

The Long-Term Impact of NIFDI-Supported Implementation of Direct Instruction  
on Reading Achievement: An Analysis of Fifth Graders  
in the Baltimore City Public School System

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

September 2008

Technical Report 2008-02



## Table of Contents

|  | Page |
|--|------|
| Executive Summary  | iii  |
| List of Tables   | iv   |
| List of Figures  | v    |
| Report   | 1    |
| Methodology  | 1    |
| Results  | 3    |
| Summary  | 7    |
| Appendix I: The Sample and Measures  | 8    |
| Measuring Student Achievement  | 8    |
| Composite Measures of School Context   | 8    |
| The Panel Sample   | 11   |
| Appendix II: Detailed Statistical Results  | 17   |
| Analysis   | 17   |
| Research Question One: Did Students Who Received Direct Instruction in First Grade Have Greater Gains in Reading Achievement from First to Fifth Grade than Students in Other Schools?   | 18   |
| Research Question Two: Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools? | 26   |
| Research Question Three. 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?  | 35   |
| References   | 36   |

# The Long-Term Impact of Direct Instruction on Reading Achievement: An Analysis of Fifth Graders in the Baltimore City Public School System

## 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. While all of these schools initially received technical support from the National Institute for Direct Instruction, five eventually ceased this support and used an alternative provider. Beginning in 1998 schools in the district that were not using Direct Instruction adopted Open Court as a standard reading curriculum. Scores were available from the Comprehensive Test of Basic Skills (CTBS) for all first graders and all fifth graders in BCPSS in the spring of 6 school years: 1997-1998 through 2002-2003.

This report examines the achievement of two BCPSS cohorts of students who were enrolled in the Baltimore City Public School System in both first grade and fifth grade during this time: in first grade in 1997-98 and fifth grade in 2001-2003 or in first grade in 1998-99 and fifth grade in 2002-2003. Almost 5,000 students who had data for both years were included in the analysis. The study compared students' achievement in 1) schools with NIFDI support, 2) schools that implemented Direct Instruction without continual 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 and a substantially lower rate of student turnover. Data for this low poverty DI school were also analyzed separately.

Key findings are listed below.

- Students who received Direct Instruction in first grade had significantly higher achievement scores in fifth grade than other students. The results were strongest with the measures of vocabulary and the composite score.
- While students in all schools had higher average achievement in fifth grade than in first grade, the increases were strongest for students in the NIFDI-supported schools and in the low-poverty Other DI school. On average, students in NIFDI-supported schools had a 25 percent gain in their composite reading achievement scores from first grade to fifth grade. Students in the control schools had a gain of only 5 percent.
- The magnitude of changes from first to fifth grade far surpassed the traditional criterion of educationally important effects for both the NIFDI-supported school and the low poverty school in the Other DI group.

## List of Tables

| Title   | Page |
|---|------|
| A-1. Average Race-Ethnic Composition of Schools and Free and Reduced Lunch Levels, Total and by Treatment Group, BCPSS  | 9    |
| A-2. Factor Analysis of School Characteristics, Elementary Schools, BCPSS, 1998-2003  | 10   |
| A-3. Schools in Treatment Groups and Start Dates of Reading Direct Instruction  | 12   |
| A-4. Turnover in the BCPSS, 1997-98, 1998-99, to 2001-02, 2002-03   | 12   |
| A-5. Percentage of Students Who Persisted in the Same Schools from First Grade to Fifth Grade by Group, BCPSS, 1998-2003  | 13   |
| A-6. Comparison of Panel Group to Other First Graders, First Grade Reading Achievement  | 14   |
| A-7. School Characteristics by Presence in Panel Sample and Group   | 15   |
| A-8: Descriptive Statistics and Paired t-test Results, 1 <sup>st</sup> and 5 <sup>th</sup> Grade Reading Achievement, BCPSS, 1998-2003  | 19   |
| A-9: Independent t-tests, Average Reading Achievement, DI and Control Schools, 1 <sup>st</sup> and 5 <sup>th</sup> Grade, BCPSS, 1998-2003  | 20   |
| A-10. Repeated Measures Analysis of Variance, Norm Equivalent Reading Achievement Scores, 1 <sup>st</sup> and 5 <sup>th</sup> Grade, DI Schools versus Control Schools  | 21   |
| A-11. Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Reading Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction           | 22   |
| A-12. Coefficients Associated with Mixed Model Results, Fifth Grade Reading Regressed on First Grade Reading, School Advantage/Disadvantage, and Attending a DI School  | 24   |
| A-13. Simulated Fifth Grade 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 | 25   |
| A-14. Descriptive Statistics and Paired t-test Results, 1 <sup>st</sup> and 5 <sup>th</sup> Grade Reading Achievement Scores, NIFDI-Supported Schools, Other DI Schools, and Control Schools, BCPSS, 1998-2003      | 27   |

|  |    |
|--|----|
| A-15. One-Way Analysis of Variance, Average Reading Achievement, Norm Equivalent Scores, NIFDI, Other DI, and Control Schools, 1 <sup>st</sup> and 5 <sup>th</sup> Grade, BCPSS, 1998-2003   | 29 |
| A-16. Repeated Measures Analysis of Variance, Normal Curve Equivalent Scores, Reading  | 30 |
| A-17. Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Reading Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction, Comparing NIFDI Supported Schools, Other DI Schools (with and without low poverty school) and Control Schools | 31 |
| A-18. Mixed Model Regressions, Fifth Grade Reading (Norm Equivalent Scores) on School Advantage/Disadvantage and School Type   | 32 |
| A-19. Effect Sizes of Change, First to Fifth Grade, by Group Reading Achievement, Norm Equivalent Scores   | 35 |

### List of Figures

|   | Page |
|---|------|
| 1. Vocabulary, Comprehension, and Composite Reading Percentile Achievement Scores of an Average Student, First Grade and Fifth Grade, DI and Control Schools, BCPSS   | 3    |
| 2. Percentage Change in Average Achievement Scores, First to Fifth Grade, DI and Control Schools  | 4    |
| 3. First and Fifth Grade Composite Achievement, NIFDI-Supported, Other DI and Control Schools   | 5    |
| 4. Percent Change in Average Achievement, First to Fifth Grade, NIFDI-Supported, High Poverty Other DI Schools, Low Poverty Other DI Schools, and Control Schools   | 6    |
| 5. Effect Sizes of Change from First to Fifth Grade in Vocabulary, Comprehension and Composite Reading Achievement Scores, NIFDI-Supported, High Poverty Other DI Schools, Low Poverty Other DI School, and Control Schools | 7    |

# The Long-Term Impact of Direct Instruction on Reading Achievement: An Analysis of Fifth Graders in the Baltimore City Public School System

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 mid 1990s, the BCPSS implemented curricular reforms throughout its elementary schools to address this low achievement.

The reforms introduced in Baltimore provided the opportunity for a natural experiment that could compare the effects of different curricula on student achievement. Some of the schools used Direct Instruction for reading instruction, specifically 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. While all of the schools using Direct Instruction started implementation with the National Institute for Direct Instruction (NIFDI), six of these schools ceased involvement with NIFDI after implementation and began to work with another provider. These schools are termed “Other DI schools” in this report.

A very large body of literature, including well designed meta-analyses, has demonstrated the effectiveness of Direct Instruction in promoting high reading achievement. A somewhat smaller literature has addressed the impact of DI instruction over the span of the elementary years. This paper addresses the question of longer-term impact by examining the change in reading achievement for two cohorts of BCPSS students: those who were in first grade in 1997-98 and fifth grade in 2001-02 and those who were in first grade in 1998-99 and fifth grade in 2002-2003. Three questions were addressed: 1) Did students who received Direct Instruction in first grade have greater gains in reading achievement from first to fifth grade than students in other schools? 2) Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools? and 3) 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?

Results are summarized in graphical form in this report, while appendices provide full methodological details and statistical results. A brief overview of the methodology is below followed by the results with each research question.

## Methodology

The analysis uses data from almost 5,000 students who were first graders in the Baltimore City Public School System in 1997-98 or 1998-99 and fifth graders in 2001-2 or 2002-2003. Measures of reading achievement were obtained each spring, using the Comprehensive Test of Basic Skills (CTBS), a widely used 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 students that would have scores lower than a given student. For instance, if a student has a score at the 60<sup>th</sup> percentile, 60 percent of all students had scores that were lower.

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 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. In addition, 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, three quarters to ninety percent of the students had free or reduced lunch.

There was substantial student turnover in the BCPSS in the targeted years. Of students who were in first grade in the years under study, only 27 percent remained in the same school in fifth grade. This percentage was not significantly different between the three treatment groups. However, the low-poverty school in the Other DI group had a substantially higher rate of persistence, with 57 percent of its first graders still in the school at the end of fifth grade. In addition, although there was no difference across the three groups in the percentage of students who remained in the same schools for the five years, the characteristics of those who remained differed significantly between the three groups. Students who remained in the NIFDI-supported schools were more likely than students in the other schools to have lower achievement levels in first grade. Note that these differences would produce conservative estimates of the impact of NIFDI support on changes in achievement, for students in the NIFDI-supported schools began first grade under arguably more difficult socio-demographic circumstances than students in the other schools. (Appendix I provide additional details on the sample and measures.)

The analysis focused on comparing students' achievement in first grade with their achievement in fifth grade and the ways in which these differences varied across the three groups of schools: those that had NIFDI support throughout the elementary careers of the students, those that only had NIFDI support in the early years, and the Control schools that did not implement Direct Instruction. Full details on the statistical techniques and results are given in Appendix II.

Three research questions were addressed:

- Did students who received Direct Instruction in first grade have greater gains in reading achievement from first to fifth grade than students in other schools?
- Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?

- 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?

