Improving First Grade Reading 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 2008-1

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

Table of Contents

	Page
Executive Summary	iv
List of Tables	v
List of Figures	vi
Report	1
Methodology	1
Results	3
1. Did students who received Direct Instruction have higher achievement scores than students in the other schools?	3
2. Did students who received Direct Instruction in schools with consistent NIFDI technical support have higher achievement scores than students in the other schools?	4
3. Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools were equalized?	5
4. Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?	6
5. What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?	10
Summary	12
Appendix I: Sample and Measures	14
Sample	14
Measuring Achievement	15
Measuring School Context	15
Appendix II: Detailed Statistical Results	19
Methodology	19

Research Question 1: Did students who received Direct Instruction have higher achievement scores than students in the other schools? 20

Research Question 2:"Did students who received Direct Instruction in schools with consistent NIFDI technical support have higher achievement scores than students in the other groups?" 22

Research Question 3: Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools are equalized? 23

Research Question 4: Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school? 30

Research Question 5: What is the Magnitude of the Effect of NIFDI-supported Direct Instruction? 35

References 38

Improving First Grade Reading Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in the Baltimore City Public School System
Technical Report 2008-1
Executive Summary

In the mid 1990s, in response to very low reading achievement scores, the Baltimore City Public School System (BCPSS) implemented curricular reforms. Sixteen schools used Direct Instruction. All the Direct Instruction schools used the *Reading Mastery Classic, Language for Learning, Language for Thinking,* and *Reasoning and Writing* curricula. Eleven of these schools received technical support from the National Institute for Direct Instruction throughout the time period of this study, while the others ceased support or used an alternative provider. The Comprehensive Test of Basic Skills (CTBS) was administered to all first graders in the spring of 6 school years: 1997-1998 through 2002-2003. Data were obtained from over 40,000 students on reading vocabulary, comprehension, and a composite score.

The study compared students' achievement in 1) schools with NIFDI support, 2) schools that implemented Direct Instruction without NIFDI support (other DI schools), and 3) schools without Direct Instruction (the Control schools). Schools in the three groups had similar demographic characteristics, although one of the Other DI schools had a markedly lower rate of poverty. Key findings are listed below.

- Combining data across all years (1998-2003) students in schools that implemented Direct Instruction had significantly higher achievement than students in other schools.
- Combining data across all years (1998-2003) students in NIFDI-supported schools had significantly higher achievement than students in the Control schools on all three measures of achievement and significantly higher comprehension scores than students in the Other DI schools.
- When the socio-economic characteristics of the schools were controlled, the students in the NIFDI-supported schools had significantly higher achievement scores than students in both the Other DI schools and the control schools on all measures of achievement.
- First grade students in all 3 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 reading achievement scores in the NIFDI-supported schools increased by 113 percent from 1998 to 2003, while those in the control schools and the Other DI schools increased by 56 percent or less.
- The magnitude of the effect on first grade achievement from attending a NIFDI-supported school was statistically significant and substantively large. An effect size of .25 has traditionally been considered educationally important. At the end of the study period (2003) the effect of attending a NIFDI-supported school versus attending an Other DI school on composite achievement was .63. The effect of attending a NIFDI-supported school versus a Control school was .82.

List of Tables

Title	Page
A-1. Schools in Treatment Groups and Start Dates of Treatment	14
A-2. Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels, Total and by Treatment Group	16
A-3. Factor Analysis of School Characteristics, Elementary Schools, BCPSS, 1998-2003	17
A-4. Average Reading Achievement Scores of Students with and without Direct Instruction, All Years, First Graders, BCPSS	21
A-5. Average Reading Achievement Scores by Group, All Years, First Graders	22
A-6. Average Reading Achievement Scores by Group, All Years, First Graders, BCPSS, Four Groups	23
A-7: Variables in Models and Measures of Fit, Analysis of First Grade Reading Achievement, BCPSS, 1998-2003	25
A-8: Coefficients Associated with Mixed Model Regressions of First Grade Reading Achievement Scores on School and Treatment Factors, BCPSS, 1998-2003	29
A-9: Predicted Achievement Scores Assuming Equal School Contexts	30
A-10 First Grade Reading Achievement Scores by Group and Years of Implementation, Adjusting for School Advantage, BCPSS	32
A-11: Reading Achievement Scores by Year and Group, First Graders, BCPSS, 1998-2003	33
A-12: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Reading Achievement, BCPSS, 1998-2003	36

List of Figures

Page

1. Vocabulary, Comprehension, and Composite Reading Percentile Achievement Scores of an Average First Grade Student, DI and Control Schools, BCPSS, All Years 3

2. Vocabulary, Comprehension, and Composite Reading Percentile Achievement Scores of an Average First Grade Student, NIFDI-Supported, Other DI, and Control Schools, BCPSS, All Years 4

3. Reading Percentile Achievement Scores of an Average First Grade Student, Adjusting for School Context, NIFDI-Supported, Other DI, and Control Schools, BCPSS, All Years 5

4. Vocabulary, Comprehension, and Composite Reading Percentile Achievement Scores of an Average First Grade Student, BCPSS, 1998-2003 6

5. Composite Reading Achievement Percentile Score of an Average First Grade Student, NIFDI-Supported, Other DI, and Control Schools, BCPSS, First Year and Seventh Year of Implementation, Adjusting for School SES 7

6. Reading Comprehension Percentile Scores of an Average First Grade Student, NIFDI-Supported, High Poverty Other DI, Low Poverty Other DI, and Control Schools, BCPSS, 1998-2003 8

7. Vocabulary Percentile Scores of an Average First Grade Student, NIFDI-Supported, High Poverty Other DI, Low Poverty Other DI, and Control Schools, BCPSS, 1998-2003 8

8. Composite Percentile Scores of an Average First Grade Student, NIFDI-Supported, High Poverty Other DI, Low Poverty Other DI, and Control Schools, BCPSS, 1998-2003 9

9. Percentage Change in Achievement of an Average First Grade Student, NIFDI-Supported, High Poverty Other DI, Low Poverty Other DI, and Control Schools, BCPSS, 1998-2003 10

10. Effect Sizes, Comparisons of Composite Reading Achievement Scores of First Grade Students in NIFDI-Supported, Other DI, and Control Schools, BCPSS, All Years and 2003 11

11. Effect Sizes, Comparisons of Composite Reading Achievement Scores of First Grade Students in NIFDI-Supported, Other DI, and Control Schools, Adjusted for School SES, All Years and 2003 12

Improving First Grade Reaching Achievement in a Large Urban District: The Effects of NIFDI-Supported Implementation of Direct Instruction in Baltimore City Public Schools

Low reading achievement of elementary school students 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.

Sixteen of the schools used Direct Instruction for reading instruction, primarily the *Reading Mastery Classic, Language for Learning, Language for Thinking,* and *Reasoning and Writing* curricula. Other schools in the district adopted the Open Court Reading curriculum. In addition, while all of the schools using Direct Instruction started implementation with the National Institute for Direct Instruction (NIFDI), five of these schools either ceased involvement with NIFDI soon after implementation or began to work with another provider. These variations provide the basis for what could be seen as a natural experiment, providing the opportunity to compare student achievement with different types of curricula and different sources of implementation support.[1]

A very large body of literature, including well designed meta-analyses, has demonstrated the effectiveness of Direct Instruction in promoting high reading achievement. A slightly smaller body of literature has highlighted the importance of consistent implementation of the program in achieving the best results. Using data on the achievement of first graders in the BCPSS from 1998 through 2003, this report compares achievement scores of students in schools with NIFDI support with those in other DI schools and in the rest of the Baltimore system. This report examines the extent to which using the Direct Instruction curriculum and consistent technical support in implementation of this curriculum can promote higher levels of achievement. Additional analyses further articulate the extent to which differences in achievement remain when changes over time and the socio-economic and demographic characteristics of the schools are controlled. These analyses disclose the extent to which the Direct Instruction curriculum, with and without NIFDI technical support, can help counter the detrimental impact of a context of poverty on student achievement. In this report results are summarized in graphs and are also expressed as effect sizes. Two appendices provide full statistical details. The methodology is briefly described below, followed by a summary of the results obtained with each research question and a discussion of the implications.

Methodology

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. Measures of reading achievement were obtained each spring using the Comprehensive Test of Basic Skills (CTBS), a widely used, nationally normed standardized test. Subtest measures of Vocabulary and Comprehension, 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

[1] MacIver and Kemper (2002) examined data for a shorter time span from 6 of the BCPSS schools included in this study, but did not differentiate between the NIFDI-supported and other DI schools.

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.

The average school in the BCPSS had large proportions of poor and minority students. On average, schools were quite poor, with 75 percent of the student bodies on free or reduced lunch. This varied however, from a low of 22 percent to a high of 93 percent. The district also had a large proportion of minority students. On average, the student body of elementary schools in the district was 84 percent African American, although the percentage varied from a minimum of three percent to almost 100 percent. Reflecting a high level of segregation, schools varied in their representation of non-Hispanic white students, from having virtually no white students to being 94 percent white, with an average value of 14 percent. Asian, Hispanic and Native American students were quite rare in the district, with an average representation of one percent in the student bodies. There was also substantial variation in the poverty status of the schools.

The NIFDI-supported schools were similar to the Other DI schools and the other BCPSS schools in both racial-ethnic composition and rates of poverty. There were no significant differences between the three groups in the average proportion of students receiving free and reduced lunch or the representation of African American or non-Hispanic white students. The NIFDI schools, however, had significantly higher proportions of Hispanic and Native American children.

One of the schools in the Other DI group had a much lower rate of poverty. On average, over the years in the study, only about one-third of the students in this school received free or reduced lunch. In all of the other schools that received Direct Instruction, seventy-five to ninety percent of the students had free or reduced lunch. Appendix I provides additional details on the sample and measures used in the analysis.

The analysis focused on 1) comparing the achievement of students who received Direct Instruction with those who did not and 2) comparing the achievement of students in the NIFDI-supported schools with those in the two other groups. For the three way comparisons, schools that implemented Direct Instruction, but without NIFDI support throughout the entire study period, are termed "Other DI schools." The remaining schools in the system are termed "Control schools."

