implementation would become stronger over time. This question was analyzed both through examination of mean scores over time, with independent sample t-tests to compare the means, as well as through adding variables regarding length of implementation to the mixed models.

The final research question asked, "What is the magnitude of the effect of Direct Instruction on mathematics achievement." To address this question, the results were translated into effect sizes. Effect sizes are commonly used to summarize the magnitude of differences between two groups. They are calculated by simply dividing the difference in the means of two groups by their common standard deviation. The resulting figure thus describes the magnitude of a difference between two groups as a percentage of the standard deviation. Unlike tests of significance, which can be heavily influenced by sample size, effect sizes are unaffected by the size of a sample. Effect sizes greater than .25 have traditionally been considered educationally significant.

Research Question 1: Did students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?

The first research question addressed was, "Do students who received Direct Instruction have higher mathematics achievement scores than students in the other schools?" This was answered through comparing the average achievement scores of students in the DI schools with students in the Control schools. In this analysis data for all years in the sample (1997-98 through 2002-03) were combined.

Table A-4 summarizes the information used for these comparisons. Panel A reports the means and standard deviations of the normal curve equivalent scores for students who received Direct Instruction and students in the Control schools on each of the measures of achievement as well as the results of t-tests examining the hypothesis that any differences occur by chance. Panel B translates the average Normal Curve Equivalent scores in Panel A to the corresponding percentiles. The values in Panel B may be interpreted as the score of an average student in each type of school over all the years in the study and were used to construct Figure 1 in the text.

Table A-4: Average Mathematics Achievement Scores of Students With and Without Direct Instruction, All Years, First Graders, BCPSS

Panel A: Descriptive Statistics, Norm Equivalent Scores, and t-test Results

Computations	Control Schools	Direct Instruction
Mean	43.4	48.6
S.D.	24.9	25.8
N	42838	2449
t-value	9.73	
d.f.	2716	
prob.	<.001	
Concepts and Applications		
Mean	42.3	45.2
S.D.	22.1	22.9
N	43072	2463
t-value	6.40	
d.f.	45533	
prob.	<.001	

Composite

Mean	43.6	48.6
S.D.	24.0	25.4
N	42340	2426
t-value	9.47	
d.f.	2679	
prob.	<.001	

Panel B: Percentile Scores Corresponding to Mean Values

	Control School Students	Direct Instruction Students
Computations	37.7	47.4
Concepts and Applications	35.7	41.0
Composite	38.0	47.4

Note: For the composite measure and the measure of computations, the F-test comparing the variances of the two groups indicated that the hypothesis of equal variances should be rejected. Thus the t-values and degrees of freedom reflect the use of the unequal variance formulas for those two analyses.

The results in Table A-4 show that students in schools with Direct Instruction had significantly higher achievement scores than those in the Control Schools on all three measures: Computations, Concepts and Applications, and the Composite score. The average student in the Control schools, across all the years in the study, had scores ranging from the 35th to the 38th percentile. In contrast, the average student in a DI school had scores ranging from the 41st to the 47th percentile. The lowest scores for both groups occurred with the measure of Concepts and Applications. In neither of the groups did the average student reach the national average of the 50th percentile when scores were combined across all years.

Research Question 2: Did students in DI Schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools are equalized?

Research Question 2 deals with the extent to which students in Direct Instruction schools have higher mathematics achievement scores than students in other schools when the demographic characteristics of their schools are equalized. (See Stockard and Mayberry, 1992, for an extensive discussion of the literature regarding such "school effects.") This question is addressed with mixed model regressions. Table A-5 summarizes the models that were tested and gives summary information on the results with each model. Panel A gives the variables that are included and Panel B gives the model fit statistics.

Table A-5: Variables in Models and Measures of Fit, Analysis of First Grade Mathematics Achievement, BCPSS, 1998-2003

Panel A: Variables in the Models

Variables	Model 1	Model 2	Model 3	Model 4
Random intercept for Schools	x	x	x	X
Year		x	x	X
Factors 1 and 2		x	x	X
DI School			x	X
DI * Year				X

Panel B: Model Fit Statistics

	Model 1	Model 2	Model 3	Model 4
Computations				
Random Effects Estimate	47.5	37.6846	38.0	37.8873
s.e.	6.5	5.1378	5.2	5.1685
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	582.9	515.3	514.5	514.3
s.e.	3.9	3.4	3.4	3.4
sig.	<.0001	<.0001	<.0001	<.0001
BIC	417320	411740.3	411646	411661.3
- 2 Log Likelihood	417306	411711.6	411660	411623.1
Change in LL		5594	52	37
Df		3	1	1
P		<.0001	<.0001	<.0001
PRE measure		0.12	0.001	0.0005
Correlation ratio	0.08			
Concepts and Applications				
Random Effects Estimate	46.4	34.8	34.5	34.6
s.e.	6.4	4.8	4.8	4.8
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	451.6	434.5	434.1	433.2
s.e.	3.0	2.9	2.9	2.9
sig.	<.0001	<.0001	<.0001	<.0001
BIC	408014	406238.6	406202.5	406112.9
-2 Log Likelihood	407999.7	406209.9	406169	406074.7
Change in LL		1790	41	94
Df		3	1	1
P		<.0001	<.0001	<.0001
PRE measure		0.04	0.001	0.0021
Correlation ratio	0.09			
Composite				

Random Effects Estimate	53.4	40.5	40.5	40.4
s.e.	7.3	5.5	5.5	5.5
sig.	<.0001	<.0001	<.0001	
Residual Estimate	533.2	474.8	474.0	473.47
s.e.	3.6	3.2	3.2	3.2
sig.	<.0001	<.0001	<.0001	
BIC	408554	403360	403293	403245
-2 Log Likelihood	508539	403331	403259	403207
Change in LL		105209	71	53
Df		3	1	1
sig.		<.0001	<.0001	<.0001
PRE measure		0.11	0.002	0.0012
Correlation ratio	0.09			

Note: The correlation ratio is computed by dividing the random effects estimate for schools by the sum of the estimate for schools and the residual. Thus, the ratio represents the proportion of total variation that is between schools. The PRE measure is the ratio of the difference of residual estimates of two models divided by the estimate from the less complex model. Thus it tells the proportionate change in the variance that occurs by adding more variables to a model.

Model 1 is the baseline "intercept only" or "random effects" model and only includes schools as a random variable. This tests the null hypothesis that the schools are equal in average reading achievement. The correlation ratio attached to Model 1 is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that from eight to nine percent 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 Model 1 and with all subsequent models. There is significant variation between schools and also between students in all models that we examine. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those available to in this analysis.