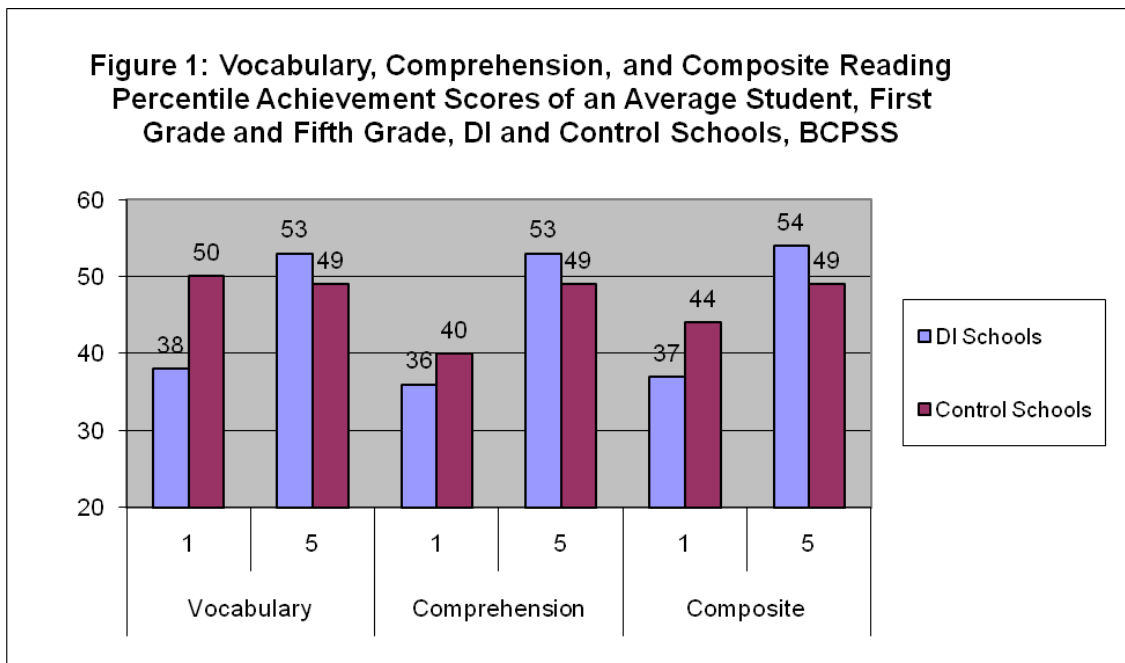
### Results

A general summary of the results related to each of the research questions is given in this section. Each research question is addressed in turn.

#### 1. Did students who received Direct Instruction in first grade have greater gains in reading achievement from first to fifth grade than students in other schools?

Changes in achievement from first to fifth grade were much stronger for students in the DI schools than for students in the Control Schools. Several statistical tests were used to examine this question, and the results confirmed that the changes from first to fifth grade were significantly stronger for students in the DI schools.

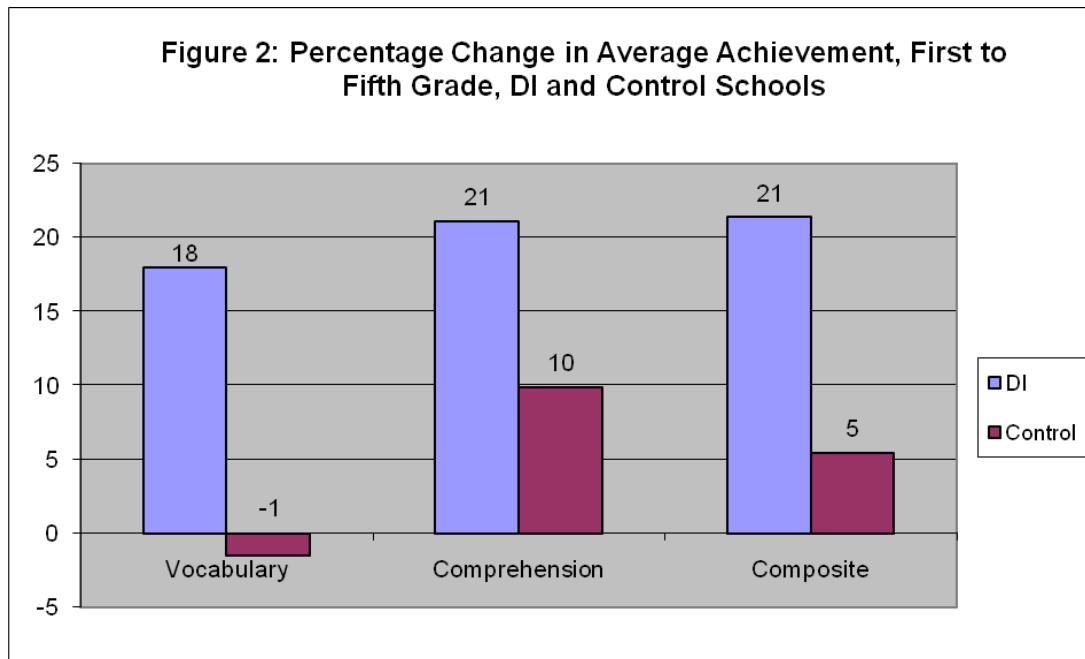
Figure 1 illustrates the magnitude of these results. It displays the percentile scores of the average student in both the Direct Instruction and the Control Schools for both first and fifth grade. In first grade, the average student in a DI school had achievement scores that were significantly lower than those of students in the Control Schools. By fifth grade, this pattern had reversed, and the students in the DI schools had scores that were significantly higher than those of students in the Control school. For instance, an average student in a Direct Instruction school had a composite reading achievement score at the 37th percentile in first grade, but a score at the 54th percentile in fifth grade. In contrast, the average student in the control schools moved from the 44<sup>th</sup> percentile in first grade to the 49<sup>th</sup> percentile in fifth grade.



Note: Figures in this graph represent the percentile score of an average student in a DI school or a Control school on each component of the CTBS in first and fifth grade. The group of DI Schools includes those that were NIFDI-supported as well as both the low poverty and high poverty Other DI schools.



Figure 2 summarizes the extent of these differences by showing the percentage of change in the average score of student from first grade to fifth grade. On average, for students in the DI schools, achievement scores rose from 18 to 21 percent from first to fifth grade. In contrast, although scores for students in the control schools rose on two of the three measures, the magnitude of this increase was less than half that of students in the DI schools.



Note: The data in this figure represent the percentage change in average scores from first to fifth grade, calculated as the difference between first and fifth grade divided by the first grade score. Calculations were made with norm equivalent scores. Appendix II provides further details.

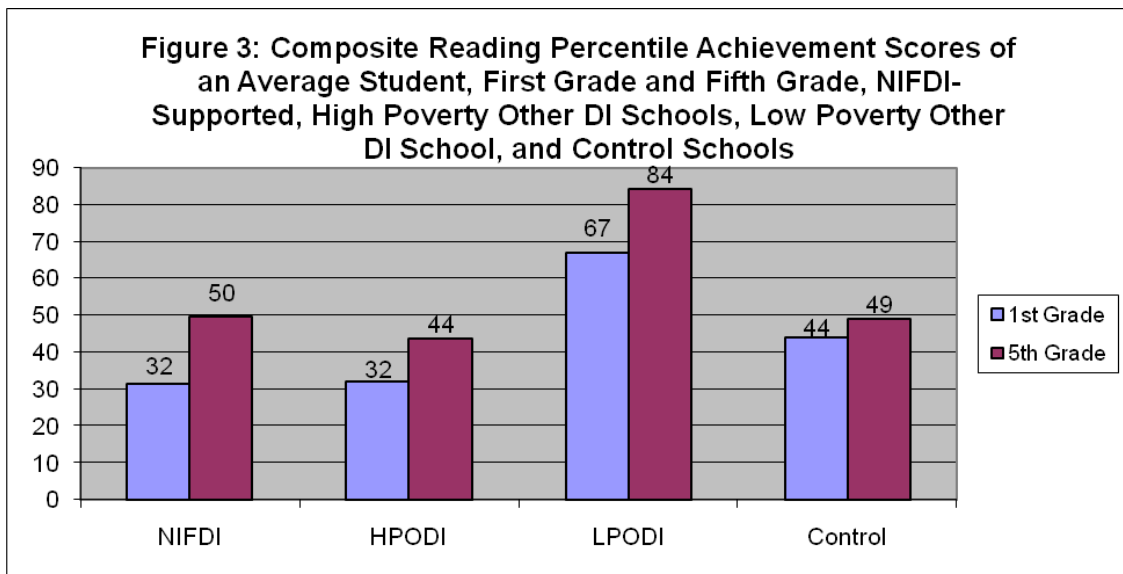
Data in Figures 1 and 2 are unadjusted for students' reading levels in first grade or the schools' contexts of disadvantage. We also used multivariate statistics to predict students' fifth grade achievement based on their achievement in first grade, the poverty context of their schools, and whether or not they had Direct Instruction in first grade. The results confirmed the finding that students who had Direct Instruction in first grade had higher achievement scores in fifth grade than in first grade and that these higher scores occurred independently of the poverty context of the school and their level of first grade achievement. These results were strongly significant for the measures of both vocabulary and the composite score. With the measure of comprehension the results were in the same direction, but just shy of traditional levels of significance.

## **2: Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?**

All of the students in the sample who received DI in first grade were in schools that were receiving technical support from the National Institute of Direct Instruction at that time. However, only students in the NIFDI-supported schools were in schools with this support throughout their elementary career. Various statistical analyses were used to examine differences in gains from first to fifth grade for students in the NIFDI supported schools, the Other DI schools, and the Control Schools.

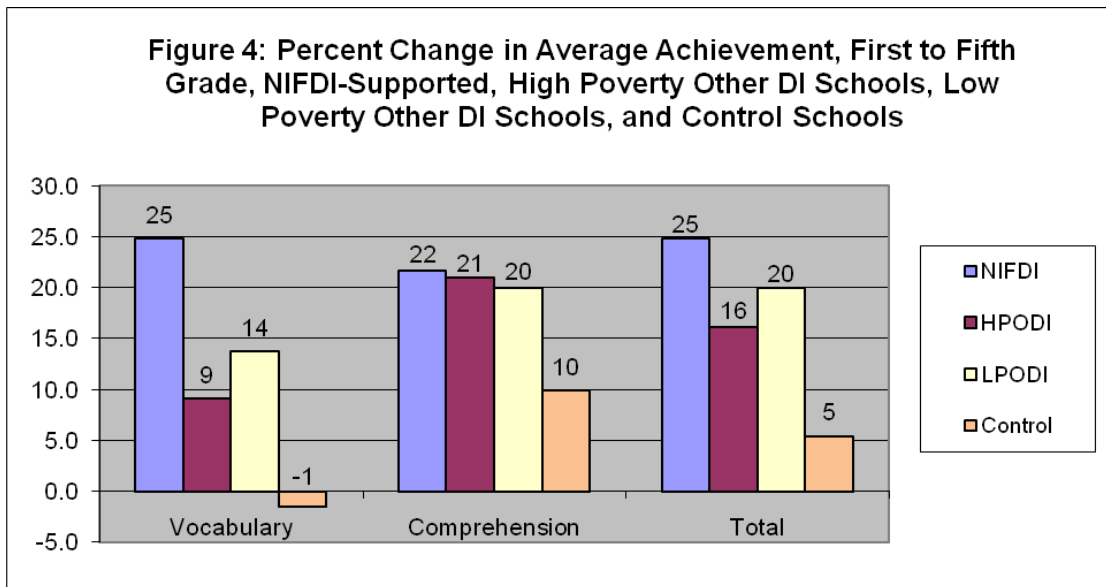
Figure 3 gives the percentile scores on the composite measure of reading achievement in first and fifth grade for an average student in 1) NIFDI supported schools, 2) the high poverty Other DI (HPODI) schools, 3) the low poverty school in the Other DI (LPODI) group, and 4) the control schools. Students in the LPODI had the highest achievement scores in first grade, followed by the students in the Control Schools. An average student in the LPODI school had a composite score at the 67<sup>th</sup> percentile in first grade, while an average student in the control schools had a composite score at the 44<sup>th</sup> percentile. Students in the demographically similar NIFDI-supported and HPODI schools had much lower first grade scores, with the average student scoring at the 32<sup>nd</sup> percentile.