Five research questions were addressed:

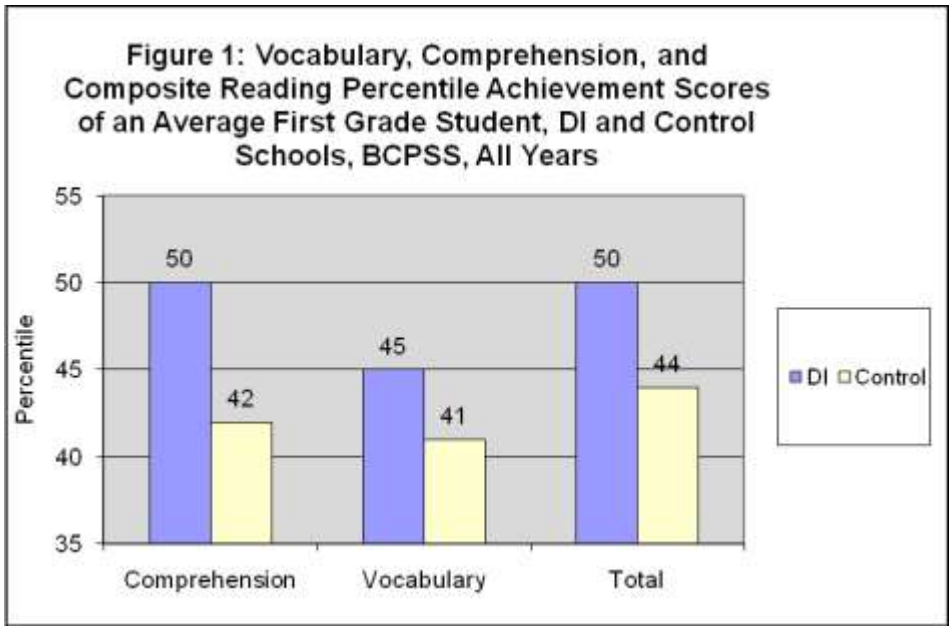
- Did students who received Direct Instruction have higher achievement scores than students in the other schools?
- Did students who received Direct Instruction in schools with consistent NIFDI technical support have higher achievement scores than students in other groups?
- Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools were equalized?
- Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?
- What is the magnitude of the effect of NIFDI-supported Direct Instruction?

Results

A general summary of the results related to each of the research questions is given below. Full statistical details are provided in Appendix II.

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

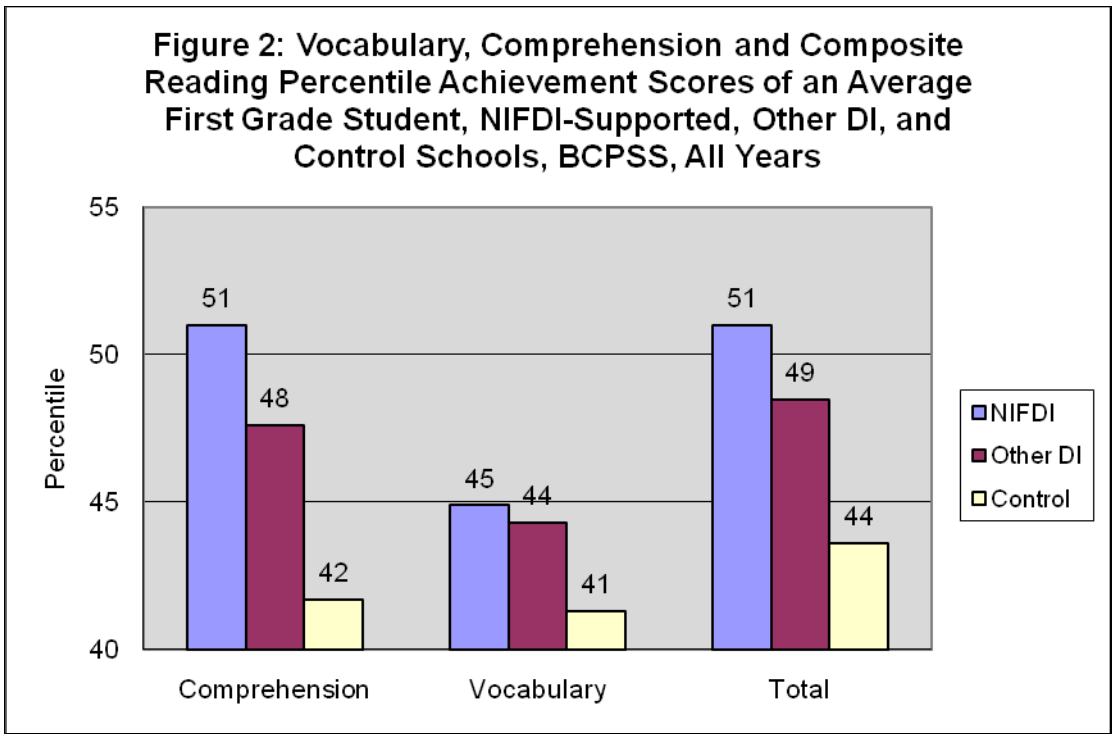
Figure 1 summarizes achievement data for students who received Direct Instruction and those who did not receive DI on each of three measures of reading achievement. These results combine the data from all the available years (1997-98 through 2002-03) and represent the percentile score of the average student in a school. Combining data from all the years in the study, the average first grader that received Direct Instruction scored at the 50th percentile, the national norm, in both the measure of comprehension and the composite score, but somewhat below the national average, at the 45th percentile, on vocabulary. The average student in the Control schools had scores that were substantially lower: at the 42nd percentile for comprehension, the 41st percentile for vocabulary, and the 44th percentile for the composite measure. These differences were statistically significant at well beyond the .001 level of significance.



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

2. Did students who received Direct Instruction in schools with consistent NIFDI technical support have higher achievement scores than students in the other groups?

As illustrated in Figure 2, students in the NIFDI-supported schools had higher average levels of achievement than students in both the Other DI schools and the Control schools. Combining data over all years in the study, the average student in a NIFDI-supported school scored at the 51st percentile on the comprehension and composite measures and at the 45th percentile on vocabulary. The average student in the Other DI schools scored at the 48th percentile on comprehension, at the 44th percentile on vocabulary, and the 49th percentile on the composite measure. Scores were lowest for students in the Control schools. The average student in a Control school had comprehension at the 42nd percentile, vocabulary scores at the 41st percentile and composite scores at the 44th percentile. Differences between the scores in NIFDI-supported schools and those in Control schools were statistically significant on all three measures. However, differences between the scores for students in NIFDI-supported schools and those in the Other DI schools were statistically significant only for the measure of comprehension.

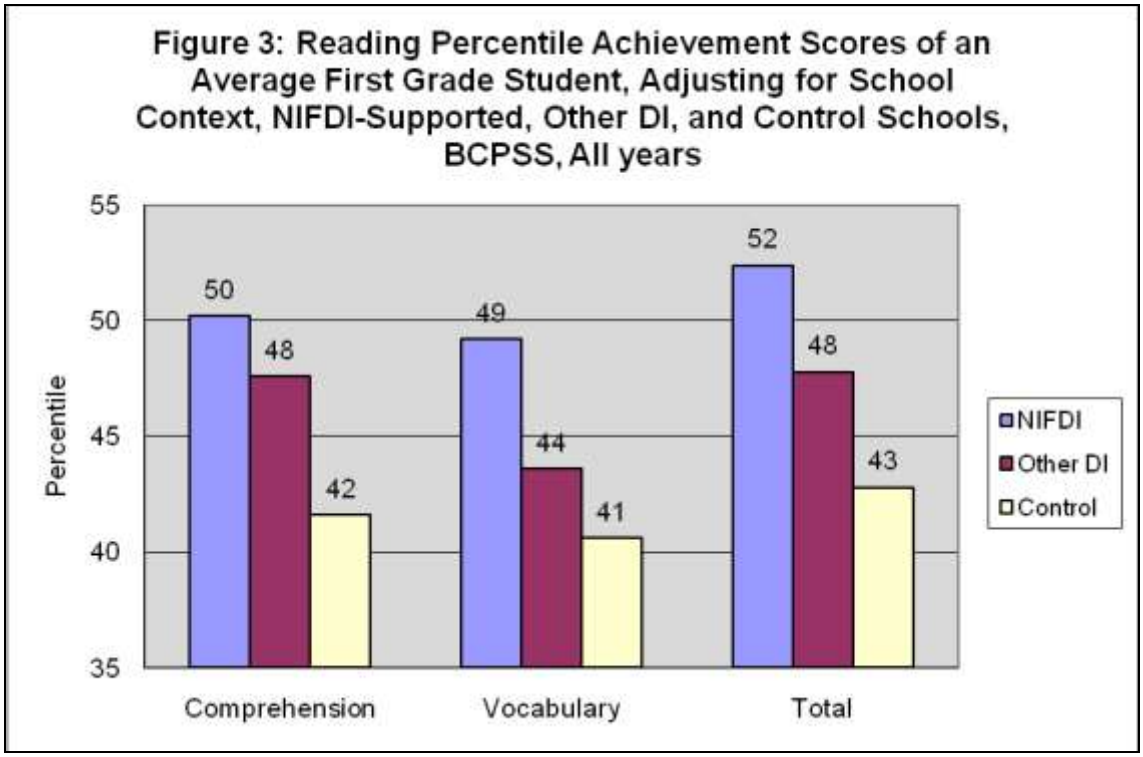


Note: Scores in this graph represent the percentile score of an average student in each type of school averaging over all the years in the analysis: 1998-2003.

3. Did students in NIFDI-supported schools have higher 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.[2] As noted above, elementary schools in the Baltimore City Public School System varied substantially in their racial-ethnic composition and poverty rate. As in other studies, students in BCPSS schools with fewer poor children had higher average levels of achievement.

When the socio-economic characteristics of the schools were statistically controlled the differences between NIFDI-supported schools and other schools became even more marked. Figure 3 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.) The major difference between these results and those in Figure 2 is that the differences between scores of students in the NIFDI-supported schools and those in the Other DI schools are now statistically significant on all three measures. The higher levels of achievement for students in the Other DI schools relative to the Control schools shown in Figure 2 reflect the high achievement of students in the low poverty school in the Other DI group. When statistical controls adjust for this disparity – making the three groups of schools more similar – any advantage for students in the Other DI group disappears.