The models become incrementally more complex, with each subsequent model adding one or more explanatory variables to test the research questions, as indicated by the "x's" associated with each model in Panel A. Model 2 adds the year in which data were collected and the two factor scores to the baseline model. Year was included to test the hypothesis that test scores vary over the years in the study (1997-98 through 2002-03). This is important to control for any general changes within the district. Including the two factor scores is important to control for the extent to which the demographic context of a school affects student achievement. It was expected that Factor 1, with higher loadings for schools with less poverty, fewer African

American students and more non-Hispanic white students, would be positively associated with achievement. It was expected that students in schools with higher scores on Factor 2, which indicated higher proportions of Hispanic and Native American students, would have lower achievement scores (i.e. a negative coefficient).

Model 3 then adds the grouping variable – the indicator of attending first grade in a DI school or a Control School. It was expected that, net of the year data were gathered and the measures of school context, students in DI schools would have higher levels of achievement (a positive coefficient).

The -2 log likelihood measures and the BIC values in Panel B of Table A-5 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 bottom part of each section of Panel B of Table A-7. For example, Model 2 provides a significantly better fit to the data than Model 1 for all measures. For computations, the change in the -2 Log Likelihood = 5594 (417,306 -41712). With three degrees of freedom (because three new variables were added to Model 2 compared to Model 1), this result is highly significant. The comparisons of Model 3 with Model 2 and of Model 4 with Model 3 also indicate that adding the variables in the more complex model significantly improves the fit. (See the significance associated with the change in the -2 log likelihood.)

The BIC values provide a descriptive summary of the fit of the models, with lower values indicating a better fit. Looking at all the models in Table A-4 it may be seen that the lowest BIC values appear for Model 4. The Proportionate Reduction of Error (PRE) measures are another descriptive measure of the incremental changes in fit of the models and simply reflect the proportionate changes in the residual variance from one model to the next. The greatest changes occur from Model 1 to Model 2 and the proportionate reduction of error in prediction is substantially less for each of the more complex models. Again, this could be expected given the large number of factors that affect student achievement.[10]

Table A-6 gives the coefficients associated with each of the models in Table A-5, beginning with Model 2. As expected, the coefficients associated with Factor 1 (where higher scores indicate schools with fewer students on free and reduced lunch and more non-Hispanic whites) are positive and highly significant and the coefficients associated with Factor 2 (where higher scores indicate more Hispanic and Native American children) are negative and also statistically significant. The coefficient for year is positive and significant in all models, indicating that, over time, students' achievement scores increased.

The coefficients in Model 3 directly test the second research question by including both a dummy variable for treatment group and the controls for the measures of school context. It should be recalled that students were only coded as having the treatment if they were in a school in a year when the school was in a treatment condition. The coefficients associated with being in a DI school are positive and highly significant ($p < .0001$) in analyses of all three measures of

[10] It should be noted that the values of BIC, the changes in the Log-Likelihood Ratio, and the PRE measures would alter if the order in which variables were introduced were changed.

achievement. In other words, once the demographic context of the school is controlled, students who received Direct Instruction in mathematics had significantly higher achievement than students in the Control Schools.

Table A-6: Coefficients Associated with Mixed Model Regressions of First Grade Mathematics Achievement, BCPSS, 1998-2003

	Model 2			Model 3			Model 4		
Computations	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-9754.1	127.1	<.0001	-9568.6	129.1	<.0001	-9442.3	131.7	<.0001
Year	4.9	0.1	<.0001	4.8	0.1	<.0001	4.7	0.1	<.0001
Factor 1	1.6	0.6	0.005	1.6	0.6	0.005	1.7	0.6	0.004
Factor 2	-0.8	0.6	0.16	-1.7	0.6	0.004	-1.5	0.6	0.009
DI	-----	-----	-----	7.9	1.0	<.0001	-3090.9	653.0	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	1.5	0.3	<.0001

Concepts and Applications	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-4902.4	116.6	<.0001	-4768.7	118.4	<.0001	-4532.7	120.7	<.0001
Year	2.5	0.1	<.0001	2.4	0.1	<.0001	2.3	0.1	<.0001
Factor 1	2.6	0.6	<.0001	2.7	0.6	<.0001	2.7	0.6	<.0001
Factor 2	-1.0	0.6	0.06	-1.7	0.6	0.003	-1.4	0.6	0.01
DI	-----	-----	-----	5.7	0.9	<.0001	-5824.1	599.9	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	2.9	0.3	<.0001

Composite	b	s.e.	p.	B	s.e.	p.	b	s.e.	p.
Intercept	-9060.7	122.7	<.0001	-8873.5	124.6	<.0001	-8687.6	127.2	<.0001
Year	4.6	0.1	<.0001	4.5	0.1	<.0001	4.4	0.1	<.0001
Factor 1	2.2	0.6	0.0002	2.3	0.6	0.0001	2.3	0.6	0.0001
Factor 2	-1.0	0.6	0.09	-1.9	0.6	0.002	-1.7	0.6	0.005
DI	-----	-----	-----	8.1	1.0	<.0001	-4568.5	631.2	<.0001
DI* Year	-----	-----	-----	-----	-----	-----	2.3	0.32	<.0001

Table A-7 shows the predicted achievement scores for students in the DI and Control Schools based on the results in Model 3. These values were calculated by substituting average values for the Factor Scores and Year in the models and calculating predicted values for students in each group. Thus, the resulting values give the achievement score that would be predicted if students were equalized on the factor scores and time. Panel A gives the Normal Curve Equivalent scores and Panel B gives the percentile equivalent of these scores. Values in Panel B may be interpreted as the percentile at which an average student in each of group would be expected to achieve if schools were equal in their demographic characteristics. The values in Panel B were used in Figure 2 in the body of the text. The results in Table A-7 indicate the magnitude of the advantage that accrued to students in the DI Schools for they have higher average scores on all three measures and surpass the national mean on both the measure of computations and the composite score.

Table A-7: Predicted Achievement Scores Assuming Equal School Contexts

	DI	Control
Panel A: Norm Equivalent Scores		
Computation	50.8	42.8
Concepts and Applications	47.7	42.0
Composite	51.1	43.0
Panel B: Percentile Scores		
Computation	51.5	36.6
Concepts and Applications	45.6	35.2
Composite	52.0	37.0

Note: Scores were computed by substituting average values for the entire sample for Year and Factors 1 and 2 into the equations associated with Model 3 in Table A-6. The average value for year was 2000.3. The average value for Factor 1 was -.03203, the average value for Factor 2 was -.04252. The value for year does not equal 2000.5 (the average of 1998-2003) because there were fewer students in the BCPSS in later years than in earlier years. The average values for Factors 1 and 2 do not equal zero because the unit of analysis for computing the average values was students rather than schools. If the alternative values were used in computations, the substantive results would not differ.