By fifth grade the situation had changed. Average fifth grade achievement scores were higher than average first grade scores in all four groups of schools, and the LPODI school again had the highest achievement score, with an average student scoring at the 84<sup>th</sup> percentile. In contrast to results in first grade, the next highest average scores occurred with students in the NIFDI-supported schools, where an average student scored at the 50<sup>th</sup> percentile. These were followed closely by the Control Schools, where an average student scored at the 49<sup>th</sup> percentile. Students in the HPODI schools had the lowest average scores in fifth grade.



Note: The data in this figure are the percentile score of an average student in each type of school in first grade and fifth grade. HPODI refers to the high poverty Other DI schools, LPODI refers to the low poverty Other DI school.

Figure 4 shows the percentage change in scores from first to fifth grade for each of the four groups. The changes were largest for students in the NIFDI-supported schools for each of the measures of achievement. On the comprehension and composite measures the changes for the both the high poverty and low poverty Other DI schools were close to the values for the NIFDI-supported schools. In contrast, the average percentage change for students in the Control schools was substantially smaller, and even negative for the measure of vocabulary.



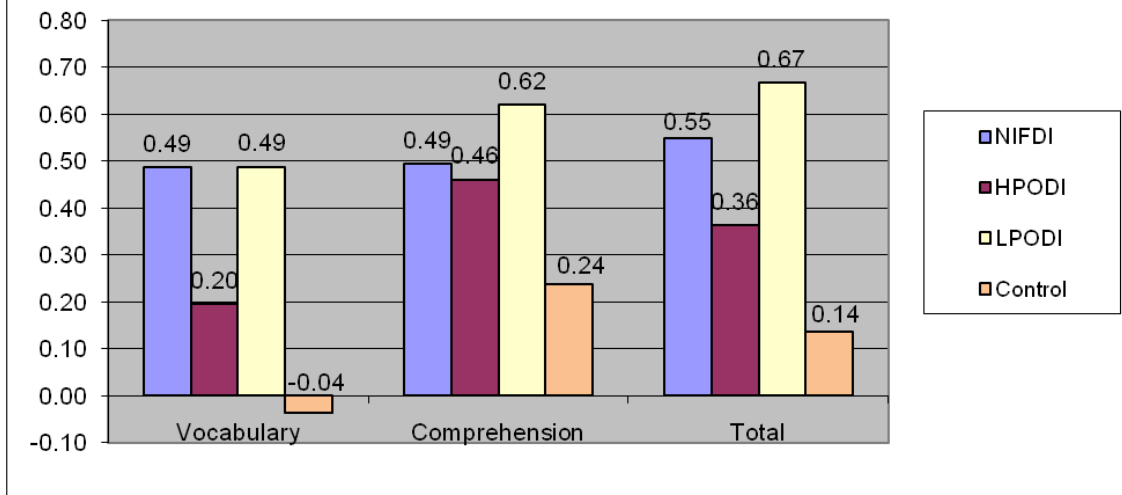
Note: The percentage change values were calculated from the normal curve equivalent scores rather than the percentile scores such as those shown in Figure 3. They are the difference in average first and fifth grade scores divided by first grade scores. Details are in Appendix II. HPODI refers to the high poverty Other DI schools, LPODI refers to the low poverty Other DI school.

### 3. 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?

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 II.) Traditionally, effect sizes of .25 or larger are considered educationally important.

Figure 5 gives the effect sizes associated with the changes from first to fifth grade for each of the measures of achievement and for the four groups considered in the analysis of research question two. All of the effect sizes associated with the NIFDI-supported schools and the LPODI school far surpass the .25 criterion, as do two of the scores for the HPODI school. Effect sizes for change for the NIFDI supported schools range from .49 (vocabulary and comprehension) to .55 (composite score). Those for the LPODI school, which had very low turnover, range from .49 to .67, and those for the HPODI schools range from .20 to .46. The effect sizes for changes for students in the Control Schools are substantially smaller, although the value for Comprehension approaches the .25 criterion ( $d = .24$ ).

**Figure 5: Effect Sizes of Change from First to Fifth Grade in Vocabulary, Comprehension and Composite Reading Achievement Scores, NIFDI-Supported, High Poverty Other DI Schools, Low Poverty Other DI School, and Control Schools**



Note: HPODI refers to the high poverty Other DI schools, LPODI refers to the low poverty Other DI school.

### Summary

Compared to other students in the BCPSS, those who were in schools with Direct Instruction in first grade had significantly higher achievement in 5<sup>th</sup> grade. These differences were both statistically significant and educationally important. On average, achievement test scores for students in DI schools rose from 18 to 21 percent over the 5 grades. In contrast, test scores for students in the Control schools rose by only 10 percent or less, and declined in one area.

While all of the students receiving DI were in schools supported by the National Association for Direct Instruction in first grade, several of those schools later ceased this involvement. Students in schools with NIFDI support throughout their elementary career had the highest percentage change in achievement scores from first to fifth grade. One of the schools that ceased involvement with NIFDI but continued using DI had a relatively low poverty rate and a very low rate of turnover of students throughout the study period. Students in this school had significantly higher achievement scores in both first and fifth grade than students in the other schools. They also had percentage gains in achievement that were second only to students in the NIFDI-supported schools and effect sizes that surpassed those of the NIFDI-supported schools.

The results presented in this analysis are important in showing the long-term impact of Direct Instruction on students' achievement as well as the role of continuous, high quality technical support for schools. Even though the students in the NIFDI-supported schools were in high poverty school environments and had lower first grade achievement than the other students, they had gains in achievement from first grade to fifth grade that were statistically significant and educationally important. The presence of NIFDI-supported Direct Instruction throughout their elementary career appears to have promoted significantly higher rates of learning. This pattern of results appeared in all the measures of reading achievement that were considered.

## Appendix I

### The Measures and Sample

This appendix provides details on the measures of achievement and school context used in the analysis and on the panel sample, the students who were in the same schools in both first and fifth grade.

#### **Measuring Student Achievement**

The BCPSS provided data on students' scores on the Comprehensive Test of Basic Skills for all first and fifth graders from the spring of 1998 to the spring of 2003. The Comprehensive Test of Basic Skills is a widely used standardized measure of academic achievement. The 4<sup>th</sup> edition was administered in the spring of 1998 and 1999 and the 5<sup>th</sup> 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 available and analyzed.

Both raw scores and Normal Curve Equivalent (NCE) scores were provided. 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 of the statistics used in analysis. This is not characteristic of percentiles or raw scores, and thus statisticians urge researchers to analyze data based on NCE scores. All statistical calculations for this report were done using NCE scores.

The meaning of NCE scores is, however, not intuitively obvious, and, for this reason, many of the results in the body of the text and the executive summary have been translated into percentile scores after the statistical calculations were completed. The resulting descriptive information (mean NCE score) was converted into percentiles for display in the graphs in the body of the report and display in the tables in the appendices. These conversions were done using a standard conversion table and are included in the tables of results in Appendix II.<sup>1</sup>

#### **Composite Measures of 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.

---

<sup>1</sup> 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.

predicting 1998 levels of proportion black from levels in 1999 through 2003).<sup>2</sup> Values for each variable were averaged across the years to produce aggregate measures of the demographic characteristics of the school.

Table A-1: 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       | 0    | 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    |
| Factor 2               | 1.18  | 2.82 | -0.34    | 0.25  | -0.11   | 0.47 | 0     | 1    |
|                        | 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                 | 0.7   | 0.68                                | 0.9              | 0.78                |
| Free and Reduced Lunch | 2.06                 | 0.13  | 0.35                                | 0.12             | 0.97                |
| Factor 1               | 0.12                 | 0.89  | 1                                   | 0.89             | 0.98                |
| Factor 2               | 9.76                 | <.001 | 0.009                               | <.001            | 0.85                |

Table A-1 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

<sup>2</sup> 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.

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.

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.<sup>3</sup> Factor scores were computed for each school and saved.

Results for the factor analysis are given in Table A-2. The communalities 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.

Table A-2: 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 Black                       | -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 many fewer Black and low income students. As can be seen in Table A-1, 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

---

<sup>3</sup> 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. N = 120 in the factor analysis for Charles Carroll was included. The results did not differ when it was excluded.

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-1 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 this school 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. In the analysis Roland Park is referred to as a Low Poverty Other DI school (LPODI) and the other schools within this group that received Direct Instruction but without NIFDI support is referred to as the High Poverty Other DI Schools (HPODI).

### **The Panel Sample**

This section provides details on the sample used in the analysis of changes in achievement from first to fifth grade. The potential pool of students included all 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.

Table A-3 lists schools in each of the treatment groups (those that implemented Direct Instruction with NIFDI-support throughout the study period and those that implemented DI but did not retain support from NIFDI throughout the study period) and the date at which the DI Reading program was implemented in each school.<sup>4</sup> In the analysis reported below schools were designated as belonging to a given condition only in 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 who were in first grade in 1998-99 were determined to be in the NIFDI reading intervention group, but those who were in first grade in 1997-98 were in the control group.

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 (53 percent) 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<sup>th</sup> grade in 2001-2 or 2002-3. The panel sample analyzed in this report included a total of 4,771 students. Of these, 355 were in NIFDI-supported Schools, 228 were

---

<sup>4</sup> There were two additional schools in the data set, 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. Both were among the control group of schools.



in Other DI schools, and the remaining 4,188 were in the Control Schools.<sup>5</sup> Six schools in the BCPSS, all within the Control Group, had no students that persisted from first to sixth grade. As a result, the final panel sample included students from 113 schools.