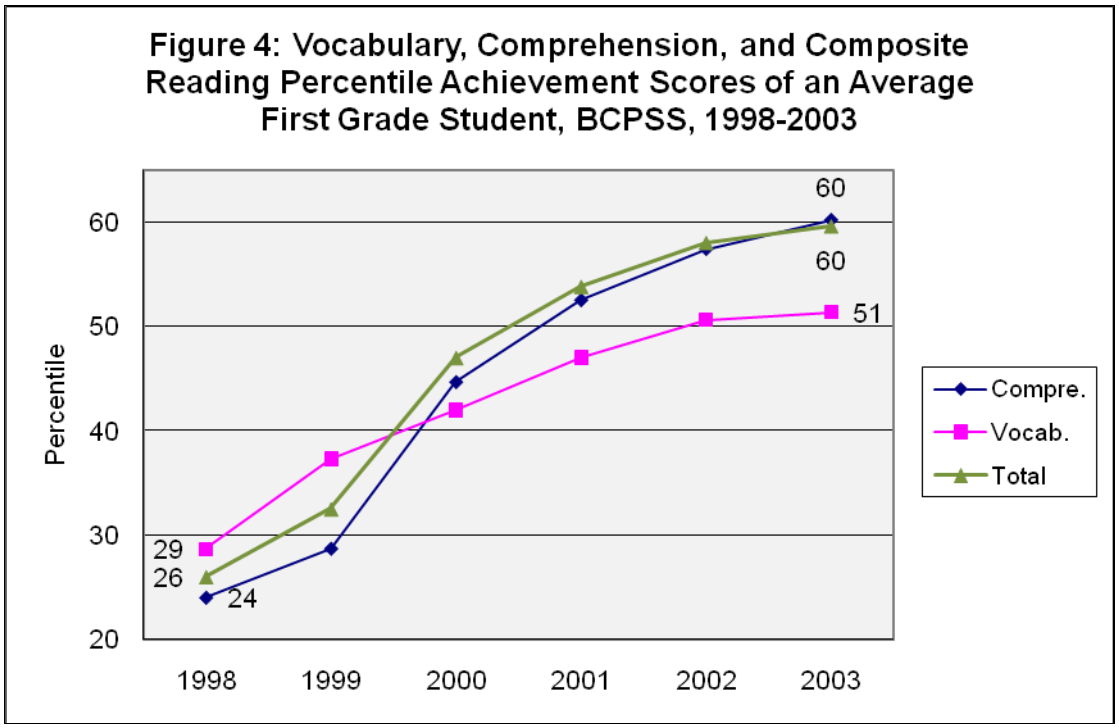


Note: Scores in this graph represent the percentile score of an average student in each type of school if the schools were equal in their socio-economic and demographic characteristics, averaging over all the years in the analysis: 1998-2003.

[2] See Stockard and Mayberry, 1992, for a complete review of this literature.

4. Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?

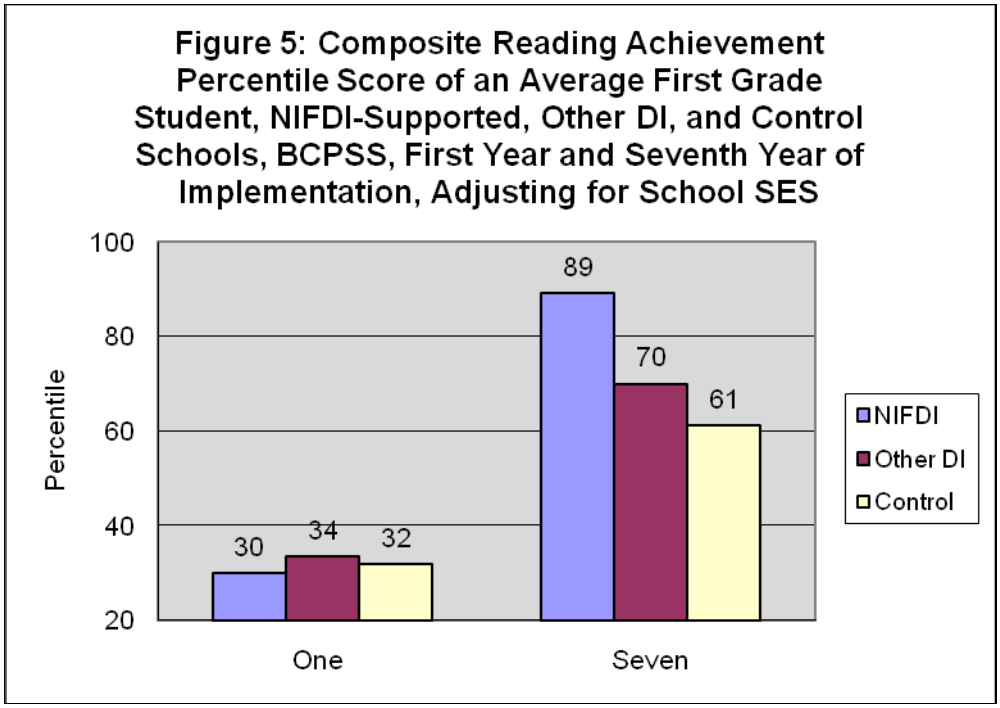
Changing school practices takes time and effort. It is reasonable to expect that the advantage of an effective instructional program would become more apparent as schools have more experience with a new curriculum, teachers gain more practice, and the procedures become part of the institutionalized and accepted practices within a school. As noted above, Direct Instruction was not the only new reading curriculum introduced to the BCPSS during the years included in the analysis. In 1998 the system adopted Open Court Reading for all K-2 schools not in the DI program, and concerted attention was paid throughout the district to enhancing student achievement. The data indicate that these efforts were fruitful. Data in Figure 4 represent the percentile at which an average student achieved at each year for each measure. It can be seen that the reading achievement scores of the average student in the BCPSS increased markedly from 1998 to 2003 on all three of the achievement measures. For instance, in 1998, the average student in BCPSS had a composite achievement score at the 26th percentile. By 2003, the average student scored at the 60th percentile.



Note: Scores in this graph represent the percentile score of an average student in the BCPSS for each year of the analysis and for each measure of achievement.

However, the change in average achievement levels was significantly greater for students in the NIFDI-supported schools than in either the Other DI or Control schools. Figure 5 summarizes the results of this analysis. The bars in the graph give the percentile score on the composite reading measure for an average student in each type of school in the first and last year of implementation of Direct Instruction. All of the calculations equalized school socio-economic status. (Details regarding these calculations are included in Appendix II.) The results show that after the first year of implementation, an average student in a NIFDI-supported school had achievement scores that were equal to or even slightly less than

an average student in the other schools. However, after seven years of implementation, an average student in a NIFDI-supported school had scores that were substantially higher than an average student in other schools.[3] The scores of students in all three groups of schools increased significantly over time, but the increase in scores of students in the NIFDI-supported schools was much greater. In other words, on average all first graders in the BCPSS had significantly higher achievement scores at the end of the study period than at the beginning, but this increase was significantly higher in the schools that were served by NIFDI.



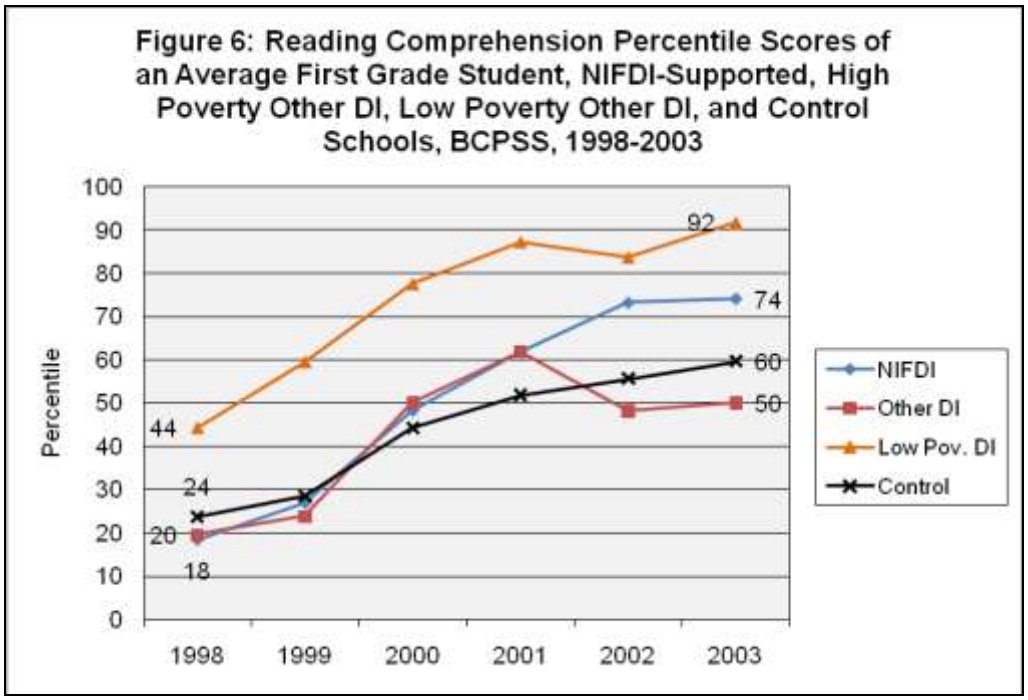
Note: Scores in this graph represent the percentile score on the composite measure of reading achievement for an average first grade student in each type of school in the first year of implementation of DI and the 7th year.

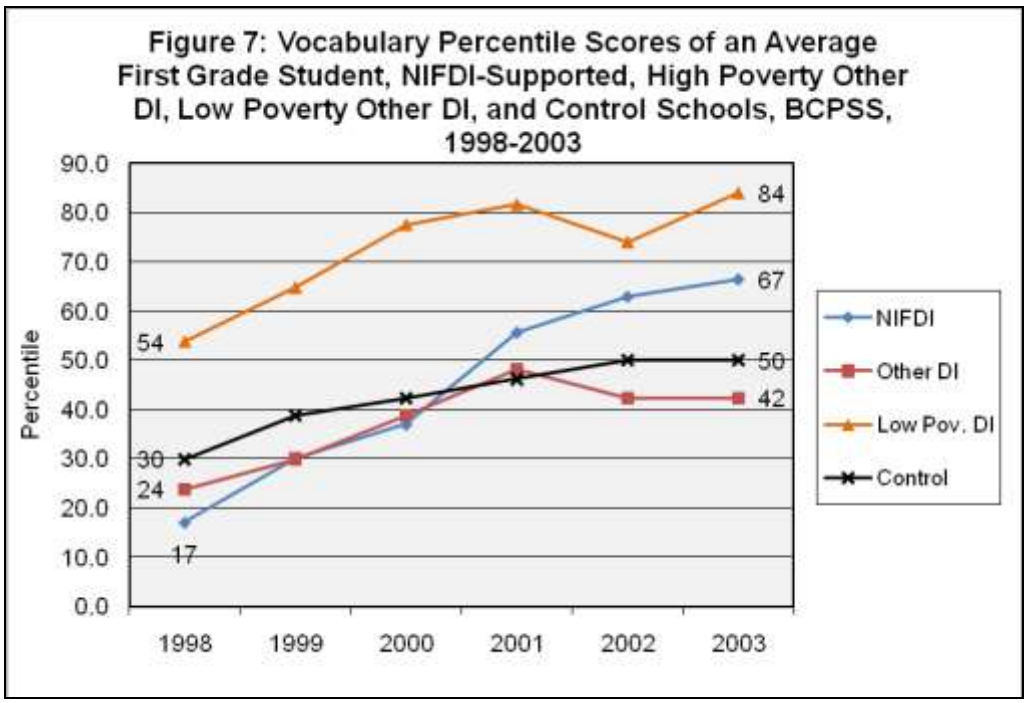
The data in Figure 5 were calculated from the results of statistical analyses that adjusted for variations in school advantage and other differences in the schools. However, the same pattern occurs when the simple raw achievement scores for schools in each group are examined over the time period in the study: 1998-2003. Figures 6 through 8 summarize these data. They show the percentile at which an average student in each group scored in each of the 6 years. In these figures the data for the low-poverty school in the Other DI group is separated from the others.