Research Question 3: Did the impact of being in a DI School increase over time as reforms became institutionalized within a school?

Research question three tests the hypothesis that the advantage to students in a DI School would become stronger as practices and procedures were institutionalized within a school. As noted in the text, it takes time for teachers and other school staff to fully adjust to and incorporate the nuances of a new curriculum. Thus, it would be expected that the advantages accruing to attending a DI School would become greater over time.

Model 4 in Tables A-5 and A-6 addresses this hypothesis by adding the interaction of attending a DI school and year. It was expected that the coefficient associated with the interaction of year and DI would be positive, indicating stronger effects of Direct Instruction as years of treatment increased. This hypothesis was confirmed. In each analysis the interaction term is positive and highly significant.

The coefficients associated with Model 4 in Table A-7 can be used to calculate the average change expected in the three groups of schools from one year to the next. For instance, first grade students in the control schools had an average gain from one year to the next of 4.4 points in their normal curve equivalent scores on the composite measure of mathematics achievement (the coefficient associated with year), controlling for differences in the demographic context of the schools. In contrast, the scores of first grade students in the DI Schools had an average gain of 6.7 normal curve equivalent points with each additional year of implementation of Direct Instruction (summing the impact of year and the impact of years of implementation: 4.4 + 2.3).

Table A-8 gives the scores that would be predicted for students in the two groups of schools if they attended schools with similar demographic characteristics in 1998, the first year for which data are available, and 2003, the last year. These scores were obtained by using the regression coefficients in Table A-6 for Model 4 and substituting the average factor scores, as was done for the calculations reported in Table A-9, and either 1998 or 2003. Panel A gives the resulting normal curve equivalent scores and Panel B translates these scores into the corresponding percentiles. The scores in Panel B can be interpreted as the percentile that an average student in a group of schools would be expected to achieve at a given year.

Table A-8: First Grade Mathematics Achievement Scores by Group and Year, Adjusting for School Advantage, BCPSS

Panel A: Norm Equivalent Scores

	Year	DI	Control
Computations	1998	35.0	32.0
	2003	66.5	55.7
Concepts and Applications	1998	33.2	36.7
	2003	59.2	48.2
Composite	1998	33.8	33.1
	2003	67.1	54.9

Panel B: Expected Percentile Score of Average Student

	Year	DI	Control
Computations	1998	24	20
	2003	78	61
Concepts and Applications	1998	21	26
	2003	67	47
Composite	1998	22	21
	2003	79	59

Even after just one year of implementation, in 1998, the average scores of students in the DI schools, after adjustments for the socio-economic status of the schools, are higher than those of students in the other schools on the computation and composite measures. By 2003 these differences become much stronger and more dramatic. Assuming that they attended schools with similar socio-economic contexts, by 2003 the average student in a DI School would be expected to score between the 67thth and 79th percentile on the various measures, while the average student at Control School would score from the 47th to the 61st percentile, depending on the measure of achievement.

Table A-9 presents the data that were used to construct Figures 3, 4, and 5 in the text. Unlike data in Tables A-7 and A-8, these data are simply the average achievement scores in each year on each measure for each group. No adjustments were included for school advantage/disadvantage. They simply compare the average scores in the schools in the two groups from 1998 to 2003. The t-tests in each row test the null hypothesis that the mean NCE scores of students in the two groups are equal.

The results show that students in the DI schools had significantly lower scores than students in the control schools in the first two years for which data are available (1998 and 1999) for the concepts and applications and the composite measures and the first year (1998) for the computation measure. However, by the end of the study period the situation had totally reversed. Students in the DI Schools had significantly higher scores than students in the Control Schools on both the computation and composite measure from 2001 to 2003 and on the concepts and applications measure in 2003.

Another way to examine the change in scores is to calculate the percentage change in average NCE scores from 1998 to 2003. The last row of each section of Panel A gives these figures for each group and measure. The percentage change in average scores was much larger for students in the DI Schools than in the Control Schools. For instance, with the composite achievement measure, average normal curve equivalent scores for students in the DI Schools increased by 135 percent, while scores for students in the Control Schools increased by only 54 percent, less than half as much. Differences between the three groups were most marked in the measure of Concepts and Applications. Scores of students in the DI Schools increased by 87% over the 6 year period, while those of students in the Control Schools increased by only 26 percent. (These data were used to create Figure 6 in the text.)

Panel B of Table A-9 translates the Normal Curve Equivalent scores in Panel A into the corresponding percentiles. (These percentiles were used to create Figures 3 through 5 in the text.) The percentile scores confirm the pattern that appears in the other analyses. Students scores increased in both groups, but the increases were substantially stronger for students in the DI Schools. For instance, on the composite measure of mathematics achievement the average student in the DI Schools scored at the 13th percentile in 1998, but by 2003 the average student in these schools scored at the 68th percentile. In contrast, the average student in the Control Schools went from the 24th percentile in 1998 to the 56th percentile in 2003. In general, the raw, unadjusted scores confirm the more efficient and parsimonious results obtained with the mixed models, which incorporate the statistical controls for school context.

Table A-9: Mathematics Achievement Scores by Year and Group, First Graders, BCPSS, 1998-2003

Panel A: Means and t-test Results

	Group				
	DI	Control	t-value	df	prob.
Computations					
1998	25	33	-4.76	176	<.001
1999	30	32	-1.27	343	0.21
2000	45	46	-0.56	7700	0.58
2001	53	50	2.26	608	0.02
2002	57	52	4.57	570	<.001
2003	59	53	5.01	530	<.001
Percent Increase	134	60			
Concepts and Applications					
1998	29	38	-5.64	176	<.001
1999	32	37	-4.42	342	<.001
2000	40	41	-0.10	7851	0.92
2001	46	45	0.91	7352	0.36
2002	53	47	5.97	6805	<.001
2003	55	48	5.98	518	<.001
Percent Increase	87	26			
Composite					
1998	26	35	-5.71	174	<.001
1999	30	33	-3.09	334	0.002
2000	44	45	-0.51	7610	0.61
2001	52	50	1.93	7248	0.05
2002	58	52	5.58	568	<.001
2003	60	53	5.89	523	<.001
Percent Increase	135	54			

Panel B: Percentile Equivalents of Norm Equivalent Scores

	DI	Control
Computations		
1998	12	21
1999	17	20
2000	41	42
2001	56	50
2002	63	54
2003	66	56

Concepts and Applications		
1998	16	29
1999	20	27
2000	32	33
2001	42	41
2002	56	44
2003	60	46
Composite		
1998	13	24
1999	17	21
2000	39	41
2001	54	50
2002	65	54
2003	68	56

Note: In Panel A, the smaller degrees of freedom 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.