Table A-3: Schools in Treatment Groups and Start Dates of Reading Direct Instruction

**NIFDI Schools**

|                   |         |
|-------------------|---------|
| 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 |

Table A-4: Turnover in the BCPSS, 1997-98, 1998-99 to 2001-02, 2002-03

|   | N      | Percent |
|---|--------|---------|
| Left BCPSS (In first grade, but not in fifth)                                 | 9,362  | 52.8%   |
| Transferred within BCPSS (In BCPSS in fifth grade, but in a different school) | 3,595  | 20.3%   |
| Panel Sample (Same School, first and fifth grade)                             | 4,771  | 26.9%   |
| Total Number of Students In First Grade 1997-98 or 1998-99                    | 17,728 | 100.0%  |

---

<sup>5</sup> Students who were retained were omitted from the analysis. In addition, 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-5 gives the percentage of students who persisted within the same school, that is, who were in the same school for both first and fifth grade, for each of the groups. It can be seen that the persistence was much higher for students within the Low Poverty Other DI school than within the other schools: 57 percent compared to 26 to 28 percent for students in the other schools. When the LPODI school was removed from the analysis (the third column of data in Table A-5) there were no differences in the persistence of students in the groups of schools.

Table A-5: Percentage of Students Who Persisted in the Same Schools from First Grade to Fifth Grade by Group, BCPSS, 1998-2003

|            | % Persisting |       |       | N persisted | Total N |
|------------|--------------|-------|-------|-------------|---------|
| NIFDI      | 28           | 28    | 28    | 280         | 985     |
| Other DI   | 32           | ----- | ----- | 228         | 718     |
| HPODI      | -----        | 26    | 26    | 156         | 591     |
| LPODI      | -----        | 57    | ----- | 72          | 127     |
| Control    | 27           | 27    | 27    | 4263        | 16025   |
| chi-square | 10.49        | 59.3  | 1.61  |             |         |
| Df         | 2            | 2     | 2     |             |         |
| P          | 0.005        | <.001 | 0.445 |             |         |

Note: The percentage of students who did not persist is omitted from the table. This can be simply calculated by subtracting the percentage given from 100.

As would be expected the students in the panel group (those that were in the same schools in 1<sup>st</sup> and 5<sup>th</sup> grade) differed from other students in both achievement and the nature of the schools that they attended. In addition, however, these differences varied from one treatment group to another. Table A-6 compares the first grade reading achievement scores of students in the panel group and the other students (combining those who moved elsewhere in the system and those for whom data for only first grade were available). Analyses of variance were calculated with treatment group (NIFDI, Other DI, and Control schools) and data availability (presence or absence of panel data) as factors. In all cases, the first grade achievement scores were higher for those in the panel sample than for those who moved. However, the differences between the panel sample and those who moved were significantly smaller for students in the NIFDI supported schools than for students in the other schools. (See the significant interaction term for all three achievement measures.) For instance, translating the norm equivalent scores to percentiles, the average student in a NIFDI supported school who was in the panel sample had a composite reading score at the 32<sup>nd</sup> percentile in first grade, compared to a score at the 42<sup>rd</sup> percentile for an average student in the panel who attended an Other DI school and a score at the 44<sup>th</sup> percentile for an average student in the panel who attended a control school. Differences are substantially smaller for the group that is not in the panel sample. (See Section B of Table A-6 for the percentiles that correspond to the norm equivalent scores.) Note that these differences would bias results in a

conservative direction, for the initial (first grade) differences between the treatment groups favor the non-NIFDI-supported groups.

Table A-6: Comparison of Panel Group to Other First Graders, First Grade Reading Achievement

*A: Reading Achievement (Norm Equivalent Scores)*

|                        | Vocabulary |       | Comprehension |       | Composite |       |
|------------------------|------------|-------|---------------|-------|-----------|-------|
|                        | Mean       | s.d.  | Mean          | s.d.  | Mean      | s.d.  |
| <b>NIFDI Supported</b> |            |       |               |       |           |       |
| Non-Panel Sample       | 34         | 21.5  | 33.1          | 19.3  | 33.3      | 19.3  |
| Panel Sample           | 40         | 21.4  | 40.4          | 18.4  | 40        | 18.3  |
| <b>Other DI</b>        |            |       |               |       |           |       |
| Non-Panel Sample       | 34.7       | 20.5  | 31.1          | 19.1  | 32.8      | 18.7  |
| Panel Sample           | 48         | 21.9  | 45            | 20.9  | 46.3      | 20.4  |
| <b>Control Schools</b> |            |       |               |       |           |       |
| Non-Panel Sample       | 37.4       | 32.2  | 33.5          | 20.6  | 35.2      | 21.1  |
| Panel Sample           | 50.1       | 21.6  | 44.8          | 20.2  | 47.1      | 19.9  |
| <b>ANOVA Results</b>   | F          | P     | F             | p     | F         | P     |
| Panel vs. Non-Panel    | 172.9      | <.001 | 219.25        | <.001 | 205.3     | <.001 |
| Treatment Group        | 35.5       | <.001 | 5.98          | 0.003 | 19        | <.001 |
| Interaction            | 8.1        | <.001 | 4.99          | 0.007 | 6.7       | 0.001 |

*B: Corresponding Percentile Scores*

|                        | Vocabulary | Comprehension | Composite |
|------------------------|------------|---------------|-----------|
| <b>NIFDI Supported</b> |            |               |           |
| Non-Panel Sample       | 22         | 21            | 21        |
| Panel Sample           | 32         | 32            | 32        |
| <b>Other DI</b>        |            |               |           |
| Non-Panel Sample       | 24         | 18            | 21        |
| Panel Sample           | 46         | 41            | 42        |
| <b>Control Schools</b> |            |               |           |
| Non-Panel Sample       | 27         | 22            | 24        |
| Panel Sample           | 50         | 41            | 44        |

Table A-7 compares the measures of school context between the panel and non-panel groups of students and the groups of schools. As explained more fully in Appendix I, higher scores on Factor 1 indicate a school with more non-Hispanic white students and fewer students who receive free and reduced lunch. A higher score on Factor 2 indicates a school with more Hispanic and Native American students. Note that data are included for the Other DI schools as a group and differentiating the HPODI and LPODI schools. The results of three different analyses of variance are included for both dependent variables: one comparing the three groups (NIFDI-supported, Other DI Schools, and the Control Schools), one comparing the four groups (NIFDI-supported, HPODI Schools, the LPODI School, and the Control Schools), and one comparing three groups (omitting the LPODI School).

The F-test associated with the interaction term is significant in all comparisons, indicating that differences between those in the panel sample and not in the panel sample varies from one group to another. Students in the panel sample were more likely to attend schools with higher scores on Factor 1, indicating a school with more non-Hispanic white children and fewer low income and African American children. Yet, when the data for the students in the Other DI group were disaggregated by the poverty status of the schools it can be seen that the differences in this group are highly influenced by the heterogeneity of the category. The results with the analysis of Factor 2 indicate that the tendency for the students in the NIFDI-supported schools to be in schools with more Hispanic and Native American students is strengthened within the panel sample. These results indicate, again, the importance of controlling for school characteristics in the analysis.

Table A-7: School Characteristics by Presence in Panel Sample and Group

|  | Factor 1(White and Higher Income) |       | Factor 2 (Hispanic and Native American) |       | N        |
|--|-----------------------------------|-------|---|-------|----------|
|  | Mean                              | s.d.  | Mean                                    | s.d.  |          |
| <b>NIFDI Supported</b>                             |                                   |       |   |       |          |
| Non-Panel Sample                                   | -0.06                             | 0.91  | 1.34                                    | 2.71  | 705      |
| Panel Sample                                       | 0.17                              | 1.07  | 1.76                                    | 2.89  | 280      |
| Total  | 0.01                              | 0.96  | 1.46                                    | 2.77  | 985      |
| <b>Other DI</b>                                    |                                   |       |   |       |          |
| Non-Panel Sample                                   | -0.28                             | 0.75  | -0.29                                   | 0.18  | 490      |
| Panel Sample                                       | 0.21                              | 1.10  | -0.41                                   | 0.25  | 228      |
| Total  | -0.12                             | 0.91  | -0.33                                   | 0.21  | 718      |
| <b>HPODI</b>                                       |                                   |       |   |       |          |
| Non-Panel Sample                                   | -0.55                             | 0.08  | -0.23                                   | 0.07  | 435      |
| Panel Sample                                       | -0.53                             | 0.09  | -0.24                                   | 0.07  | 156      |
| Total  | -0.54                             | 0.08  | -0.23                                   | 0.07  | 591      |
| <b>LPODI</b>                                       |                                   |       |   |       |          |
| Non-Panel Sample                                   | 1.82                              | 0.00  | -0.77                                   | 0.00  | 55       |
| Panel Sample                                       | 1.82                              | 0.00  | -0.77                                   | 0.00  | 72       |
| Total  | 1.82                              | 0.00  | -0.77                                   | 0.00  | 127      |
| <b>Control Schools</b>                             |                                   |       |   |       |          |
| Non-Panel Sample                                   | -0.13                             | 0.85  | -0.10                                   | 0.48  | 11762    |
| Panel Sample                                       | 0.14                              | 1.09  | -0.15                                   | 0.46  | 4263     |
| Total  | -0.06                             | 0.93  | -0.12                                   | 0.47  | 16025    |
| <b>ANOVA Results</b>                               |                                   |       |   |       |          |
|  | F                                 | p     | F                                       | p     | df       |
| <b>Comparison 1 (NIFDI, Other DI, and Control)</b> |                                   |       |   |       |          |
| Panel vs. Non-Panel                                | 100.44                            | <.001 | 8.40                                    | 0.004 | 1, 17722 |
| Treatment Group                                    | 1.80                              | 0.16  | 1743.22                                 | <.001 | 2, 17722 |
| Interaction  | 7.52                              | 0.01  | 33.30                                   | <.001 | 2, 17722 |
| <b>Comparison 2 (4 groups)</b>                     |                                   |       |   |       |          |
| Panel vs. Non-Panel                                | 7.21                              | 0.01  | 4.37                                    | 0.04  | 1, 17720 |
| Treatment Group                                    | 225.11                            | <.001 | 1177.85                                 | <.001 | 3, 17720 |
| Interaction  | 3.98                              | 0.01  | 21.54                                   | <.001 | 3, 17720 |

**Comparison 3 (3 groups, omitting LPODI)**

|                     |       |       |         |       |          |
|---------------------|-------|-------|---------|-------|----------|
| Panel vs. Non-Panel | 23.55 | <.001 | 14.25   | <.001 | 1, 17595 |
| Treatment Group     | 82.53 | <.001 | 1701.09 | <.001 | 2, 17595 |
| Interaction         | 4.63  | 0.01  | 32.07   | <.001 | 2, 17595 |

## Appendix II: Detailed Statistical Results

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.