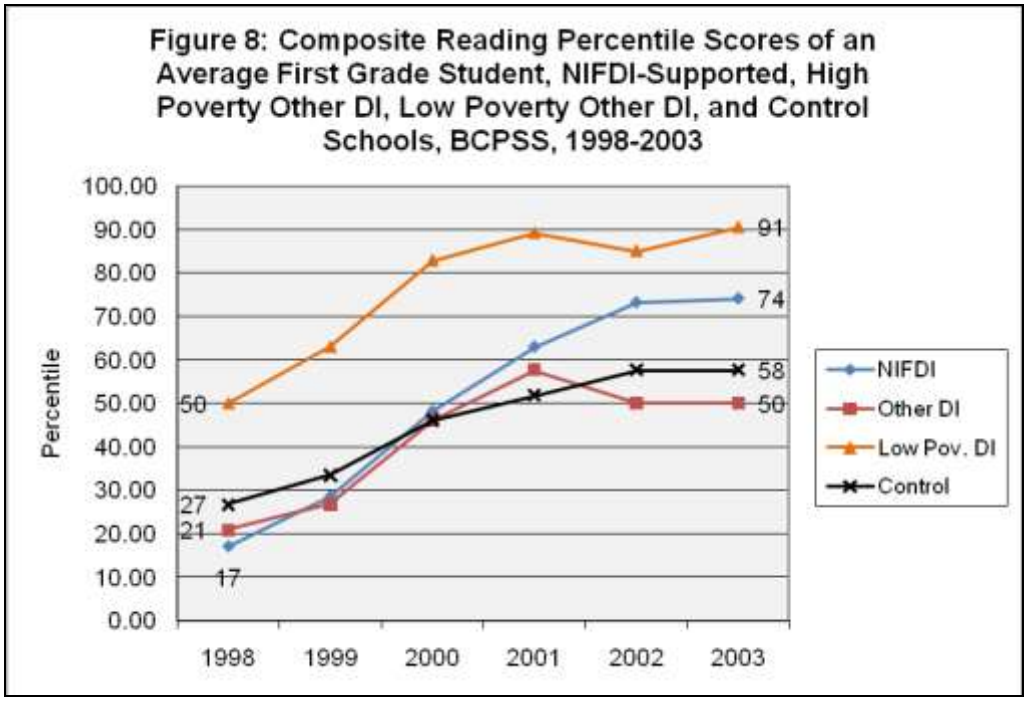
At the beginning of implementation in 1998, an average student in NIFDI-supported schools had lower scores than average students in the other schools, scoring at the 17th percentile in vocabulary and the composite score and at the 18th percentile in comprehension. An average student in the high poverty Other DI schools (those that were demographically similar to the NIFDI-supported schools) had scores only slightly higher than the average student in a NIFDI-supported school, ranging from the 20th to 24th percentile, while an average student in the control schools scored from the 24th to the 30th percentile. In contrast,

[3] All but four of the NIFDI-supported schools and all but one of the Other DI schools began implementation in 1996-97, one year before the start of the data set, thus having 7 years of implementation by the spring of 2003.

in 1998 an average student in the low-poverty school that implemented DI scored from the 44th percentile (on comprehension) to the 54th percentile (on vocabulary). In general, in 1998, an average BCPSS student had very low achievement scores.





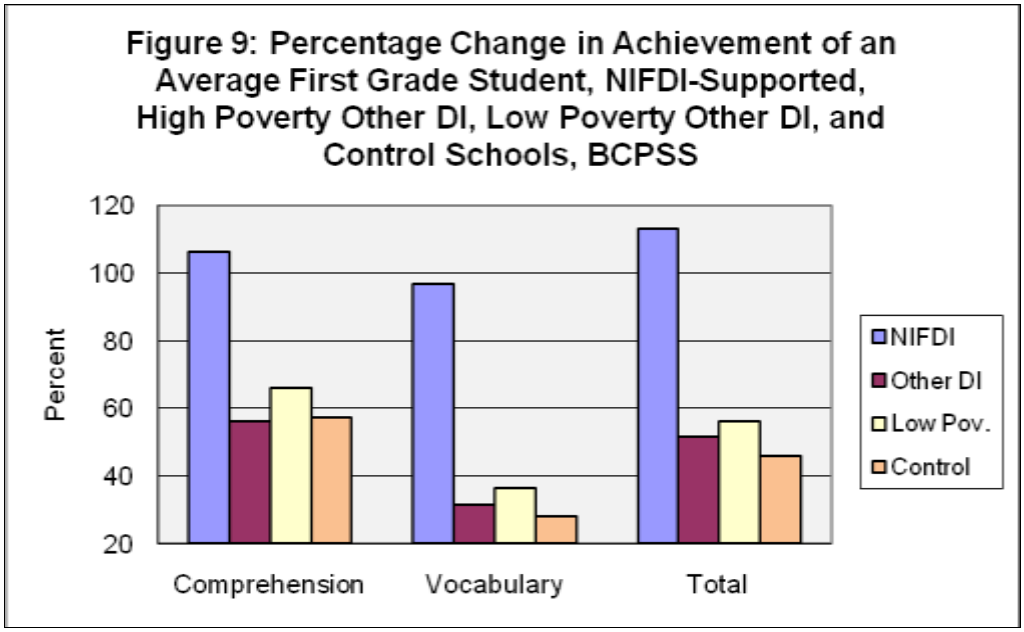


Note: Scores in Figures 6, 7 and 8 represent the percentile score of an average first grade student in each type of school for each year in the analysis. Data for students in the low poverty school in the Other DI group are separated from other schools in that group.

By 2003 the situation had changed markedly. The average student had substantially higher scores than in 1998. For instance, scores for an average student in the Other DI, high poverty schools, ranged from the 42nd (vocabulary) to the 50th percentile (comprehension and composite), while scores for an average student in the Control schools ranged from the 50th (vocabulary) to the 60th percentile (comprehension). An average student in the low-poverty Other DI school had very high scores, ranging from the 84th percentile in vocabulary to the 92nd percentile in comprehension. The changes were even more marked, however, for students in the NIFDI-supported schools. By 2003, the average student in a NIFDI-supported school had percentile scores that ranged from 66 (for vocabulary) to 74 (for comprehension and the composite score). These scores were significantly higher than those of students in both the Other DI schools (excluding the low poverty school) and the control schools.

Figure 9 illustrates the magnitude of these changes. Values in Figure 9 represent the difference in the 1998 and 2003 scores as a percentage of the 1998 score, thus showing the percentage increase over time. The differences between the groups are very large. For instance, for the composite scores, the percentage increase in achievement for first grade students in the NIFDI-supported schools was over twice that for students in the other three groups: 113 percent versus 46 percent (the control schools) to 56 percent (the low poverty school in the Other DI group).[4]

[4] 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.



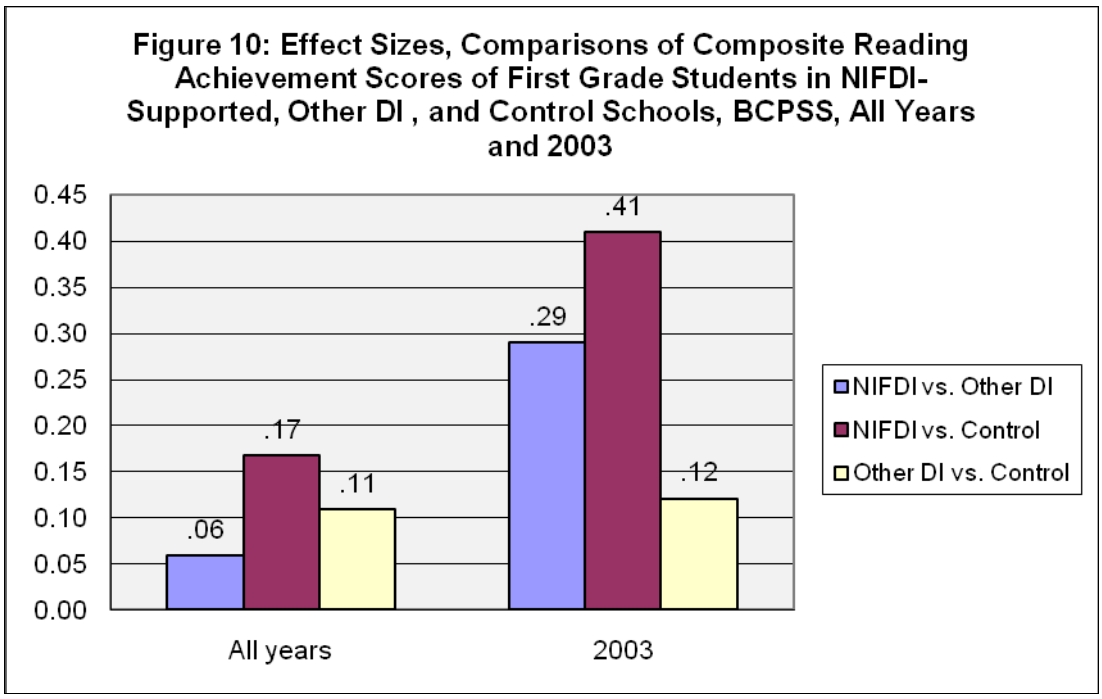
Note: The figures in this graph represent the percentage change in achievement from 1998 to 2003 for each of the groups of schools, separating the low poverty school in the Other DI group from others.