Research Question 4: What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?

The fourth and final question asked, "What was the magnitude of the effects that were found?" Cohen's d, a standard measure of effect size, was calculated to provide a descriptive measure of the extent to which students in the NIFDI supported schools had achievement scores that were superior to students in the other schools. Cohen's d is simply the difference between the average score of two groups divided by the common standard deviation. Traditionally, d scores of .25 or larger have been deemed educationally significant, although McLean and associates (2000) caution that effect sizes calculated with normal curve equivalent scores are conservative in nature.

Table A-10 gives the d scores comparing the average achievement of students in the DI and Control Schools**.** Effect sizes are given in Panel A for the unadjusted scores. Effect sizes in Panel B were calculated with school context controlled, based on the results in Tables A-7 and A-8. The standard deviations for the total group were used for all calculations, and all calculations were made with the norm equivalent scores.

With the unadjusted scores combined over all years, the effect sizes are positive, but below the .25 threshold. The effect sizes for 1998 are negative and surpass the .25 criterion, indicating that the students in the DI Schools had substantially lower achievement than students in the Control Schools at the beginning of the implementation. However, after 2000 all of the effect sizes are positive and by 2003 they all surpass the .25 threshold. In other words, while the d score at the beginning of the study period indicates an educationally significant disadvantage for students in the Direct Instruction Schools, this had totally reversed by the end of the study period and the students in the DI Schools had an educationally significant advantage.

Table A-10: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Mathematics Achievement, BCPSS, 1998-2003

Panel A: Gross Effect Sizes

	Computations	Concepts and Applications	Composite
Total	0.21	0.13	0.21
1998	-0.33	-0.38	-0.39
1999	-0.07	-0.22	-0.16
2000	-0.03	0.00	-0.02
2001	0.11	0.04	0.09
2002	0.23	0.28	0.27
2003	0.25	0.32	0.31

Panel B: Effect Sizes Independent of School Context

	Computations	Concepts and Applications	Composite
Total	0.32	0.26	0.33
1998	0.12	-0.15	0.03
2003	0.48	0.52	0.54

Similar, but somewhat stronger, results occur with the measures derived from the results where school context was statistically controlled (see Panel B). These effect sizes far surpass the threshold of educational significance by the 7^{th} year of implementation, ranging from .48 to .54.

Thus, the results with the effect sizes parallel the results with the other statistical analyses. Students in DI schools had higher achievement than students in the Control Schools, and the magnitude of this advantage increased substantially over the years of implementation.

Appendix III
Panel Sample: Changes from First to Fifth Grade

This appendix provides details on the sample used in the analysis of changes in achievement from first to fifth grade. 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. Thus, data for this study included student cohorts who began first grade in the BCPSS in 1997-98 or 1998-99 and were in the same schools in fifth grade in either 2001-02 (for the first graders in 1997-98) or 2002-03 (for the first graders in 1998-99). Students who were held back were not included.

There was substantial student turnover during the study period, as summarized in Table A-4. There were almost 18,000 students in first grade in the BCPSS in 1997-98 and 1998-99. Five years later over half of these students (52.8%) had left the BCPSS, and an additional 20 percent had transferred to another school within the system. Only 27 percent of the students who were in first grade in the system in 1997-98 and 1998-99 were still in the same school in 5^{th} grade in 2001-2 or 2002-3. The panel sample included a total of 4,771 students. Of these, 164 were in DI Schools and the remaining 4,607 were in the Control Schools.

Table A-11: Turnover in the BCPSS, 1997-98, 1998-99 to 2001-02, 2002-03

	N	Percent
Left BCPSS (In first grade, but not in fifth)	9362	52.80%
Transferred within BCPSS (In BCPSS in fifth grade, but in a different school)	3595	20.30%
Panel Sample (Same School, first and fifth grade)	4771	26.90%
Total Number of Students In First Grade 1997-98 or 1998-99	17,728	100.00%

Students were considered to have received an intervention only if they could be considered to have begun receiving Direct Instruction in first grade.[11] As summarized in Table A-12, this limitation of the sample resulted in students in several of the schools listed in Table A-1 not being included within the intervention group. For example, students who were fifth graders

[11] Analyses were also conducted that considered students who began DI instruction in second grade as part of the implementation group as well as at any time in their elementary career. The results with these comparisons were substantively similar to those reported here. It was decided, however, to limit the intervention sample only to those who had received DI throughout their first 5 elementary years to provide the least contaminated test possible of the effect of the curriculum.

in 2001-02 or 2002-03 at Collington Square would not have had Direct Instruction in first grade (the 1997-98 or 1998-99 school year) and thus were included in the control condition. Students who were fifth graders in 2001-02 in City Springs, Barrister, and Federal Hill did not have Direct Instruction in mathematics in first grade and were included in the control group. But, students who were fifth graders in 2002-03 in these schools did have first grade Direct Instruction and were included in the treatment group.

Table A-12: Schools in Treatment Group by Year for Panel Analysis

School	Year Started DI in Math	Fifth Graders in 2001-02	Fifth Graders in 2002-03	N in Panel Sample
Arundel	97 – 98	yes	yes	37
General Wolfe	97 – 98	yes	yes	23
Hampstead Hill	97 – 98	yes	yes	55
CC Barrister	98 – 99	no	yes	23
City Springs	98 – 99	no	yes	16
Federal Hill	98 – 99	no	yes	10
Collington Square	00 – 01	no	no	
Dickey Hill	99 – 00	no	no	
Langston Hughes	99 – 00	no	no	
Margaret Brent	99 – 00	no	no	
Dr. Rayner Browne	99 – 00	no	no	

The pattern of persistence within the same school differed somewhat between the groups of schools, as shown in Table A-13, with the Direct Instruction Schools having a higher proportion of their students remaining within the same schools from first to fifth grade[12] About a third of the students in first grade remained in the same schools in fifth grade, compared to slightly over one-fourth of those in the Control Schools. These differences were statistically significant.

As would be expected the students in the panel group (those that were in the same schools in 1st and 5th grade) differed from other students in their achievement levels in first grade. As shown in Table A-14, students who persisted in the schools in which they attended first grade had significantly higher first grade mathematics achievement on all three scores. These differences appeared for both the Direct Instruction Schools and the Control Schools, but were slightly stronger for the Control Schools than for the DI Schools (see the marginally significant interaction effects for concepts and applications and composite scores).