### Analysis

The analysis of data focused on changes in students' 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 reading achievement from first to fifth grade than students in other schools?" Several comparisons were used to examine this question. First, the average change from first grade to fifth grade was computed for each of the three measured areas and in both groups (DI and Control Schools). Simple paired t-tests were used to examine changes from first to fifth grade. If students in DI schools had greater gains than those in the Control schools, the t-values associated with their average change would be expected to be larger. Second, to provide more insight into the results, 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. It was expected that results would favor the students in the DI schools much more strongly in fifth grade than in first 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. Finally, the first research 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). This analysis is important to control for possible influences of school context and initial differences in achievement on the results. Successively more complex models were used and differences between these models were examined. It was expected that students in the DI schools would continue to have an advantage in fifth grade achievement when these controls were introduced.

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 second research question was, "Do students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?" It should be recalled that all of the students who received DI instruction were in schools supported by NIFDI in first grade. The Other DI schools, however, ceased this involvement soon after the first grade year of students in the panel sample. To analyze this question three treatment groups were used in the

comparisons: NIFDI supported schools, Other DI schools and the Control schools. In addition, within the Other DI group results for the higher income school were differentiated from those for the Other DI schools. The various techniques used to analyze the first research question were also used with this question, except that, because of the number of comparisons involved, one-way analysis of variance was used instead of independent t-tests.

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 unaffected 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.

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

Four different approaches were used to address this question: paired t-tests, independent sample t-tests, repeated measures analysis of variance, and mixed-model regressions.

#### **Difference Scores and Paired t-tests**

Table A-8 summarizes the data used to examine differences in achievement gains from first to fifth grade for students in the Direct Instruction (whether or not they were in NIFDI-supported schools) and Control Schools. Panel A gives descriptive statistics for each measure of reading 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 1 in the text.

Results indicate that the changes in achievement were much larger for students in the DI schools than in the Control Schools. Students in the DI schools had norm equivalent achievement scores in fifth grade that were significantly higher than their scores in first grade on all three dimensions. Students in the Control schools had comprehension and composite scores that were significantly higher in fifth grade, but vocabulary scores that were significantly lower. Examination of the average difference (Panel B) shows that the changes were much larger for students in the DI schools than in the Control Schools.

Table A-8: Descriptive Statistics and Paired t-test Results, 1st and 5th Grade Reading Achievement, BCPSS, 1998-2003

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

|               |                 | DI Schools |      | Control Schools |      |
|---------------|-----------------|------------|------|-----------------|------|
|               |                 | Mean       | S.D. | Mean            | S.D. |
| Vocabulary    | 1 <sup>st</sup> | 43.7       | 22   | 50.1            | 21.6 |
|               | 5 <sup>th</sup> | 51.5       | 19.9 | 49.4            | 17.5 |
|               | % gain          | 18         |      | -1.5            |      |
| Comprehension | 1 <sup>st</sup> | 42.5       | 19.7 | 44.9            | 20.2 |
|               | 5 <sup>th</sup> | 51.4       | 18.1 | 49.3            | 16.8 |
|               | % gain          | 21.1       |      | 9.9             |      |
| Composite     | 1 <sup>st</sup> | 42.8       | 19.5 | 47.1            | 19.9 |
|               | 5 <sup>th</sup> | 51.9       | 19.1 | 49.7            | 16.9 |
|               | % gain          | 21.4       |      | 5.4             |      |

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

|                        | Ave. Dif. | R    | T     | p     | N    |
|------------------------|-----------|------|-------|-------|------|
| <b>DI Schools</b>      |           |      |       |       |      |
| Vocabulary             | 7.8       | 0.54 | 8.70  | <.001 | 497  |
| Comprehension          | 9         | 0.53 | 10.89 | <.001 | 501  |
| Composite              | 9.1       | 0.62 | 12.03 | <.001 | 495  |
| <b>Control Schools</b> |           |      |       |       |      |
| Vocabulary             | -0.7      | 0.47 | -2.31 | 0.021 | 4172 |
| Comprehension          | 4.5       | 0.52 | 15.67 | <.001 | 4207 |
| Composite              | 2.6       | 0.56 | 9.48  | <.001 | 4136 |

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

|               |                 | DI Schools | Control Schools |
|---------------|-----------------|------------|-----------------|
|               |                 | Vocabulary | 38              |
|               | 5 <sup>th</sup> | 53         | 49              |
| Comprehension | 1 <sup>st</sup> | 36         | 40              |
|               | 5 <sup>th</sup> | 53         | 49              |
| Composite     | 1 <sup>st</sup> | 37         | 44              |
|               | 5 <sup>th</sup> | 54         | 49              |

The numbers in Panel C illustrate the magnitude of these changes. An average student in a DI school scored from the 36<sup>th</sup> to 38<sup>th</sup> percentile on all three measures in first grade, but by 5<sup>th</sup> grade, the average student scored above the national norms, at the 53<sup>rd</sup> or 54<sup>th</sup> percentile. In contrast, although doing better on two of the three measures in 5<sup>th</sup> grade than in 1<sup>st</sup> grade, the average student in the Control school still had scores just below the national norm (at the 49<sup>th</sup> percentile) on all measures. The lines in panel A labeled percent gain also illustrate the magnitude of the changes over time. These were calculated from the norm



equivalent scores and are simply the percentage change from first to fifth grade. It can be seen that the values are much larger, on average, for students in the DI schools than students in the Control schools. These data were used to create Figure 2 in the text.

### Independent t-tests

Table A-9 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. In first grade students in the Control schools had norm equivalent scores that were significantly higher than those of students in the DI schools. By fifth grade, however, this had reversed, with students in the Control schools having scores that were significantly lower than those of students in the DI schools. (See the mean scores in Table A-8.)

Table A-9: Independent t-tests, Average Reading Achievement, DI and Control Schools, 1st and 5<sup>th</sup> Grade, BCPSS, 1998-2003

|                    | t-value | df   | P     |
|--------------------|---------|------|-------|
| <b>First Grade</b> |         |      |       |
| Vocabulary         | 6.34    | 4711 | <.001 |
| Comprehension      | 2.51    | 4713 | 0.01  |
| Composite          | 4.62    | 4676 | <.001 |
| <b>Fifth Grade</b> |         |      |       |
| Vocabulary         | -2.3    | 600  | 0.02  |
| Comprehension      | -2.45   | 614  | 0.02  |
| Composite          | -2.57   | 602  | 0.01  |

Note: F tests indicated that the t-value assuming unequal variances between the groups should be used for all comparisons with fifth grade scores. The degrees of freedom reported are adjusted for that result. 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-10 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-8 and A-9. If exposure to Direct Instruction in the first grade is related to higher achievement in fifth grade, a significant interaction effect would be expected. The results are as expected. The interaction effect is significant for each of the measures of reading achievement, indicating that the change from grade 1 to grade 5 is significantly different for students in the DI schools than in the Control schools. As noted above, the changes were greater for students in the DI schools than in the Control schools.

Table A-10: Repeated Measures Analysis of Variance, Norm Equivalent Reading Achievement Scores, 1st and 5th Grade, DI Schools versus Control Schools

|             | Vocabulary | Comprehension | Composite |
|-------------|------------|---------------|-----------|
| F – Reading | 54.24      | 237.71        | 202.02    |
| P           | <.001      | <.001         | <.001     |
| F- Group    | 7.1        | 0.038         | 1.79      |
| P           | 0.008      | 0.84          | 0.18      |
| Interaction | 78.76      | 26.91         | 63.86     |
| P           | <.001      | <.001         | <.001     |
| df          | 1, 4668    | 1, 4707       | 1, 4630   |

Note: The F test associated with reading 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 reading score of students in the DI schools and that of students in the Control schools are equal. The F associated with the Interaction tests the hypothesis that the differences between groups of schools are the same for first and fifth grade or, alternatively, that the differences between first and fifth grade are the same for the two groups of schools.

### Mixed Model Regressions

Tables A-11 and A-12 report the results of our final method of testing differences in achievement gains between the two groups of schools: the mixed model regressions. Panel A of Table A-11 summarizes the models that were used and Panel B gives summary statistics associated with each model. 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 reading 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, depending on the dependent measure, between 7 and 9 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,

Table A-11, Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Reading 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*

|                      |        | Model 1 |        |       | Model 2 |        |       | Model 3 |        |  |
|----------------------|--------|---------|--------|-------|---------|--------|-------|---------|--------|--|
|                      | Effect | s.e.    | p.     | b     | s.e.    | p.     | b     | s.e.    | p.     |  |
| <b>Vocabulary</b>    |        |         |        |       |         |        |       |         |        |  |
| Random Effects Est.  | 28.6   | 4.9     | <.0001 | 23.76 | 4.2     | <.0001 | 21.4  | 3.8     | <.0001 |  |
| Residual Estimate    | 286.6  | 6       | <.0001 | 218.8 | 4.6     | <.0001 | 218.7 | 4.6     | <.0001 |  |
| BIC                  | 40347  |         |        | 38621 |         |        | 38615 |         |        |  |
| - 2 Log Likelihood   | 40333  |         |        | 38593 |         |        | 38583 |         |        |  |
| Change in LL         |        |         |        | 1740  |         |        | 10    |         |        |  |
| Df                   |        |         |        | 3     |         |        | 1     |         |        |  |
| P                    |        |         |        | <.001 |         |        | <.01  |         |        |  |
| Correlation ratio    | 0.09   |         |        |       |         |        |       |         |        |  |
| <b>Comprehension</b> |        |         |        |       |         |        |       |         |        |  |
| Random Effects Est.  | 19.9   | 3.6     | <.0001 | 17.0  | 3.0     | <.0001 | 16.0  | 2.9     | <.0001 |  |
| Residual Estimate    | 266.4  | 5.5     | <.0001 | 193.3 | 4.0     | <.0001 | 193.4 | 4.0     | <.0001 |  |
| BIC                  | 40299  |         |        | 38343 |         |        | 38345 |         |        |  |
| - 2 Log Likelihood   | 40285  |         |        | 38315 |         |        | 38312 |         |        |  |
| Change in LL         |        |         |        | 1970  |         |        | 2.5   |         |        |  |
| Df                   |        |         |        | 3     |         |        | 1     |         |        |  |
| P                    |        |         |        | <.001 |         |        | >.10  |         |        |  |
| Correlation ratio    | 0.07   |         |        |       |         |        |       |         |        |  |
| <b>Composite</b>     |        |         |        |       |         |        |       |         |        |  |
| Random Effects Est.  | 25.6   | 4.4     | <.0001 | 22.2  | 3.8     | <.0001 | 20.1  | 3.5     | <.0001 |  |
| Residual Estimate    | 269.1  | 5.6     | <.0001 | 177.9 | 3.7     | <.0001 | 178.0 | 3.7     | <.0001 |  |
| BIC                  | 40012  |         |        | 37363 |         |        | 37360 |         |        |  |
| - 2 Log Likelihood   | 39998  |         |        | 37335 |         |        | 37327 |         |        |  |
| Change in LL         |        |         |        | 2663  |         |        | 7.5   |         |        |  |
| Df                   |        |         |        | 3     |         |        | 1     |         |        |  |
| P                    |        |         |        | <.001 |         |        | <.01  |         |        |  |
| Correlation ratio    | 0.09   |         |        |       |         |        |       |         |        |  |

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 other 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.