5. What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?

Unless specifically noted, all of the results presented earlier 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.[5]

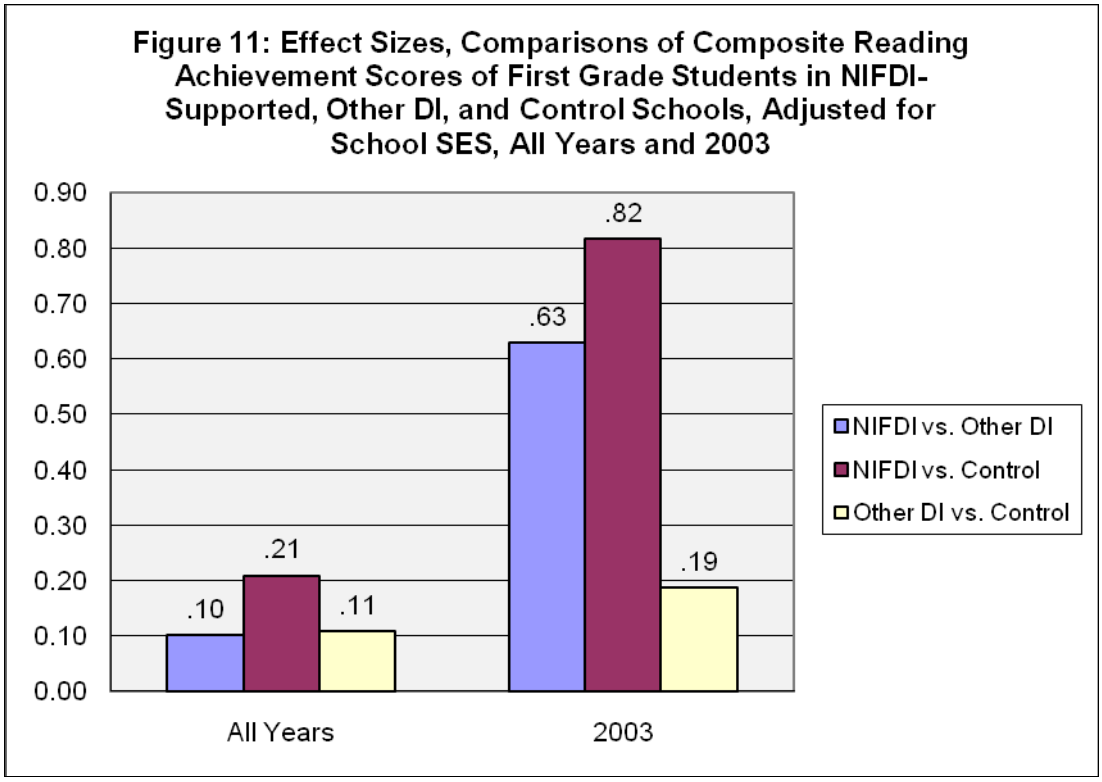
Figures 10 and 11 summarize the magnitude of the various effects reported earlier in this report. They compare the three groups: NIFDI-supported schools, Other DI schools, and the Control Schools. Figure 10 summarizes the effects without any adjustment for school characteristics. When data from all the years are examined together, the effect sizes comparing achievement of students in the NIFDI-supported schools and those in the two other groups of schools are lower than the .25 criterion (.06 for the comparison of NIFDI and the other DI schools and .17 for the comparison of the NIFDI and Control schools). However, by 2003 the effect sizes were substantially larger: .29 for the comparison of students in the NIFDI schools and in the other DI schools and .41 for the comparison with students in the Control schools. The comparisons of students in the Other DI group with those in the Control schools, while positive, were substantially lower in both comparisons (.11 for all years and .12 for 2003).

[5] This criterion is taken from Fashola and Slavin (1997).



Note: Values in the table are effect sizes, a measure of the magnitude of the difference between two groups. Effect sizes of .25 or larger are traditionally considered educationally significant.

Figure 11 gives the effect sizes for the measure of composite reading achievement for all years and 2003 after the impact of school socio-economic status has been controlled. Averaged across all years, the effect sizes in comparison to other DI schools was .10 and in comparison to the Control schools was .19. However, at the end of the study period, the effect of being in a NIFDI-supported school relative to other DI schools was .63 and relative to the Control schools was .82. The effect size of Other DI relative to the Control schools remained small and below the level of educational significance (values of 0.11 for all years and .19 in 2003).



Note: Values in the table are effect sizes, a measure of the magnitude of the difference between two groups. Effect sizes of .25 or larger are traditionally considered educationally significant.

Summary

This report analyzed differences in first grade reading achievement of students in the Baltimore City Public School System from 1997-98 through 2002-03. The BCPSS is typical of large urban school districts in serving large numbers of poor and minority students with low levels of academic achievement. Eleven schools in the BCPSS implemented Direct Instruction programs under the guidance of NIFDI during this time period. Five other schools used DI material but had NIFDI support for only the initial years. Beginning in 1998 the remaining schools used the Open Court curriculum. In the spring of each year all first graders were administered the Comprehensive Test of Basic Skills. Data from these tests were analyzed for this report.

Despite an extremely large body of empirical evidence, some continue to suggest that Direct Instruction is not effective in enhancing student achievement. The results from this study, with over 40,000 students and 6 years of data, would appear to contradict that statement. The results confirm previous research that has found the Direct Instruction curriculum to be effective in promoting student achievement. Averaging across all the years in the study, students exposed to Direct Instruction had higher achievement scores than students in the control schools.

The results also, however, provide an important caveat regarding the role of strong implementation and technical support. Once the socio-economic characteristics of schools were controlled (a necessary step given the low poverty rate of one of the Other DI schools), students in the NIFDI-supported schools had significantly higher achievement than both the students in the Control schools and those in the Other DI schools. By the end of the study

period the achievement scores of students in the NIFDI-supported schools were well above national averages and close to those of students in the low poverty Other DI group. Thus, while these results confirm the role of Direct Instruction in promoting high achievement, they also suggest that consistent, well-designed technical support may be important to ensure that these superior results occur.

The analysis may address some common objections to Direct Instruction. For instance, it has been claimed that Direct Instruction involves a relatively mindless application of a script with little intellectual engagement or skill required by the teacher. The results presented in this report showed that strong gains in achievement only appeared several years after schools began to implement DI and then only with consistent technical support. One could ask, "If DI simply employs mindless, rote tasks, why does it take so much time for the effects to appear?" A logical reply could be that it is not a mindless, rote task at all. Instead, 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.

Second, some have suggested that Direct Instruction is useful for teaching rote, elementary skills such as decoding, but has more limited utility in teaching higher level skills such as comprehension. The results in this paper would contradict that view as well. Increases in achievement for students in the NIFDI-supported schools on comprehension were larger than those in vocabulary, the area more likely to be "rote" in nature.

Third, some have suggested that DI works only with low performers, but holds back high performing students. The results with the low poverty school that implemented DI, albeit without NIFDI support after the first few years, belie this suggestion. Even though students in this school had scores that averaged at the national mean at the beginning of the study, their scores increased markedly during the 6 years for which data were gathered. In addition, by the later years of the study, first grade students in the NIFDI schools had joined the ranks of high performers, with achievement scores well above the national norms. The scores of first graders in these schools continued to be maintained at these high levels and to even improve, providing no indication that their progress was impeded.

Appendix I
Sample and Measures

This appendix provides details on the sample and the measures used in the analysis.

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. An additional 5 schools started implementation with NIFDI but then ceased involvement or began to work with another provider. The former group is referred to below as the NIFDI-supported schools, and the latter group as Other DI schools.[6] The remaining 103 schools are termed the Control schools. Schools in the two intervention groups are listed in Table A-1[7]

Table A-1: Schools in Treatment Groups and Start Dates of Treatment

NIFDI Schools	Starting Year
Arundel	96 - 97
CC Barrister	97 - 98
City Springs	96 - 97
Collington Square	98 - 99
Dickey Hill	98 - 99
Federal Hill	97 - 98
General Wolfe	96 - 97
Hampstead Hill	96 - 97
Langston Hughes	98 - 99
Margaret Brent	98 - 99
Dr. Rayner Browne	98 - 99
Other DI Schools	
William Pinderhughes	96 - 97
George Kelson	96 - 97
Robert Coleman	96 - 97
Roland Park	96 - 97
Westport	97 - 98

[6] 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.

[7] 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 also lists the date at which programs were implemented in each school. In the analysis schools were designated as belonging to a given condition only during the years in which their school was receiving an implementation. For instance, Collington Square had no intervention in the first year of data collection (1997-98), but began implementing Direct Instruction reading programs in 98-99. Thus, students in Collington Square were determined to be in the NIFDI reading intervention group in 98-99 and subsequent years, but in the control condition in 1997-98. 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 reading in 1998-99, its second in 1999-2000, etc.

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, Reading Comprehension and Reading Vocabulary, and a Composite Reading Achievement score were analyzed.

Normal Curve Equivalent (NCE) scores were used for all statistical analyses. NCE scores are calculated by translating the raw or 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 virtually all of the common statistics used in analysis. 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.[8]

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.

[8] The conversion can also be accomplished manually 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.

predicting 1998 levels of proportion African-American from levels in 1999 through 2003).[9] Values for each variable were averaged across the years to produce aggregate measures of the demographic characteristics of the school.

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

A: Descriptive Statistics

	NIFDI		Other DI		Control		Total	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Asian	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01
African-American	0.75	0.34	0.93	0.13	0.85	0.27	0.84	0.27
Hispanic	0.05	0.12	0.002	0.003	0.01	0.01	0.01	0.04
Native American	0.01	0.02	0.001	0.001	0.00	0.00	0.003	0.01
Non-Hispanic White	0.17	0.25	0.06	0.12	0.14	0.25	0.14	0.25
Free and Reduced Lunch	0.83	0.06	0.72	0.21	0.74	0.15	0.75	0.15
Factor 1	-0.12	0.96	-0.07	1.06	0.02	1.01	0.01	1.00
Factor 2	1.18	2.82	-0.34	0.25	-0.11	0.47	0.00	1.00
N	11		5		103		119	

B: Analysis of Variance and Tukey Post-Hoc Tests

	Analysis of Variance		Post-Hoc Comparisons (significance)		
	F	p	NIFDI v. Other DI	NIFDI v. Control	Other DI v. Control
Asian	2.01	0.14	0.61	0.12	0.96
African-American	0.85	0.43	0.46	0.54	0.78
Hispanic	5.51	0.01	0.08	0.004	0.95
Native American	6.99	0.001	0.04	0.001	0.92
Non-Hispanic White	0.35	0.70	0.68	0.90	0.78
Free and Reduced Lunch	2.06	0.13	0.35	0.12	0.97
Factor 1	0.12	0.89	1.00	0.89	0.98
Factor 2	9.76	<.001	0.009	<.001	0.85

[9] 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.