[12] There was no way to control for the possibility that a student had attended multiple other schools between first and fifth grade. It is possible that some of the students in the panel sample were in the targeted schools in both first and fifth grade, but had attended other schools in the interim. Assuming that such children would have less of the "full treatment," this would bias results in a conservative direction.

Table A-13: Persistence in the Same School, by School Group

	DI	Control Schools	Total
Not in Panel Sample	66.8	73.3	73.1
In Panel Sample	33.2	26.7	26.9
Total (%)	100	100	100
N (first grade)	494	17234	17728
N (panel)	164	4607	4771

chi-square = 10.209, df = 2, p = .001

The results in Table A-14 also indicate that the students in the DI Schools had substantially lower average levels of first grade achievement on each of the measures than students in the Control Schools. These differences were highly significant.

Table A-14: Comparison of Achievement Scores of Panel Group and Other First Graders, DI and Control Schools

	Computations		Concepts and Applications		Composite	
DI Schools	Mean	s.d.	Mean	s.d.	Mean	s.d.
Non-Panel Sample	26.2	22.9	28.8	19.5	25.7	20.1
Panel Sample	33.3	21.9	36.2	19.0	33.0	20.1
Control Schools						
Non-Panel Sample	29.9	24.1	34.7	22.8	30.9	22.7
Panel Sample	40.2	25.1	46.2	22.2	42.2	23.0
ANOVA Results	F	p	F	p	F	p
Panel vs. Non-Panel	53.16	<.001	71.66	<.001	68.34	<.001
Treatment Group	20.14	<.001	50.89	<.001	40.49	<.001
Interaction	1.79	0.18	3.25	0.07	3.21	0.07

Note: Mean and standard deviation values in the table are computed from Normal Curve Equivalent (NCE) Scores.

In order to minimize the differences between the Control Schools and the DI schools the sample of schools included within the control group was also limited to produce a subsample that had a mean as close as possible to that of the DI schools. This was accomplished by rank ordering all schools by their average composite mathematics achievement score and then selecting the subsample of schools (beginning with the lowest) that produced an average as close as possible to that of the Direct Instruction schools. The resulting sample includes the 164 students in the DI Schools, as noted in Tables A-12 through A-14, and 1,820 students in the Control Schools. The sample includes 50 different schools. The minimum number of cases from a school is 16, and the maximum is 101. There are no significant differences in first grade achievement between these two groups. Results with both groups of Control Schools are presented in Appendix IV and in the body of the text.

Appendix IV: Detailed Statistical Results – Fifth Grade Achievement

This appendix expands upon the material presented in the text, including a more extended discussion of the analysis techniques employed and the results related to each of the three research questions regarding the impact of DI instruction in first grade on mathematics achievement in fifth grade.

Analysis Methodology

The analysis focused on changes in students' mathematics achievement from first to fifth grade and the extent to which students in schools with Direct Instruction had greater gains over time. Three research questions were addressed. The first was, "Do students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools?" Several comparisons were used to examine this question. Simple paired t-tests were used to examine changes from first to fifth grade in both sets of schools. Second, independent sample t-tests were conducted to compare the average achievement scores of students in the DI and Control schools in both first and fifth grade. Third, a repeated measures analysis of variance was conducted with the first and fifth grade scores as repeated measures and the school as a factor. If gains were greater in the DI schools, the interaction between the repeated measures would be expected to be significant.

The second research question was, "Do students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school is controlled?" This question was addressed by employing a mixed model regression, having schools as a random variable and regressing students' fifth grade scores on first grade scores, the measures of a school context of advantage/ disadvantage, and school type (DI or Control). Successively more complex models were used and differences between these models were examined.

The third research question asked, "What is the magnitude of the effects of being in a school with Direct Instruction on changes in reading achievement from first to fifth grade?" To answer this question, the results were translated into effect sizes. Unlike tests of significance, effect sizes are not influenced by the size of a sample. Because the issue of interest in this report is 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 and associates (1996) is used. The calculations involve an adjustment for the correlation between the two scores. Results may be interpreted in the same manner as Cohen's d. Effect sizes greater than .25 have traditionally been considered educationally significant (Fashola and Slavin 1990). However, McLean and associates (2000) caution that effect sizes calculated with norm equivalent scores, as with this analysis, are inherently smaller than with other metrics. Thus the effect sizes presented are a conservative estimate of the results.

For all analyses results are presented comparing the students in the Direct Instruction Schools with students in all the Control Schools as well as those in the schools with more similar levels of achievement in first grade. The latter group is referred to as the "reduced sample" of control schools.

Research Question One: Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Mathematics Achievement from First to Fifth Grade than Students in Other Schools?

Three different approaches were used to address the first research question: paired t-tests, independent sample t-tests, and repeated measures analysis of variance.

Difference Scores and Paired t-tests

Table A-15 summarizes the data used to examine differences in achievement gains from first to fifth grade for students in the Direct Instruction and Control Schools using paired t-tests. Panel A gives descriptive statistics for each measure of mathematics achievement using norm equivalent scores. Panel B gives the average change from first to fifth grade, the correlations of first and fifth grade scores, and results of paired t-tests that examine the null hypothesis that there was no change, on average, in achievement from one grade to the other. Panel C translates the averages in Panel A into percentile scores. The data in Panel C can be interpreted as the percentile score that the average student obtained in the two groups of schools in first grade and fifth grade. The data in Panel C were used to construct Figure 9 in the text.

Table A-15: Descriptive statistics and Paired t-test Results, 1st and 5th Grade Mathematics Achievement Norm Equivalent Scores, DI and Control Schools, Full and Reduced Samples, BCPSS, 1998-2003

A: Descriptive Statistics (Norm Equivalent Scores)

		DI Schools Full Sample		**Control Schools Full Sample**		**Control Schools Reduced Sample**	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
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	
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	
Composite	1st.	32.8	20.2	42.3	22.9	33.3	19.9
	5th	48.5	17.8	50.6	20.3	47.9	19.9
	% ch.	0.48		0.20		0.44	

B: Average Differences, Correlations, paired (dependent sample) t-scores

	Ave. Dif.	r	t	p	N
DI Schools					
Computations	17.1	0.45	9.85	<.001	159
Concepts and Applications	10.9	0.46	7.44	<.001	159
Composite	15.7	0.54	10.58	<.001	155