Two other hypotheses were tested. The first was that the impact of first grade achievement on fifth grade achievement would vary significantly between the two groups of schools. This was tested by adding the interaction of school group (DI versus control schools) with first grade scores to the variables in Model 3. The second hypothesis was that the impact of school advantage/disadvantage would differ between the two groups. This was tested by adding the interaction of school group with factor 1 to Model 3. Neither of these interaction effects was significant either when added singly or when added jointly with any of the measures of reading achievement. Thus, those results are not included in Tables A-11 or A-12.

The -2 log likelihood measures and the BIC values in Panel B of Table A-11 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 in Table A-8. 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 = 1970 (40,285-38,315). With three degrees of freedom (because three new variables were added to Model 2 compared to Model 1), this result is highly significant. The comparison of Model 2 with Model 1 indicates that adding the variables in the more complex model significantly improves the fit with all three measures. (See the significance associated with the change in the -2 log likelihood.) However the significance of the change from Model 2 to Model 3 does not reach statistical significance for the analysis of reading comprehension. The BIC values provide a descriptive summary of the fit of the models, with lower values indicating a better fit. The lowest BIC values appear for Model 3 for the analyses of vocabulary and the composite scores, but for Model 2 for the analysis of comprehension scores.

Table A-12 gives the coefficients associated with variables in Models 2 and Model 3 for each of the measures of achievement. 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 achievement scores in first grade also have higher achievement scores in fifth grade. The coefficients associated with the two factor scores are also as expected, although only those associated with Factor 1 are statistically significant. Students in schools with higher scores on Factor 1 (indicating higher proportions of white students and fewer low income and African American students) had significantly higher achievement scores. Students in schools with higher scores on Factor 2 (indicating more Hispanic and Native American students) had lower achievement scores, although this did not reach traditional levels of significance.

Table A-12: Coefficients Associated with Mixed Model Results, Fifth Grade Reading Regressed on First Grade Reading, School Advantage/Disadvantage, and Attending a DI School

|                       | Model 2 |       |        | Model 3 |      |        |
|-----------------------|---------|-------|--------|---------|------|--------|
|                       | b       | s.e.  | p.     | b       | s.e. | p.     |
| <b>Vocabulary</b>     |         |       |        |         |      |        |
| Intercept             | 29.6    | 0.7   | <.0001 | 29      | 0.7  | <.0001 |
| First grade score     | 0.4     | 0.01  | <.0001 | 0.4     | 0.01 | <.0001 |
| Factor 1              | 1.3     | 0.5   | 0.0081 | 1.3     | 0.5  | 0.005  |
| Factor 2              | -0.5    | 0.5   | 0.3442 | -1      | 0.5  | 0.061  |
| Had DI in First Grade | -----   | ----- | -----  | 4.9     | 1.5  | 0.0009 |
| <b>Comprehension</b>  |         |       |        |         |      |        |
| Intercept             | 29.2    | 0.7   | <.0001 | 29      | 0.7  | <.0001 |
| First grade score     | 0.4     | 0.01  | <.0001 | 0.4     | 0.01 | <.0001 |
| Factor 1              | 1       | 0.4   | 0.02   | 1.1     | 0.4  | 0.01   |
| Factor 2              | -0.1    | 0.4   | 0.83   | -0.3    | 0.5  | 0.5    |
| Had DI in First Grade | -----   | ----- | -----  | 2.1     | 1.3  | 0.1    |
| <b>Composite</b>      |         |       |        |         |      |        |
| Intercept             | 25.9    | 0.7   | <.0001 | 25.4    | 0.7  | <.0001 |
| First grade score     | 0.5     | 0.01  | <.0001 | 0.5     | 0.01 | <.0001 |
| Factor 1              | 1.2     | 0.5   | 0.01   | 1.2     | 0.5  | 0.01   |
| Factor 2              | -0.3    | 0.5   | 0.56   | -0.7    | 0.5  | 0.18   |
| Had DI in First Grade | -----   | ----- | -----  | 3.8     | 1.4  | 0.01   |

As expected, the coefficient associated with attending a DI school in first grade is positive in all cases. The results with both the measure of vocabulary and the composite score are statistically significant. Independent of students' first grade reading score and characteristics of their school context, students who had Direct Instruction in first grade had higher vocabulary and composite achievement scores in fifth grade. However, in the analysis of comprehension scores, the result does not reach traditional levels of statistical significance ( $p = .10$ ).

Table A-13 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-9 for these calculations. Results are given both in the norm equivalent scores and as the corresponding percentile score for a student scoring at this average. It is clear that, after equalizing for first grade scores and school context, BCPSS students who had Direct Instruction in first grade would be expected to score at or above the national average in fifth grade, while students in the control schools would have scores that were still substantially below the national norm.

Table A-13: Simulated Fifth Grade 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

| <b>Norm Equivalent Scores</b>    | <b>Direct Instruction</b> | <b>Control Schools</b> |
|----------------------------------|---------------------------|------------------------|
| Vocabulary                       | 51                        | 46                     |
| Comprehension                    | 50                        | 48                     |
| Composite                        | 51                        | 47                     |
| <b>Corresponding Percentiles</b> |                           |                        |
| Vocabulary                       | 51                        | 42                     |
| Comprehension                    | 50                        | 46                     |
| Composite                        | 51                        | 44                     |

Note: Scores were calculated by substituting the average values of first grade achievement and factor scores in the equation for Model 3 in Table A-12. For Vocabulary the value was 43.6, for comprehension the value was 42.5, for the composite score the value was 42.8. The value for Factor 1 was .19 and the value for Factor 2 was .79. The values for the Factors were not equal to zero both because the panel sample had a slightly different sample of schools than the group for which the factor scores were calculated and, more importantly, because the unit of analysis for calculating the means was students rather than schools.

### **Summary**

To summarize, four different methods were used to examine the first research question, which asked if students who were in DI schools in first grade had higher achievement scores in fifth grade than students in the control schools. First, paired (dependent sample) t-tests were used to compare students' scores in first and fifth grade. As expected, changes from first to fifth grade were much larger for students in the DI schools than in the Control Schools. Second, independent t-tests compared scores of students in the DI schools and students in the Control schools in both first and fifth grade. While students in the Control Schools had significantly higher scores in first grade, students in the DI schools had significantly higher scores in fifth grade. Third, a repeated measures analysis of variance, with achievement as a repeated measure and group as factor, resulted in significant interaction effects, again reflecting the larger gains from first to fifth grade for the students in the DI schools. Finally, mixed model regressions predicted fifth grade scores, controlling for students' first grade scores and characteristics of the schools. As expected, the coefficient associated with attending a DI school in first grade was positive in all cases and statistically significant for the measures of vocabulary and the composite measure.

## **Research Question Two: Did students who were in schools with consistent support from the National Institute of Direct Instruction have greater reading achievement gains from first to fifth grade than students in other schools?**

Four analysis techniques were used to address the second research question: paired t-tests within each group, one-way analysis of variance, repeated measures analysis of variance, and mixed model regressions. Note that this analysis differs from that with Research Question 1 by differentiating students in NIFDI-supported schools from students in the Other DI schools.

### **Descriptive Statistics and Paired t-tests**

Table A-14 summarizes the data used to examine differences in achievement gains from first to fifth grade for students in the NIFDI supported schools, the Other DI schools, and the Control Schools. Scores are also presented separately for students in the low poverty Other DI (LPODI) school. This school had an extraordinarily higher rate of persistence from first to fifth grade and a much lower poverty rate than the other schools. It should also be recalled that all of the students studied in schools with NIFDI support during first grade, but only students in the NIFDI-supported schools had this support throughout their elementary career.

Panel A of Table A-14 gives descriptive statistics for each measure of reading achievement using norm equivalent scores. It also includes the percentage change from first to fifth grade. 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 for each group. Panel C translates the average norm equivalent scores in Panel A into percentile scores. The data in Panel C can be interpreted as the percentile score that the average student obtained in each of the groups of schools in first grade and fifth grade. The data in Panel C were used to construct Figure 3 in the text, and the data on percentage change in Panel A were used to construct Figure 4.

Results indicate, as in Table A-8, that all students had higher norm equivalent scores in fifth grade than in first grade. These changes were significant for students in all groups. Yet, the percentage change in scores was greatest for students in the NIFDI-supported schools for all measures.<sup>6</sup> Values were lowest for students in the control schools for both the measures of absolute change (Panel B) and percentage change (Panel A).

The percentile scores for students scoring at the average (panel C) indicate that, by 5<sup>th</sup> grade, the average student in schools that received NIFDI support throughout their elementary careers scored at the national norm (49<sup>th</sup> or 50<sup>th</sup> percentile) on all three measures, substantially above the level that the average student had scored in 1<sup>st</sup> grade (the 32<sup>nd</sup> and 33<sup>rd</sup> percentile). The average student in the Control Schools also scored at around the national norm at 5<sup>th</sup> grade, but this represented very little change from first grade scores. The average score for the students in the Other DI group translated to percentiles above the mean (56 to 59), but the overall results for this group mask differences between the high poverty schools and the low poverty school. The average student in the low poverty DI school scored well

---

<sup>6</sup> The absolute change in scores was greater for students in the LPODI school for the measure of comprehension and the composite score, but the percentage change was actually lower because these students began at a higher level of achievement.

within the 8<sup>th</sup> decile, while the average student in the remaining Other DI schools scored well below the national norm.