Panel A of Table A-2 gives means and standard deviations on each of the demographic characteristics for each group of schools. Analysis of variance indicated no significant differences between the schools in their proportion of African American, non-Hispanic White, or Asian students, as well as no difference in the proportion receiving free or reduced lunch. The NIFDI schools, however, had significantly more Hispanic and Native American students than the students in the other two groups of 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 white students as well as the average proportion of students receiving free and reduced lunch. A principle components extraction method was used, with varimax rotation. Two significant factors (with eigenvalues greater than 1.0) were found.[10] 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.[11]

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 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 first factor, which accounts for 44 percent of the total variance, has a strong positive loading of the proportion of 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 white and Asian students and

[10] 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.
[11] N = 120 for these analyses, for Carroll School was included. Results are identical when Carroll is omitted.

many fewer African-American and poor students. As can be seen in Table A-2, there are no significant differences between the three treatment groups on this factor.

The second factor, which accounts for 29 percent 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 indicate that the NIFDI schools had significantly higher scores on this factor.

In addition, one of the schools in the Other DI group, Roland Park, was strikingly different from the other schools in its rate of children receiving free and reduced lunch. On average only .34 of the students in Roland Park had free or reduced lunch, the third lowest rate in the BCPSS. The other schools in the Other DI group had free lunch proportions ranging from .76 to .84, while schools in the NIFDI group had proportions ranging from .74 to .90. While the presence of Roland Park in one of the treatment groups allows the examination of the effects of DI in more advantaged schools, it was also important to control for the presence of this school in the various analyses.

Appendix II: Detailed Statistical Results

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 analysis of data focused on comparing the achievement of students who received Direct Instruction, in the NIFDI schools and in the Other DI schools, with those in the Control Schools. Six research questions were addressed, each providing a slightly more complex method of control to help ensure the accuracy of the results.

The first research question was, "Did students who received Direct Instruction have higher 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 NIFDI and other 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 in schools that received consistent NIFDI technical support have higher achievement scores than students in the other groups?" The analysis for this question compared the average scores of students in the three groups of schools (NIFDI-supported, Other DI, and Control Schools) across all the years of data using simple one-way analysis of variance and post-hoc tests to compare subgroup means.

The third research question asked, "Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools are equalized?" This analysis is especially important given the unique low poverty status of one school in the Other DI group. Two approaches were used. The first was a simple one-way analysis of variance, separating data from the low poverty Other DI school from other schools within that group. The second method was mixed model regression, including measures of school advantage and disadvantage from the factor analysis as predictors. To the extent that Direct Instruction, as implemented and supported by NIFDI, is more effective at promoting student achievement, it would be expected that students in the NIFDI group would have higher achievement scores than those in the other two groups, 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 fourth research question asked, "Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?" It was expected that the impact of the NIFDI implementation would become stronger over time. This question was analyzed both through examination of mean scores over time, with analysis of variance and post-hoc comparison of 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 NIFDI-supported Direct Instruction." 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 (Fashola and Slavin 1997).

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

The first research question addressed was, **"**Do students who received Direct Instruction have higher achievement scores than students in the other schools?" This was answered through comparing the average achievement scores of students in the combined NIFDI and Other DI groups of 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 addressing the hypothesis that any differences occur by chance. Panel B translates the average Normal Curve Equivalent scores in Panel A to the corresponding percentiles. Data in Panel B were used to create Figure 1 in the text. 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.

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: Comprehension, Vocabulary, and the Composite score. The average student in the Control schools, across all the years in the study, had scores ranging from the 41st to 43rd percentile. In contrast, the average student in a DI school had scores ranging from the 45th to the 50th percentile. On both comprehension and the composite score the average DI student had a score at the national average. The lowest scores for both groups occurred with the measure of Vocabulary.

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

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

Comprehension	Control Schools	Had Direct Instruction
Mean	45.6	49.8
S.D.	23.7	24.7
N	41043	4820
t-value	11.43	5905.60
d.f.	5906	
Prob.	<.001	
Vocabulary		
Mean	45.4	47.2
S.D.	23.1	24.7
N	40825	4754
t-value	4.78	5764.08
d.f.	5764	
Prob.	<.001	
Composite		
Mean	46.6	50.1
S.D.	23.0	24.5
N	40425	4705
t-value	9.23	5711.99
d.f.	5711	
Prob.	<.001	

Panel B: Percentile Scores Corresponding to Mean Values

	Control School Students	Direct Instruction Students
Comprehension	42	50
Vocabulary	41	45
Composite	44	50

Research Question 2:"Did students who received Direct Instruction in schools that received consistent NIFDI technical support have higher achievement scores than students in the other groups?"

Simple one-way analysis of variance was used to analyze differences in average achievement between the three groups: NIFDI-supported schools, Other DI schools, and the Control schools. Table A-5 summarizes the results of this analysis. Panel A gives the means and standard deviations on the three measures of achievement. Panel B includes results of one-way analyses of variance and Tukey post-hoc tests comparing each pair of schools. Finally, Panel C translates the average normal curve equivalent scores in Panel A into the corresponding percentiles. The results in Panel C were used to create Figure 2 in the text.

Table A-5: Average Reading Achievement Scores by Group, All Years, First Graders, BCPSS

Panel A: Descriptive Statistics (NCE Scores) and Analysis of Variance Results

Comprehension	NIFDI	Other DI	Control	Total
Mean	50.5	48.7	45.6	46.0
S.D.	24.4	25.2	23.7	23.8
N	2991	1829	41043	45863
Vocabulary				
Mean	47.3	47	45.4	45.6
S.D.	25.2	24	23.1	23.3
N	2948	1806	40825	45579
Composite				
Mean	50.6	49.2	46.6	47.0
S.D.	24.5	24.6	23	23.2
N	2911	1794	40425	45130

Panel B: Analysis of Variance and Post-Hoc Pair-Wise Comparisons

	Analysis of Variance		Tukey Post-Hoc Tests (p)		
	F Value	prob.	NIFDI v. Other DI	NIFDI v. Control	Other DI v.Control
Comprehension	73.17	<.001	0.03	<.001	<.001
Vocabulary	12.85	<.001	0.86	<.001	0.01
Composite	49.19	<.001	0.09	<.001	<.001

Panel C: Percentile Scores Corresponding to Means in Panel A

	NIFDI	Other DI	Control	Total
Comprehension	51	48	42	42
Vocabulary	45	44	41	42
Composite	51	48	44	44

Inspection of the means shows that students in the NIFDI-supported schools had higher achievement than students in the Other DI schools or the Control schools on all three

measures. The analysis of variance results indicate that there are significant differences between these three groups on all three measures. All of the pair-wise comparisons with the control schools, for both the NIFDI-supported and the Other DI schools, are statistically significant. However, the differences between the NIFDI schools and the Other DI schools are only significant for the measure of comprehension.

Panel C, which translates the Normal Curve Equivalent scores to percentiles, indicates that the average student in a NIFDI-supported school, across all the years in the study, scored slightly above the national average on Comprehension and the Composite score, but somewhat below the average in Vocabulary. Percentiles for the average student in the Other DI group and the Control schools were below the national average for all measures.

Research Question 3: Did students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools are equalized?

Research Question 3 deals with the extent to which students in NIFDI-supported schools have higher achievement scores than students in other schools when the demographic characteristics of their schools are equalized. This question is answered in two ways. The first is with descriptive data and analyses of variance and the second uses mixed model regressions.

Descriptive Data and Analyses of Variance

The descriptive statistics, analysis of variance, and post-hoc comparisons of means are in Table A-6. The analysis parallels that of Table A-5 except that, within the Other DI group, the results for the low poverty school are separated from those for the other schools.

Table A-6: Average Reading Achievement Scores by Group, All Years, First Graders, BCPSS, Four Groups
Panel A: Descriptive Statistics (NCE Scores) and Analysis of Variance Results

Comprehension	NIFDI	HPODI	LPODI	Control	Total
Mean	50.5	44.2	65.3	45.6	46.0
S.D.	24.4	24.1	22.0	23.7	23.8
N	2991	1434	395	41043	45863
Vocabulary					
Mean	47.3	42.2	63.8	45.4	45.6
S.D.	25.2	22.8	20.4	23.1	23.3
N	2948	1411	395	40825	45579
Composite					
Mean	50.6	44.2	66.8	46.6	47.0
S.D.	24.5	23.1	21.7	23	23.2
N	2911	1399	395	40425	45130

Panel B: Analysis of Variance and Post-Hoc Pair-Wise Comparisons (probabilities)

	F Value	prob.	NIFDI v. HPODI	NIFDI v. LPODI	NIFDI v. Control	HPODI v. LPODI	HPODI v. Control	LPODI v. Control
Comprehension	130.85	<.001	<.001	<.001	<.001	<.001	0.131	<.001
Vocabulary	96.95	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Composite	130.65	<.001	<.001	<.001	<.001	<.001	0.001	<.001

Panel C: Percentile Scores Corresponding to Means in Panel A

	NIFDI	HPODI	LPODI	Control	Total
Comprehension	51	39	76	42	42
Vocabulary	45	36	74	41	42
Composite	51	39	79	44	44

Note: HPODI signifies the high poverty Other DI schools, which are demographically similar to the NIFDI-supported schools, LPODI refers to the low poverty Other DI school.

The results in Table A-6 show that when the low poverty other DI (LPODI) school is removed from the Other DI group, the differences between the achievement scores of students in that group and students in the NIFDI-supported schools become even larger. All of the differences between the students in the NIFDI-supported schools and those in the high poverty Other DI (HPODI) schools are highly statistically significant, as are all the comparisons with the control schools. Comparisons of students in NIFDI-supported schools with those in the LPODI school are also highly significant, but the result is in favor of the low poverty school. In fact, averaging over all years in the study, first grade students in this low poverty, DI school had the second highest average score in the BCPSS on both the vocabulary and the composite measures and the highest measure of all schools in comprehension.