Control Schools Full Sample					
Computations	13.1	0.38	33.26	<.001	4495
Concepts and Applications	1.3	0.53	4.59	<.001	4498
Composite	8.3	0.52	25.80	<.001	4437
Reduced Sample					
Control Schools Reduced Sample					
Computations	20.4	0.34	33.94	<.001	1755
Concepts and Applications	6.1	0.53	14.06	<.001	1756
Composite	14.6	0.49	3.02	<.001	1726

Results indicate that students in both the DI and the Control Schools had significantly higher achievement from in fifth grade than in first grade. For both groups and each measure the average change in norm equivalent scores from first to fifth grade was statistically significant. Changes on all three scores were larger for students in the DI group than for those in the full sample of Control Schools. Changes for students in the DI group were larger than those for students in the reduced sample of Control Schools on both the measure of Concepts and Applications and the Composite Score. This can be seen in the average differences reported in Panel B as well as the percentage gain reported in Panel A. The largest differences between students in the DI Schools and those in the Control Schools occurred with the measure of concepts and applications. On average students in the DI schools had concepts and application scores that were 30 percent higher in fifth grade than in first grade compared to an increase of only 3 percent for students in the total group of Control Schools and 16 percent for those in the reduced sample of Control Schools. These data were used to create Figures 9 and 10 in the text.

Independent t-tests

Table A-16 provides an alternative way to examine the data by looking at the difference in average achievement scores of students in the DI schools and Control schools in first grade and in fifth grade. While the computations reported in Table A-15 compared the average difference between students' first and fifth grade scores, the computations in Table A-16 compare average scores of students in DI and Control Schools in first grade and in fifth grade.

Panel A includes comparisons of students in the DI Schools with those in the full sample of Control Schools. Results indicate that scores of those in the Control Schools were significantly higher than those in the DI schools on all measures in first grade but only on the measure of computations in the fifth grade. Panel B includes comparisons with the reduced sample of Control Schools. None of the comparisons in Panel B are statistically significant, although the results for Concepts and Applications in fifth grade approaches significance (p = .12).

Table A-16: Independent t-tests, Average Mathematics Achievement, DI and Control Schools (Full and Reduced Samples), 1st and 5th Grade, BCPSS, 1998-2003

Panel A: DI Schools vs. Full Sample Control Schools

First Grade	t	df	p
Computations	3.94	177	<.001
Concepts & Apps.	6.46	174	<.001
Composite	5.625	172	<.001
Fifth Grade			
Computations	2.01	174	0.05
Concepts & Apps.	0.37	4756	0.71
Composite	1.42	174	0.16

Panel B: DI Schools vs. Reduced Sample Control Schools

First Grade	t	df	p
Computations	-1.44	1934	0.15
Concepts & Apps.	1.46	1920	0.15
Composite	0.09	186	0.93
Fifth Grade			
Computations	0.53	197	0.60
Concepts & Apps.	-1.56	1974	0.12
Composite	-0.42	1957	0.68

Note: F tests indicated that the t-value assuming unequal variances between the groups should be used for all comparisons with the total group of control schools but the one for concepts and applications for fifth grade. The t-value assuming unequal variances was used with comparisons with the reduced sample for the composite score for first grade and the computations score for fifth grade. Negative t values indicate that students in the DI schools had higher average scores than students in the Control schools. Positive values indicate that the Control school students had higher scores.

Repeated Measures ANOVA

Table A-17 gives the results of a repeated measures analysis of variance with the first and fifth grade scores as a repeated measure and the type of school (DI or Control) as a factor. This approach is more parsimonious than the separate t-tests reported in Tables A-15 and A-16. Panel A gives results with comparisons with the total set of Control Schools and Panel B gives results with the reduced sample of Control Schools. If exposure to Direct Instruction in the first grade is related to higher achievement in fifth grade, a significant interaction effect would be expected. With comparisons with the total sample of Control Schools (Panel A) the results are as expected with the measure of concepts and applications and the composite score. With comparisons with the reduced sample, the results are as expected only with the measure of concepts and applications.

Table A-17: Repeated Measures Analysis of Variance, Norm Equivalent Mathematics Achievement Scores, 1st and 5th Grade, DI Schools versus Control Schools

Panel A: DI Schools versus Full Set of Control Schools

	Computations	Concepts and Applications	Composite
F-Mathematics	203.16	60.02	189.09
P	<.001	<.001	<.001
F- Group	10.76	13.94	14.32
P	0.001	<.001	<.001
Interaction	3.59	36.56	17.87
P	0.06	<.001	<.001
Df	1,4652	1, 4655	1, 4590

Panel B: DI Schools versus Reduced Set of Control Schools

	Computations	Concepts and Applications	Composite
F-Mathematics	329.69	126.76	327.31
P	<.001	<.001	<.001
F- Group	0.20	0.004	0.003
P	0.66	0.95	0.95
Interaction	2.59	10.03	0.43
P	0.11	0.002	0.51
Df	1, 1912	1,1913	1,1879

Note: The F test associated with mathematics tests the null hypothesis that the change from first grade to fifth grade equals zero. The F test associated with group tests the null hypothesis that the average mathematics score of students in the DI schools. The F test associated with the interaction tests the null hypothesis that the differences between groups are the same for each grade or, alternatively, that the difference between the two grades is the same for each group.

Research Question Two: Did students who received Direct Instruction in first grade have greater gains in mathematics achievement from first to fifth grade than students in other schools when the socio-demographic context of the school and first grade achievement are controlled?

Tables A-18 and A-19 report the results of the mixed model regressions. Panel A of Table A-18 summarizes the models that were used and Panels B and C give summary statistics associated with each model for the full set of Control Schools (Panel B) and the reduced set (Panel C). The models become increasingly more complex. Model 1 is the baseline “intercept only” or “random effects” model and only includes schools as a random variable. This tests the null hypothesis that the schools are equal in average mathematics achievement in fifth grade. The correlation ratio attached to Model 1 is the proportion of variance in the dependent variable that is between schools as opposed to between students. It can be seen that between 9 and 12 percent of the variance in fifth grade 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 three models. There is significant variation between schools and also between students even when the explanatory variables are included. This is as we would expect, for there are undoubtedly many factors that can influence student achievement in addition to those in this analysis.

The models become incrementally more complex, with each subsequent model adding more explanatory variables to test the research questions, as indicated by the "x's" associated with each model in Panel A. Model 2 adds the norm equivalent reading achievement score for year one and the two factor scores measuring the advantage/disadvantage of the schools the students attended. It was expected that students' first grade achievement would be positively and significantly associated with their achievement in fifth grade. Including the two factor scores is important to control for the extent to which the demographic context of a school affects student achievement. It was expected that Factor 1, with higher loadings for schools with less poverty, fewer African American students and more non-Hispanic white students, would be positively associated with achievement. It was expected that students in schools with higher scores on Factor 2, which indicated higher proportions of Hispanic and Native American students, would have lower achievement scores (i.e. a negative coefficient).