Table A-14: Descriptive Statistics and Paired t-test Results, 1st and 5th Grade Reading Achievement Scores, NIFDI Supported Schools, Other DI Schools, and Control Schools, BCPSS, 1998-2003

*A: Descriptive Statistics*

|               |       | <b>NIFDI</b> |      | <b>Other DI</b> |      | <b>Control</b> |      | <b>HPODI</b> |      | <b>LPODI</b> |      |
|---------------|-------|--------------|------|-----------------|------|----------------|------|--------------|------|--------------|------|
|               |       | Mean         | S.D. | Mean            | S.D. | Mean           | S.D. | Mean         | S.D. | Mean         | S.D. |
| Vocabulary    | 1st   | 40           | 22   | 48              | 22   | 50             | 22   | 42           | 21   | 62           | 19   |
|               | 5th   | 50           | 19   | 53              | 21   | 49             | 18   | 46           | 18   | 70           | 16   |
|               | % ch. | 25           |      | 11              |      | -2             |      | 9            |      | 14           |      |
| Comprehension | 1st   | 40           | 17   | 45              | 21   | 45             | 20   | 39           | 19   | 57           | 20   |
|               | 5th   | 49           | 18   | 54              | 19   | 49             | 17   | 48           | 17   | 69           | 15   |
|               | % ch. | 22           |      | 21              |      | 10             |      | 21           |      | 20           |      |
| Composite     | 1st   | 40           | 18   | 46              | 21   | 47             | 20   | 40           | 18   | 59           | 19   |
|               | 5th   | 50           | 18   | 55              | 20   | 50             | 17   | 47           | 17   | 71           | 16   |
|               | % ch. | 25           |      | 18              |      | 5              |      | 16           |      | 20           |      |

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

|                         | Ave. Dif. | r    | t     | p     | N    |
|-------------------------|-----------|------|-------|-------|------|
| <b>NIFDI Schools</b>    |           |      |       |       |      |
| Vocabulary              | 9.9       | 0.44 | 7.61  | <.001 | 275  |
| Comprehension           | 8.7       | 0.42 | 7.62  | <.001 | 277  |
| Composite               | 9.9       | 0.5  | 9.04  | <.001 | 273  |
| <b>Other DI Schools</b> |           |      |       |       |      |
| Vocabulary              | 5.3       | 0.65 | 4.43  | <.001 | 223  |
| Comprehension           | 9.2       | 0.62 | 7.86  | <.001 | 225  |
| Composite               | 8.2       | 0.72 | 7.99  | <.001 | 223  |
| <b>Control Schools</b>  |           |      |       |       |      |
| Vocabulary              | -0.7      | 0.47 | -2.31 | 0.021 | 4172 |
| Comprehension           | 4.5       | 0.52 | 15.67 | <.001 | 4207 |
| Composite               | 2.6       | 0.56 | 9.48  | <.001 | 4136 |
| <b>HPODI</b>            |           |      |       |       |      |
| Vocabulary              | 3.8       | 0.55 | 2.52  | 0.013 | 151  |
| Comprehension           | 8.2       | 0.54 | 5.92  | <.001 | 153  |
| Composite               | 6.5       | 0.64 | 5.29  | <.001 | 151  |
| <b>LPODI</b>            |           |      |       |       |      |
| Vocabulary              | 8.4       | 0.56 | 4.43  | <.001 | 72   |
| Comprehension           | 11.4      | 0.49 | 5.22  | <.001 | 72   |
| Composite               | 11.8      | 0.62 | 6.51  | <.001 | 72   |



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

|               |     | Other |    |         |       |       |
|---------------|-----|-------|----|---------|-------|-------|
|               |     | NIFDI | DI | Control | HPODI | LPODI |
| Vocabulary    | 1st | 32    | 46 | 50      | 35    | 71    |
|               | 5th | 50    | 56 | 49      | 42    | 83    |
| Comprehension | 1st | 33    | 41 | 40      | 30    | 63    |
|               | 5th | 49    | 58 | 49      | 45    | 81    |
| Composite     | 1st | 32    | 43 | 44      | 32    | 67    |
|               | 5th | 50    | 59 | 49      | 44    | 84    |

**One-Way Analyses of Variance**

Table A-15 summarizes the results of one-way analyses of variance with each of the measures of reading achievement. This analysis differs from that above by focusing on comparing scores between the groups of schools in both first grade and then in fifth grade. Two sets of comparisons are included. The first, in Panel A, involves a comparison between the three groups (NIFDI supported schools, Other DI schools, and Control Schools). The second, in Panel B, includes four categories, separating the students in the Other DI group into those in the low poverty school from those in other schools. Results of Tukey post-hoc tests for pair-wise comparisons are also included.

The analysis of variance results indicate that the groups of schools had significantly different average achievement scores in both first grade and in fifth grade. These results occur with both the comparison of three groups (in Panel A) and of four groups (in Panel B). However, the post-hoc tests show that these differences only occur between some pairs. In addition, paralleling the results reported above, the direction of the difference varies markedly from first grade to fifth grade.

In first grade, the students in the NIFDI-supported schools had significantly lower achievement scores than students in either the Other DI or the Control Schools on all three measures (see the first two columns in the post-hoc section of Panel A). The results in Panel B, however, show that this difference occurs only because of the presence of low poverty school in the Other DI group. First grade achievement scores of students in schools supported by NIFDI throughout the span of the study are virtually identical, on average, to those of students in the high poverty Other DI schools.

Results for Fifth Grade achievement indicate that there is no longer any difference between average scores of students in the NIFDI-supported schools and students in the control schools on any of the measures. Students in the NIFDI supported schools have significantly lower scores than students in the Other DI group, but this disappears when data for the LPODI school are analyzed separately. Students in the LPODI school have significantly higher scores than students in any of the other three groups, both in first grade and in fifth grade. Inspection of the means in Table A-14 indicates that students in the NIFDI-supported schools have higher scores than those in the HPODI schools on all three measures in fifth grade, but the results of the pair-wise post-hoc tests, given in Table A-15, indicate that these differences are significant only with the measure of vocabulary.

Table A-15: One-Way Analysis of Variance, Average Reading Achievement, NIFDI, Other DI, and Control Schools, 1st and 5th Grade, BCPSS, 1998-2003

*A. Comparison of NIFDI, Other DI, and Control Schools*

|                    | Analysis of Variance Results |         |       | Tukey Post-Hoc Tests (prob.) |                  |                     |
|--------------------|------------------------------|---------|-------|------------------------------|------------------|---------------------|
|                    | F                            | df      | p     | NIFDI v. Other DI            | NIFDI v. Control | Other DI v. Control |
| <b>First Grade</b> |                              |         |       |                              |                  |                     |
| Vocabulary         | 28.66                        | 2, 4170 | <.001 | <.001                        | <.001            | 0.34                |
| Comprehension      | 6.3                          | 2, 4712 | 0.002 | 0.03                         | 0.001            | 1.00                |
| Composite          | 16.98                        | 2, 4675 | <.001 | 0.001                        | <.001            | 0.81                |
| <b>Fifth Grade</b> |                              |         |       |                              |                  |                     |
| Vocabulary         | 4.91                         | 2, 4724 | 0.007 | 0.16                         | 0.73             | 0.01                |
| Comprehension      | 8.53                         | 2, 4762 | <.001 | 0.004                        | 1.00             | <.001               |
| Composite          | 7.59                         | 2, 4720 | 0.001 | 0.02                         | 0.91             | <.001               |

*B. Comparison of NIFDI, HPODI, LPODI, and Control Schools*

|                    | Analysis of Variance Results |         |       | Tukey Post-Hoc Tests (prob.) |                  |                |                  |                |                  |
|--------------------|------------------------------|---------|-------|------------------------------|------------------|----------------|------------------|----------------|------------------|
|                    | F                            | df      | p     | NIFDI v. HPODI               | NIFDI v. Control | NIFDI v. LPODI | HPODI v. Control | HPODI v. LPODI | LPODI v. Control |
| <b>First Grade</b> |                              |         |       |                              |                  |                |                  |                |                  |
| Vocabulary         | 33                           | 3, 4709 | <.001 | 0.87                         | <.001            | <.001          | <.001            | <.001          | <.001            |
| Comprehension      | 17.24                        | 3, 4711 | <.001 | 0.94                         | 0.002            | <.001          | 0.004            | <.001          | <.001            |
| Composite          | 26.65                        | 3, 4674 | <.001 | 1.00                         | <.001            | <.001          | <.001            | <.001          | <.001            |
| <b>Fifth Grade</b> |                              |         |       |                              |                  |                |                  |                |                  |
| Vocabulary         | 35.67                        | 3, 4723 | <.001 | 0.03                         | 0.87             | <.001          | 0.02             | <.001          | <.001            |
| Comprehension      | 31.93                        | 3, 4761 | <.001 | 0.68                         | 1.00             | <.001          | 0.47             | <.001          | <.001            |
| Composite          | 39.84                        | 3, 4719 | <.001 | 0.13                         | 0.98             | <.001          | 0.08             | <.001          | <.001            |

**Repeated Measures Analysis of Variance**

Table A-16 provides additional controls by using a repeated measures analysis of variance with the first and fifth grade scores as a repeated measure and the type of school (NIFDI, DI and Control for panel A and NIFDI, HPODI, LPODI, and control for panel B) as a factor. As noted in the analysis of the first research question, this approach is more parsimonious than the separate one-way analyses of variance or groups of paired t-tests. If exposure to Direct Instruction in first grade is related to higher achievement in fifth grade, a significant interaction effect would be expected. As expected, the interaction effects are significant in all cases. The changes in scores from first to fifth grade were not the same in the various groups. As demonstrated in Table A-14, these changes were substantially stronger in the NIFDI-supported schools than the other schools.

Table A-16: Repeated Measures Analysis of Variance, Normed Scores, Reading

*A: Comparison of NIFDI, Other DI, and Control Schools*

|             | Vocabulary | Comprehension | Composite |
|-------------|------------|---------------|-----------|
| F - reading | 60.45      | 178.7         | 169.64    |
| Sig         | <.001      | <.001         | <.001     |
| F- Group    | 10.77      | 5.48          | 7.98      |
| Sig         | <.001      | 0.004         | <.001     |
| Interaction | 42.59      | 13.5          | 32.54     |
| Sig.        | <.001      | <.001         | <.001     |

*B: Comparison of NIFDI, HPODI, LPODI, and Control Schools*

|             | Vocabulary | Comprehension | Composite |
|-------------|------------|---------------|-----------|
| F - reading | 45.37      | 130.65        | 129.21    |
| Sig         | <.001      | <.001         | <.001     |
| F- Group    | 35.16      | 27.82         | 34.57     |
| Sig         | <.001      | <.001         | <.001     |
| Interaction | 29.25      | 9.48          | 23.26     |
| Sig.        | <.001      | <.001         | <.001     |

**Mixed Model Analyses**

Tables A-17 and A-18 give the results of the mixed model analysis. Panel A in Table A-17 summarizes the models that were used and Panel B gives summary statistics associated with each model. Four models are tested. The first is the simple, random effects model. The second adds first grade achievement scores and Factor 1 as a predictor.<sup>7</sup> The third model adds dummy variables for being in a NIFDI-supported school throughout one's elementary career and for being in an Other DI school during the latter portions of the career. The fourth model is like Model 3, but separates the Other DI category into the LPODI and HPODI schools. For both Model 3 and Model 4 the omitted category of schools is the Control Schools. Using the procedures described more fully in the discussion of results with the first research question, it can be seen that the model that best fits the data is Model 4, which includes first

---

<sup>7</sup> Factor 2 was omitted from this analysis because it had no significant effect in the analyses of Hypothesis One and to help conserve degrees of freedom. Results do not differ when it is included.