Mixed-Model Results

The second method of addressing this research question was through the use of mixed model regressions. Table A-7 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.

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 nine to ten 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.

Table A-7: Variables in Models and Measures of Fit, Analysis of First Grade Reading 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
Group (NIFDI, Other DI)			x	x
Years of Treatment				x
Treatment * NIFDI				x
Factor 1 * NIFDI				
Factor 1 * Other DI				
Panel B: Model Fit Statistics				
Comprehension				
Random Effects Estimate	49.2	38.2	36.3	35.8
s.e.	6.7	5.2	4.9	4.9
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	522.9	467.63	467.52	466.09
s.e.	3.5	3.1	3.1	3.1
sig.	<.0001	<.0001	<.0001	<.0001
BIC	417661	412535	412528	412396
- 2 Log Likelihood	417647	412506	412490	412348
Change in LL		5140.2	16.4	141.9
df		3	2	2
P		<.001	<.001	<.001
PRE measure		0.11	0.0002	0.003
Correlation ratio	0.09			
Vocabulary				
Random Effects Estimate	52.3	40.2	39.8	40.1
s.e.	7.1	5.4	5.4	5.5
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	497.7	480.3	480.2	477.9
s.e.	3.3	3.2	3.2	3.2
sig.	<.0001	<.0001	<.0001	<.0001

BIC	412833	411204	411198	410994
-2 Log Likelihood	412818.7	411175	411160	410946
Change in LL		1643.6	15.1	214
df		3	2	2
P		<.001	<.001	<.001
PRE measure		0.03	0.0003	0.005
Correlation ratio	0.10			
Composite				
Random Effects Estimate	54.7	41.8	40.4	40.2
s.e.	7.4	5.6	5.4	5.4
sig.	<.0001	<.0001	<.0001	<.0001
Residual Estimate	490.7	444.95	444.8	442.9
s.e.	3.3	3.0	3.0	3.0
sig.	<.0001	<.0001	<.0001	<.0001
BIC	408143	403719	403710	403524
-2 Log Likelihood	408128	403690	403672	403476
Change in LL		4437.8	18.7	195.9
df		3	2	2
sig.		<.0001	<.001	<.0001
PRE measure		0.09	0.0003	0.004
Correlation ratio	0.10			

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.

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). Thus, the results with Models 2 and 3 provide a second way of testing Research Question three.

The -2 log likelihood measures and the BIC values in Panel B of Table A-7 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 comprehension, the change in the -2 Log Likelihood = 5174 (417,647 - 412,506). 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.[12]

Table A-8 gives the coefficients associated with each of the models in Table A-7, beginning with Model 2. As expected, the coefficient associated with Factor 1 (where higher scores indicate schools with fewer students on free and reduced lunch and more non-Hispanic whites) is positive and highly significant. Also as expected, the coefficient associated with Factor 2 (where higher scores indicate more Hispanic and Native American children) is negative, but is not significant in Model 2. 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 Research Question 3 by including both the dummy variables for treatment group and the controls for the measures of school context. Two dummy variables were used to designate students in schools receiving the two treatments: NIFDI directed interventions and DI interventions not administered by NIFDI. The Control schools are the omitted category for both variables. If NIFDI support has a unique impact we would expect stronger associations with that variable. 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 coefficient associated with being in a NIFDI-supported school is positive and highly significant ($p<.0001$) for all three measures of achievement. In contrast, the coefficient associated with being in an Other DI school, while positive, is not significant in any of the models. In other words, once the demographic context of the school is controlled, students in the NIFDI-supported schools had significantly higher achievement than both students in the Control schools and students in the Other DI schools.

Table A-9 shows the predicted achievement scores for students in each of the three groups of schools based on the results in Model 3. These values were calculated by substituting average values for the entire sample for year and the two factor scores into Model 3 and calculating predicted values for students in each of the three groups. Thus, the

[12] 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. For example, the PRE measure associated with adding the treatment groups to the model is 10 times as large when the groups are entered before the measures of year and factor, but still not as large as when year and factor are entered.

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 the three groups would be expected to achieve if schools were equal in their demographic characteristics. The values in Panel B were used in Figure 3 in the body of the text.

Table A-8: Coefficients Associated with Mixed Model Regressions of First Grade Reading Achievement, BCPSS, 1998-2003

	Model 2			Model 3			Model 4		
Comprehension	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	-8827.2	120.6	<.0001	-8776.0	121.2	<.0001	-8363.1	127.6	<.0001
Year	4.4	0.1	<.0001	4.4	0.1	<.0001	4.2	0.1	<.0001
Factor 1	2.3	0.6	<.0001	2.3	0.6	<.0001	2.4	0.6	<.0001
Factor 2	-0.5	0.6	0.39	-1.0	0.6	0.10	-1.5	0.6	0.01
NIFDI	-----	-----	-----	4.6	1.2	<.0001	-3.8	1.4	0.01
Other DI	-----	-----	-----	3.2	2.8	0.26	0.4	3.0	0.91
Years of Treatment	-----	-----	-----	-----	-----	-----	0.7	0.3	0.020
Treatment * NIFDI	-----	-----	-----	-----	-----	-----	2.3	0.4	<.0001
Vocabulary	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	-4920.2	122.3	<.0001	-4868.9	123.0	<.0001	-4372.0	129.4	<.0001
Year	2.5	0.1	<.0001	2.5	0.1	<.0001	2.2	0.1	<.0001
Factor 1	2.8	0.6	<.0001	2.9	0.6	<.0001	2.9	0.6	<.0001
Factor 2	-0.8	0.6	0.16	-1.3	0.6	0.03	-2.0	0.6	0.001
NIFDI	-----	-----	-----	4.6	1.2	0.0001	-5.9	1.4	<.0001
Other DI	-----	-----	-----	1.6	2.9	0.59	-0.9	3.2	0.78
Years of Treatment	-----	-----	-----	-----	-----	-----	0.6	0.3	0.04
Treatment * NIFDI	-----	-----	-----	-----	-----	-----	3.1	0.4	<.0001
Composite	b	s.e.	p.	b	s.e.	p.	b	s.e.	p.
Intercept	-8011.6	118.3	<.0001	-7956.1	119.0	<.0001	-7490.8	125.2	<.0001
Year	4.0	0.1	<.0001	4.0	0.1	<.0001	3.8	0.1	<.0001
Factor 1	2.7	0.6	<.0001	2.7	0.6	<.0001	2.8	0.6	<.0001
Factor 2	-0.7	0.6	0.24	-1.2	0.6	0.04	-1.8	0.6	0.003
NIFDI	-----	-----	-----	5.0	1.2	<.0001	-4.7	1.4	0.001
Other DI	-----	-----	-----	2.6	3.0	0.38	-0.1	3.2	0.98
Years of Treatment	-----	-----	-----	-----	-----	-----	0.7	0.3	0.03
Treatment * NIFDI	-----	-----	-----	-----	-----	-----	2.8	0.4	<.0001

Table A-9: Predicted Achievement Scores Assuming Equal School Contexts

Panel A: Normal Curve Equivalent Scores

	NIFDI	Other DI	Control
Comprehension	50.1	48.7	45.5
Vocabulary	49.6	46.6	45.0
Composite	51.3	48.8	46.2
Panel B: Percentile Scores			
Comprehension	50.2	47.6	41.6
Vocabulary	49.2	43.6	40.6
Composite	52.4	47.8	42.8

Note: Scores were computed by substituting average values for entire sample for Year and Factors 1 and 2 into the equations associated with Model 3 in Table A-8. 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 did 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 pattern of substantive results would not differ.

Research Question 4: Did the impact of being in a NIFDI-supported school increase over time as reforms became institutionalized within a school?

Research question four tests the hypothesis that the advantage to students in the NIFDI schools 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 NIFDI school would become greater over time.

Model 4 in Tables A-7 and A-8 addresses this hypothesis by adding the interaction of membership in the NIFDI group and the years of treatment. Years of treatment measures how long DI had been implemented in a school.[13] It was expected that the coefficient associated with the interaction of years of implementation and NIFDI would be positive, indicating stronger effects of NIFDI as years of treatment increased.

The results show very strong support for this hypothesis. The positive and significant coefficients associated with year indicate that achievement increased for all students in the BCPSS over the years of the study, and the positive and significant coefficients associated with the variable "years of treatment" indicate that, besides this general increase, students in

[13] The interaction of year of treatment and membership in the Other DI schools is not needed in the equation because this is captured by the years of treatment variable.

the schools that implemented DI had additional increases. Yet, students in a NIFDI-supported school had even stronger increases in achievement from year to year. With all three measures, the results with Model 4 indicate a strong and significant impact of the interaction of being in a NIFDI school and years of implementation.

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 3.8 points in their normal curve equivalent scores on the composite measure of reading achievement (the coefficient associated with year), controlling for differences in the demographic context of the schools. First grade students in the Other DI schools had an average gain of 4.5 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: 3.8 + 0.7). In contrast, the average change in the NIFDI schools, net of school context, was 7.3 NCE points from one year of implementation to the next (3.8 + 0.7 + 2.8), a yearly expected increase in the average scores of first graders that was almost twice the magnitude of that of the Control schools.

Table A-10 gives the scores that would be predicted for students in the three groups of schools if they attended schools with similar demographic characteristics for both one year of implementation of Direct Instruction and 7 years of implementation. These scores were obtained by using the regression coefficients in Table A-8 and substituting the average factor scores and year for all cases in the sample, as was done for the calculations reported in Table A-9.[14] 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 and were used to create Figure 5 in the text.

Results show that with only one year of implementation the average scores of students in the three groups are very similar. However, after seven years of implementation, differences are dramatic. Assuming that they attended schools with similar socio-economic contexts, the average student in a NIFDI-supported school would be expected to score between the 84th and 89th percentile, the average student at an Other DI school would score from the 58th to the 72nd percentile, and the average student at a control school would score from the 52nd to the 63rd percentile, depending on the measure of achievement.

[14] The average values for Factor 1 and Factor 2 that were used for calculations in Table A-9 were also used for the calculations for Table A-10. For year, the average value used in the calculations was 1998.7 for year 1 of implementation and 2003 for year 7. The value of 1998.7 was determined by calculating the average year in which schools were in their first year of implementation of Direct Instruction. The value 2003 was used for year 7 because that was the only year in which DI schools had 7 years of implementation.