Model 3 adds the treatment group, testing the hypothesis that students in DI Schools have higher achievement than students in the Control Schools when first grade achievement and school advantage/disadvantage are controlled. If having Direct Instruction in first grade influences students' fifth grade achievement, independent of their first grade achievement and school context, we would expect positive associations with this variable. It should be recalled that students were only coded as having Direct Instruction if they were in a school in a year when the school was in a treatment condition.

Table A-18, Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Mathematics Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction

Panel A: Variables in the Models

Variables	Model 1	Model 2	Model 3
Random intercept for Schools	x	x	x
First Grade Achievement		x	x
Factors 1 and 2		x	x
DI in First Grade			x

Panel B: Model Fit Statistics Comparing DI Schools and Full Sample of Control Schools

	Model 1			Model 2			Model 3		
Computations	Effect	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Random Effects Estimate	54.8	9.0	<.0001	51.4	8.4	<.0001	51.4	8.4	<.0001
Residual Estimate	434.6	9.0	<.0001	363.3	7.6	<.0001	363.2	7.6	<.0001
BIC	42452			40873			40877		
- 2 Log Likelihood	42438			40845			40845		
Change in LL				1593			0.0		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.11								
Concepts and Applications									
Random Effects Estimate	35.1	5.8	<.0001	34.5	5.5	<.0001	34.7	5.5	<.0001
Residual Estimate	267.7	5.6	<.0001	181.3	3.8	<.0001	181.0	3.8	<.0001
BIC	40303			37690			37686		
- 2 Log Likelihood	40297			37662			37653		
Change in LL				2634			9.0		
df				3			1		
p				<.001			<.01		
Correlation ratio	0.12								
Composite									
Random Effects Estimate	49.8	8.1	<.0001	53.2	8.3	<.0001	53.2	8.3	<.0001
Residual Estimate	361.6	7.5	<.0001	247.3	5.2	<.0001	247.2	5.2	<.0001
BIC	41535			38603			38606		
- 2 Log Likelihood	41522			38575			38573		
Change in LL				2947			1.6		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.12								

Panel C: Model Fit Statistics Comparing DI Schools and Reduced Sample of Control Schools

	Model 1			Model 2			Model 3		
Computations	Effect	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Random Effects Estimate	53.2	12.8	<.0001	57.4	13.5	<.0001	57.4	13.5	<.0001
Residual Estimate	422.0	13.6	<.0001	361.6	11.8	<.0001	361.6	11.8	<.0001
BIC	17527			16825			16829		
- 2 Log Likelihood	17515			16801			16801		
Change in LL				714			0		
df				3			1		
p				<.001			n.s.		

Correlation ratio	0.11								
Concepts and Applications									
Random Effects Estimate	26.4	6.7	<.0001	36.7	8.3	<.0001	38.3	8.7	<.0001
Residual Estimate	262.5	8.5	<.0001	168.8	5.5	<.0001	167.9	5.5	<.0001
BIC	16704			15389			15384		
- 2 Log Likelihood	16692			15365			15357		
Change in LL				1327			8.5		
df				3			1		
p				<.001			<.01		
Correlation ratio	0.09								
Composite									
Random Effects Estimate	42.7	10.4	<.0001	57.9	13.0	<.0001	58.6	13.2	<.0001
Residual Estimate	350.4	11.3	<.0001	237.6	7.9	<.0001	237.4	7.8	<.0001
BIC	17134			15764			15767		
- 2 Log Likelihood	17123			15741			15740		
Change in LL				1382			1.1		
df				3			1		
p				<.001			n.s.		
Correlation ratio	0.11								

The -2 log likelihood measures and the BIC values in Panels B and C of Table A-18 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 also included. For both sets of comparisons the results indicate that Model 2 is the best fitting model for the analyses of computations and the composite scores, but Model 3 is the best fitting for the analysis of concepts and applications.

Table A-19 gives the coefficients associated with variables in Models 2 and Model 3 for each of the measures of achievement and for both comparisons. The coefficients associated with first grade achievement are positive and highly significant in all models for all three dependent variables, indicating that, as expected, students with higher mathematics achievement scores in first grade also have higher mathematics achievement scores in fifth grade. As expected, the coefficients associated with factor 1 are positive and statistically significant with all analyses for the total group of Control Schools, indicating that students have higher achievement scores in schools with more non-Hispanic white children and fewer African American children and fewer children receiving free or reduced lunch. Associations with Factor 2 are insignificant in all analyses. Coefficients associated with Factor 1 are not significant for the analyses in Panel B, perhaps because the variation in the factor is somewhat smaller in this group.

Table A-19: Coefficients Associated with Mixed Model Results, Fifth Grade Mathematics Achievement Regressed on First Grade Achievement, School Advantage/Disadvantage, and Attending a DI School

Panel A: Coefficients for Comparisons of DI Schools and Total Sample of Control Schools

	Model 2			Model 3		
Computations	B	s.e.	p.	b	s.e.	p.
Intercept	38.5	0.9	<.0001	38.5	0.9	<.0001
First grade score	0.4	0.01	<.0001	0.4	0.01	<.0001
Factor 1	1.4	0.7	0.055	1.4	0.7	0.06
Factor 2	-0.2	0.7	0.795	-0.3	0.8	0.76
Had DI in First Grade	------	------	------	0.5	3.2	0.87
Concepts and Applications						
Intercept	27.1	0.7	<.0001	26.9	0.8	<.0001
First grade score	0.4	0.01	<.0001	0.44	0.01	<.0001
Factor 1	1.7	0.6	0.004	1.7	0.6	0.004
Factor 2	0.1	0.6	0.90	-0.7	0.6	0.25
Had DI in First Grade	------	------	------	7.0	2.3	0.003
Composite						
Intercept	29.5	0.9	<.0001	29.4	0.9	<.0001
First grade score	0.5	0.01	<.0001	0.5	0.01	<.0001
Factor 1	1.5	0.7	0.03	1.5	0.7	0.03
Factor 2	0.1	0.7	0.89	-0.31	0.8	0.70
Had DI in First Grade	------	------	------	3.6	2.8	0.20

Panel B: Coefficients for Comparisons of DI Schools and Reduced Sample of Control Schools