Table A-10: First Grade Reading Achievement Scores by Group and Years of Implementation, Adjusting for School Advantage, BCPSS

Panel A: Normal Curve Equivalent Scores

	Years of Implementation	NIFDI	Other DI	Control
Comprehension	1	38	40	39
	7	74	62	57
Vocabulary	1	39	41	41
	7	71	54	51
Composite	1	39	41	40
	7	76	61	56

Panel B: Expected Percentile Score of Average Student

	Years of Implementation	NIFDI	Other DI	Control
Comprehension	1	29	32	30
	7	87	72	63
Vocabulary	1	30	34	34
	7	84	58	52
Composite	1	30	34	32
	7	89	70	61

Table A-11 presents the data that were used to construct Figures 6, 7, and 8 in the text. Unlike data in Tables A-9 and A-10, these data are simply the average achievement scores in each year on each measure for each of the three groups. Data for the low poverty school in the Other DI group are presented separately. No other adjustments were included for school advantage/disadvantage, nor are there any adjustments for how long the schools had been implementing DI. They simply compare the average scores in the schools in the four different groups from 1998 to 2003. The F-tests summarize the analysis of variance results testing the null hypothesis that the mean scores of students in the 4 groups are equal. The pair-wise tests, using Tukey post-hoc comparisons, test the hypothesis that each pair of means is equal.[15]

The results show that in the first year for which data are available (1998) students in the NIFDI schools had the lowest average scores of students in any of the four groups. Pair-wise tests show that these differences with students in the control schools and in the low poverty Other DI school were statistically significant on all three measures and differences with students in the high poverty Other DI schools were significant on the measure of

[15] Recall that a number of the schools (all but one of the Other DI schools and four of the NIFDI-supported schools) began using DI in 1996-97, the year before the first year of data in the data set (see Table A-1).

vocabulary. By the second year students in the NIFDI schools had begun to catch up, and by the last two years they had scores that were significantly higher than students in both the Control schools and the high poverty Other DI schools. Students in the low poverty Other DI school had significantly higher scores than students in the other schools on all three measures throughout the scope of the study.

Table A-11: Reading Achievement Scores by Year and Group, First Graders, BCPSS, 1998-200

Panel A: Means(NCE Scores) and Analysis of Variance Results

	Group				4-Way Comparisons	
	NIFDI	HPODI	LPODI	Control	F	prob.
Comprehension						
1998	31	32	47	35	14.59	<.001
1999	37	35	55	38	14.91	<.001
2000	49	50	66	47	16.92	<.001
2001	56	56	73	51	32.854	<.001
2002	63	49	70	53	42.108	<.001
2003	64	50	78	55	49.59	<.001
Percent Increase	106	56	66	57		
Vocabulary						
1998	30	35	52	39	26.78	<.001
1999	39	39	58	44	19.45	<.001
2000	43	44	66	46	19.99	<.001
2001	53	49	69	48	28.457	<.001
2002	57	46	64	50	26.46	<.001
2003	59	46	71	50	45.39	<.001
Percent Increase	97	31	37	28		
Composite						
1998	30	33	50	37	21.286	<.001
1999	38	37	57	41	16.98	<.001
2000	49	48	70	48	19.20	<.001
2001	57	54	75	51	35.044	<.001
2002	63	50	71	54	39.9	<.001
2003	64	50	78	54	55.29	<.001
Percent Increase	113	52	56	46		

Panel B: Pair-Wise Comparisons (sig.)

	NIFDI v. HPODI	NIFDI v. LPODI	NIFDI v. Control	HPODI v. LPODI	HPODI v. Control	LPODI v.Control

Comprehension						
1998	1.00	<.001	0.003	<.001	0.01	<.001
1999	0.43	<.001	0.77	<.001	0.07	<.001
2000	0.93	<.001	0.14	<.001	0.13	<.001
2001	1.00	<.001	<.001	<.001	0.003	<.001
2002	<.001	0.08	<.001	<.001	0.12	<.001
2003	<.001	<.001	<.001	<.001	0.05	<.001
Vocabulary						
1998	0.04	<.001	<.001	<.001	0.02	<.001
1999	1.00	<.001	<.001	<.001	0.004	<.001
2000	1.00	<.001	0.07	<.001	0.50	<.001
2001	0.14	<.001	<.001	<.001	0.91	<.001
2002	<.001	0.066	<.001	<.001	0.18	<.001
2003	<.001	<.001	<.001	<.001	0.07	<.001
Composite						
1998	0.48	<.001	<.001	<.001	0.01	<.001
1999	0.95	<.001	0.011	<.001	0.04	<.001
2000	1.00	<.001	0.97	<.001	1.00	<.001
2001	0.48	<.001	<.001	<.001	0.22	<.001
2002	<.001	0.05	<.001	<.001	0.13	<.001
2003	<.001	<.001	<.001	<.001	0.03	<.001

Panel C: Percentile Equivalents of Normal Curve Equivalent Scores

	NIFDI	HPODI	LPODI	Control
Comprehension				
1998	18.3	19.6	44.3	23.8
1999	26.8	23.8	59.6	28.6
2000	48.2	50.0	77.6	44.3
2001	61.2	61.2	86.2	51.8
2002	73.2	48.2	82.9	55.7
2003	74.1	50.0	90.8	59.6
Vocabulary				
1998	17.1	23.8	53.8	30.0
1999	30.0	30.0	64.8	38.8
2000	37.0	38.8	77.6	42.4
2001	55.7	48.2	81.7	46.2
2002	63.0	42.4	74.1	50.0
2003	66.5	42.4	84.1	50.0
Composite				
1998	17.1	21.0	50.0	26.8
1999	28.6	26.8	63.0	33.5
2000	48.2	46.2	82.9	46.2
2001	63.0	57.6	88.3	51.8

2002	73.2	50.0	84.1	57.6
2003	74.1	50.0	90.8	57.6

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 each measure. The percentage change in average scores was much stronger, often more than twice as great, for students in the NIFDI-supported schools than for students in the other schools. For instance, with the composite achievement measure, average normal curve equivalent scores for students in the NIFDI-supported schools increased by 113 percent. The comparable figures for students in the other groups ranged from 46 percent to 56 percent. (These data were used to create Figure 9 in the text.)

Finally, Panel C of Table A-11 translates the Normal Curve Equivalent scores in Panel A into the corresponding percentiles. (These percentiles were used to create Figures 6 through 8 in the text.) The percentile scores confirm the pattern that appears in the other analyses. In each of the four groups the average student had substantially higher scores in 2003 than in 1998, but the increases were strongest for students in the NIFDI-supported schools. For instance, on the composite measure of reading achievement the average student in the NIFDI-supported school scored at the 17th percentile, but by 2003 the average student in these schools scored at the 74th percentile. The average student in the high poverty Other DI schools scored at the 21st percentile in 1998 but at the 50th percentile by 2003. The average student in the Control schools had similar gains, from the 27th percentile in 1998 to the 58th percentile in 2003. An average student at the low poverty Other DI school was at the national mean in 1998, but by 2003 was at the 91st percentile. 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.

Research Question 5: What is the Magnitude of the Effect of NIFDI-supported Direct Instruction?

The fifth 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.

Table A-12 gives the d scores comparing average achievement of students in NIFDI-supported schools with those in Other DI and Control schools as well as the comparison of students in Other DI schools with students in Control Schools**.** Effect sizes are given in Panel A for the unadjusted scores and in Panel B for scores calculated with school context controlled. With the unadjusted scores combined over all years, the effect sizes comparing students in NIFDI-supported schools with other students are positive, but below the .25 threshold. The effect sizes for 1998 are negative, indicating that the students in the NIFDI schools had substantially lower achievement than students in the other schools at the beginning of the implementation. However, by 2003, the d scores comparing students in the NIFDI schools with the other two groups surpass the .25 threshold for all measures of achievement. In other words, while the d score at the beginning of the study period indicates an educationally significant disadvantage for students in the NIFDI group, this had totally

reversed by the end of the study period and the students in the NIFDI-supported schools had an educationally significant advantage.

Similar, but even stronger, results occur with the measures derived from the results that are independent of school context (see Panel B). The effect sizes for students in schools that had experienced NIFDI-supported intervention far surpass the threshold of educational significance by the 7th year of implementation, ranging from .50 to .71 for comparisons to the Other DI schools and .72 to .85 for comparisons with the Control Schools. In contrast, the effect sizes involving comparisons of the Other DI schools and the Control Schools never reach the .25 level of educationally meaningful.

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

Table A-12: Effect Sizes, All Measures, Gross Figures and Net of School Context, First Grade Reading Achievement, BCPSS, 1998-2003

Panel A: Gross Effect Sizes

		Gross		
		All years	1998	2003
Vocabulary	NIFDI vs. Other DI	0.01	-0.33	0.28
	NIFDI vs. Control	0.08	-0.36	0.39
	Other DI vs. Control	0.07	-0.03	0.11
Comprehension	NIFDI vs. Other DI	0.07	-0.13	0.27
	NIFDI vs. Control	0.20	-0.16	0.39
	Other DI vs. Control	0.13	-0.04	0.12
Composite	NIFDI vs. Other DI	0.06	-0.22	0.29
	NIFDI vs. Control	0.17	-0.26	0.41
	Other DI vs. Control	0.11	-0.03	0.12

Panel B: Effect Sizes Independent of School Context

		All Years	1 year	7 years
Vocabulary	NIFDI vs. Other DI	0.13	-0.08	0.71
	NIFDI vs. Control	0.19	-0.09	0.85
	Other DI vs. Control	0.07	-0.01	0.14
Comprehension	NIFDI vs. Other DI	0.06	-0.08	0.50
	NIFDI vs. Control	0.19	-0.03	0.72
	Other DI vs. Control	0.13	0.04	0.22

Composite	NIFDI vs. Other DI	0.06	-0.08	0.63
	NIFDI vs. Control	0.17	-0.05	0.82
	Other DI vs. Control	0.11	0.02	0.19

Note: Effect sizes for the total "gross" figures were calculated from the raw achievement scores in Table A-5. Those for the "gross figures for the control schools and the NIFDI-supported schools for 1998 and 2003 are in Table A- . Figures for Other DI schools include Roland Park and the other schools. For the composite scores the means were 35.8 for 1998 and 57.1 for 2003, for vocabulary 37.8 for 1998 and 52.4 for 2003, and for comprehension 34.3 for 198 and 57.4 for 2003. Those for the effects "independent of school context" were calculated from the scores in Tables A-9 and A-10. The common standard deviation of 24.0 was used for all calculations.

References

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.

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.

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