	Model 2			Model 3		
Computations	b	s.e.	p.	b	s.e.	p.
Intercept	39.4	1.4	<.0001	39.4	1.4	<.0001
First grade score	0.4	0.02	<.0001	0.4	0.02	<.0001
Factor 1	-0.04	1.3	0.97	-0.04	1.3	0.97
Factor 2	-0.2	0.8	0.79	-0.2	0.9	0.80
Had DI in First Grade	------	------	------	0.1	3.3	0.97
Concepts and Applications						
Intercept	26.5	1.1	<.0001	26.1	1.1	<.0001
First grade score	0.5	0.02	<.0001	0.47	0.02	<.0001
Factor 1	1.0	1.0	0.32	0.9	1.0	0.35
Factor 2	0.3	0.7	0.69	-0.6	0.7	0.43
Had DI in First Grade	------	------	------	6.8	2.3	0.004
Composite						
Intercept	29.9	1.3	<.0001	29.7	1.3	<.0001
First grade score	0.5	0.02	<.0001	0.5	0.02	<.0001
Factor 1	0.4	1.2	0.76	0.4	1.2	0.77
Factor 2	0.1	0.8	0.91	-0.3	0.9	0.75
Had DI in First Grade	------	------	------	3.0	2.8	0.29

The coefficients associated with attending a DI school in first grade are positive in all analyses, indicating that students in the DI schools have higher fifth grade achievement when other variables are equalized. However, the coefficients are significant only in the analyses of the measure of concepts and applications. These results appear both in the comparison with the total group of Control schools and with the reduced sample. As expected, students who attended a DI school in first grade had significantly higher scores in the measure of concepts and applications than other children, even when the socio-demographic context of the school and their first grade achievement scores were controlled.

Table A-20 summarizes the impact of these effects assuming that students had the same first grade reading score and that all schools were equivalent in their socio-economic characteristics. Average scores for the total group were substituted into the regression equations for Model 3 in Table A-19 for these calculations.[13] Results are given both in the norm equivalent

[13] The average value for the total group for computations was 39.97; for concepts and applications, 45.81; and for the composite score, 41.9. The average value for factor 1 was .1495 and for factor 2 was -.0497. These values do not equal zero for two reasons. The unit of analysis for these computations is students, rather than schools; and the population of students in the panel sample, as described above, is more advantaged than the general population and thus comes from schools with higher scores on Factor 1 and lower scores on Factor 2 than did the original sample of students in first grade.

scores and as the corresponding percentile score for a student scoring at this average. It can be seen, with both sets of comparisons, that, if they were to have equal scores at first grade and to attend schools with similar socio-demographic characteristics, students in the Direct Instruction schools and the Control Schools would be expected to have almost identical computation scores. However, students in the Direct Instruction school would have scores on the composite and concepts and applications measure that were higher than those of students in the Control Schools. The difference in the concepts and applications measure is the largest. With calculations based on the entire population of students, an average student in a DI school would score at the 58th percentile on the measure of concepts and applications and the composite score, well above the national average, while an average student in the Control Schools would only score at the 44th percentile, well below the national average. If calculations are based on the analysis that used the reduced sample of Control schools, the average student in the DI Schools would score at the 51st percentile on the measure of concepts and applications and at the 50th percentile on the composite measure, while the average student in the control schools would be at the 39th and 44th percentile on these measures.

Table A-20: Simulated Fifth Grade Mathematics Achievement Scores (Norm Equivalent and Percentile of Hypothetical Average Student), DI and Control Schools, Controlling for School Context and First Grade Achievement, BCPSS, 2001-03

	Full Sample of Control Schools		Reduced Sample of Control Schools	
A. Norm Equivalent Scores	DI Schools	Control Schools	DI Schools	Control Schools
Computations	53.5	53.0	50.7	50.5
Concepts and Applications	54.1	47.1	50.7	43.9
Composite	53.8	50.2	50.0	47.0
B. Corresponding Percentiles				
Computations	56.6	55.7	51.3	51.0
Concepts and Applications	57.8	44.5	51.3	38.6
Composite	57.2	50.4	50.0	44.3

Note: Scores were calculated by substituting the average values of first grade achievement and factor scores in the equations for Model 3 in Table A-11. For the sample that includes all the Control Schools the mean of Factor 1 was 1.4 and the mean for Factor 2 was -.3. For the sample with the reduced set of Control Schools the mean of Factor 1 was -.08 and the mean for Factor 2 was .23.

Research Question Three. What is the magnitude of the effects of being in a school with Direct Instruction on changes in mathematics achievement from first to fifth grade?

Because the measures of first and fifth grade achievement are statistically dependent, Cohen's d, the most often used measure of effect size, is not appropriate. To control for the dependence of these measures, effect sizes were calculated adjusting for the correlation between scores in the two years (Dunlap et al 1996). It was anticipated that effect sizes would be larger for the students who received Direct Instruction in the first grade than for other students. The effect sizes were calculated using the descriptive statistics from the norm equivalent scores. McLean and associates, 2000, have demonstrated that effect sizes for NCE scores are lower than those that would be obtained for equivalent raw and scaled scores. Thus, the calculations presented here should be seen as conservative estimates.

Table A-21 summarizes effect scores for the change in achievement from first grade to fifth grade on each of the measures. Using the criterion of .25 as an educationally significant effect size, it may be seen that all of the effect sizes associated with the DI schools clearly surpass that criterion as do several for the other groups. Both sets of Control Schools had educationally significant changes on the measure of computations, as did the full sample of Control Schools on the composite measure and the reduced set of Control Schools on the concepts and applications measure. The largest differences between the groups occur with the measure of concepts and applications and the composite measure, where the effect sizes are substantially larger for the DI schools. The very small effect sizes for the concepts and applications measure for the full set of control schools and for the composite measure for the reduced sample of control schools reflect the much lower t-values for those difference scores.

Table A-21: Effect Sizes of Change, First to Fifth Grade, by Group, Mathematics Achievement Norm Equivalent Scores

	Computations	Concepts and Applications	Composite
DI Schools	0.82	0.61	0.82
Control Schools Full Sample	0.55	0.07	0.38
Control Schools Redeuced Sample	0.93	0.32	0.07

Note: All effect sizes are based on norm equivalent scores.

References

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

MacIver and Kempner. 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.

McLean, James E., Marcia R. O'Neal, and J. Jackson Barnette. 2000. Are All Effect Sizes Created Equal? Paper presented at the 28th Annual Meeting of the Mid-South Educational Research Association, November, 2000, ED 448 188.

Stockard, Jean and Maralee Mayberry. 1992. *Effective Educational Environments.* Newbury Park, California: Corwin/Sage, 1992.