Table A-17: Components of Models and Measures of Model Fit, Mixed Model Regressions of Fifth Grade Reading Achievement on First Grade Achievement, School Context of Advantage/Disadvantage, and DI Instruction, Comparing NIFDI Supported Schools, Other DI Schools (with and without low poverty school) and Control Schools

A: Variables in the Models

| Variables            | Model 1 | Model 2 | Model 3 | Model 4 |
|----------------------|---------|---------|---------|---------|
| Rand. Int. - Schools | x       | x       | x       | x       |
| 1st Gr. Achievement  |         | x       | x       | x       |
| Factor 1             |         | x       | x       | x       |
| NIFDI support        |         |         | x       | x       |
| Other DI             |         |         | x       |         |
| LPODI                |         |         |         | x       |
| HPODI                |         |         |         | x       |

B: Model Fit Statistics

|                      | Model 1 |      |        | Model 2 |      |        | Model 3 |      |        | Model 4 |      |        |
|----------------------|---------|------|--------|---------|------|--------|---------|------|--------|---------|------|--------|
| <b>Vocabulary</b>    | Effect  | s.e. | p.     | b       | s.e. | p.     | b       | s.e. | p.     | b       | s.e. | p.     |
| Random Effects       | 28.6    | 4.9  | <.0001 | 24      | 4.2  | <.0001 | 22.4    | 4    | <.0001 | 20.7    | 3.8  | <.0001 |
| Residual             | 286.6   | 6    | <.0001 | 218.8   | 4.6  | <.0001 | 218.7   | 4.6  | <.0001 | 218.7   | 4.6  | <.0001 |
| BIC                  | 40347   |      |        | 38618   |      |        | 38619   |      |        | 38617.8 |      |        |
| - 2 Log Likelihood   | 40333   |      |        | 38594   |      |        | 38586   |      |        | 38580.5 |      |        |
| Change in LL         |         |      |        | 1739    |      |        | 8.1     |      |        | 14      |      |        |
| df                   |         |      |        | 2       |      |        | 2       |      |        | 3       |      |        |
| p                    |         |      |        | <.0001  |      |        | <.05    |      |        | <.01    |      |        |
| Correlation ratio    | 0.09    |      |        |         |      |        |         |      |        |         |      |        |
| <b>Comprehension</b> |         |      |        |         |      |        |         |      |        |         |      |        |
| Random Effects       | 19.9    | 3.6  | <.0001 | 17      | 3    | <.0001 | 16      | 2.9  | <.0001 | 14.9    | 2.8  | <.0001 |
| Residual             | 266.4   | 5.5  | <.0001 | 193.3   | 4    | <.0001 | 193.3   | 4    | <.0001 | 193.4   | 4    | <.0001 |
| BIC                  | 40299   |      |        | 38338   |      |        | 38343   |      |        | 38343   |      |        |
| - 2 Log Likelihood   | 40285   |      |        | 38315   |      |        | 38311   |      |        | 38306   |      |        |
| Change in LL         |         |      |        | 1970    |      |        | 4       |      |        | 8.7     |      |        |
| df                   |         |      |        | 2       |      |        | 2       |      |        | 3       |      |        |
| p                    |         |      |        | <.0001  |      |        | >.05    |      |        | <.05    |      |        |

|                    |       |     |        |        |     |        |       |      |        |       |     |        |
|--------------------|-------|-----|--------|--------|-----|--------|-------|------|--------|-------|-----|--------|
| Correlation ratio  | 0.07  |     |        |        |     |        |       |      |        |       |     |        |
| <b>Composite</b>   |       |     |        |        |     |        |       |      |        |       |     |        |
| Random Effects     | 25.6  | 4.4 | <.0001 | 22.2   | 3.8 | <.0001 | 20.6  | 3.6  | <.0001 | 19.3  | 3.4 | <.0001 |
| Residual           | 269.1 | 5.6 | <.0001 | 177.9  | 3.7 | <.0001 | 177.9 | 47.6 | <.0001 | 178   | 3.7 | <.0001 |
| BIC                | 40012 |     |        | 37358  |     |        | 37361 |      |        | 37361 |     |        |
| - 2 Log Likelihood | 39998 |     |        | 37335  |     |        | 37329 |      |        | 37324 |     |        |
| Change in LL       |       |     |        | 2663   |     |        | 6.4   |      |        | 11.2  |     |        |
| df                 |       |     |        | 2      |     |        | 2     |      |        | 3     |     |        |
| p                  |       |     |        | <.0001 |     |        | <.05  |      |        | <.02  |     |        |
| Correlation ratio  | 0.09  |     |        |        |     |        |       |      |        |       |     |        |

Note: The comparison of the log-likelihood values for Model 4 is with Model 2.

Table A-18: Mixed Model Regressions, Fifth Grade Reading (norm equivalent scores) on School Advantage/Disadvantage and School Type (NIFDI Supported, Other DI, and Control Schools)

|                      | Model 2 |       |        | Model 3 |       |        | Model 4 |       |        |
|----------------------|---------|-------|--------|---------|-------|--------|---------|-------|--------|
| <b>Vocabulary</b>    | b       | s.e.  | p.     | b       | s.e.  | p.     | b       | s.e.  | p.     |
| Intercept            | 29.6    | 0.7   | <.0001 | 29.1    | 0.8   | <.0001 | 29.2    | 0.7   | <.0001 |
| First grade score    | 0.4     | 0     | <.0001 | 0.4     | 0     | <.0001 | 0.4     | 0     | <.0001 |
| Factor 1             | 1.3     | 0.5   | 0.008  | 1.3     | 0.5   | 0.01   | 1.1     | 0.5   | 0.02   |
| NIFDI                | -----   | ----- | -----  | 4.3     | 1.7   | 0.01   | 4.2     | 1.7   | 0.01   |
| Other DI             | -----   | ----- | -----  | 3.8     | 2.4   | 0.12   | -----   | ----- | -----  |
| HPODI                | -----   | ----- | -----  | -----   | ----- | -----  | 0.8     | 2.6   | 0.77   |
| LPODI                | -----   | ----- | -----  | -----   | ----- | -----  | 14.3    | 5     | 0.004  |
| <b>Comprehension</b> |         |       |        |         |       |        |         |       |        |
| Intercept            | 29.2    | 0.7   | <.0001 | 29      | 0.7   | <.0001 | 29      | 0.7   | <.0001 |
| First grade score    | 0.4     | 0     | <.0001 | 0.4     | 0     | <.0001 | 0.4     | 0     | <.0001 |
| Factor 1             | 1       | 0.4   | 0.02   | 1.1     | 0.4   | 0.01   | 0.9     | 0.4   | 0.04   |

|                   |       |       |        |       |       |        |       |       |        |
|-------------------|-------|-------|--------|-------|-------|--------|-------|-------|--------|
| NIFDI             | ----- | ----- | -----  | 0.7   | 1.5   | 0.63   | 0.8   | 1.5   | 0.6    |
| Other DI          | ----- | ----- | -----  | 4.17  | 2.1   | 0.05   | ----- | ----- | -----  |
| HPODI             | ----- | ----- | -----  | ----- | ----- | -----  | 1.8   | 2.3   | 0.44   |
| LPODI             | ----- | ----- | -----  | ----- | ----- | -----  | 12.4  | 4.3   | 0.004  |
| <b>Composite</b>  |       |       |        |       |       |        |       |       |        |
| Intercept         | 25.9  | 0.7   | <.0001 | 25.4  | 0.7   | <.0001 | 25.5  | 0.7   | <.0001 |
| First grade score | 0.5   | 0     | <.0001 | 0.5   | 0.01  | <.0001 | 0.5   | 0     | <.0001 |
| Factor 1          | 1.2   | 0.5   | 0.013  | 1.2   | 0.5   | 0.009  | 1     | 0.5   | 0.03   |
| NIFDI             | ----- | ----- | -----  | 2.8   | 1.6   | 0.07   | 2.9   | 1.5   | 0.06   |
| Other DI          | ----- | ----- | -----  | 4.3   | 2.3   | 0.05   | ----- | ----- | -----  |
| HPODI             | ----- | ----- | -----  | ----- | ----- | -----  | 1.7   | 2.5   | 0.5    |
| LPODI             | ----- | ----- | -----  | ----- | ----- | -----  | 13.7  | 4.8   | 0      |

year reading achievement, factor 1 (the measure of school advantage/disadvantage), and the three dummy variables of treatment group (NIFDI supported, HPODI, and LPODI).

The coefficients in Table A-18 confirm the results found in conjunction with Hypothesis 1 regarding the impact of first grade reading scores and Factor 1. Students with higher reading achievement scores in first grade have significantly higher reading scores in fifth grade. Students in schools with more non-Hispanic white children and fewer children receiving free and reduced lunch also have higher achievement scores. These results appear with all three measures of achievement.

Results with model three, comparing the three groups of schools, indicate that students in schools supported by NIFDI throughout their elementary careers had higher achievement scores than other students, even after controlling for first grade scores and school context, but this difference is significant only for vocabulary scores ( $p = .01$ ), although near significance ( $p = .07$ ) for the composite score. Students in the Other DI schools also had higher scores than other students on all three measures, and these differences were significant ( $p=.05$ ) for both comprehension and the composite score.

Results with Model 4, the best fitting model, which separates the low poverty and high poverty Other DI schools, show that the positive effect for the Other DI schools in Model 3 was due to the presence of the LPODI school. In Model 4, students in the LPODI school had significantly higher achievement scores than other students, even after controlling for their first grade scores and school context, on all three measures. Compared to the other coefficients associated with group membership, the magnitude of the coefficient associated with the LPODI school is quite large. Students in this school, once first grade scores and the socioeconomic characteristics of the school are controlled, are predicted to have fifth grade scores that range from 12 to 14 NCE points higher than students in the Control Schools. The advantage for students in the NIFDI-supported schools ranges from one to four points, while that for students in the Other DI schools ranges from one to two points. Even with such a low magnitude, however, the advantage for the NIFDI students in vocabulary scores remains statistically significant (.01), and the advantage in composite scores is close to traditional significance levels ( $p = .06$ ).

### **Summary**

As with research question number 1, the results indicate that students who had Direct Instruction in first grade had higher achievement scores in 5<sup>th</sup> grade than students in the Control Schools. However, the impact of receiving Direct Instruction was strongest for students in the low poverty school in the Other DI group. It should be recalled, however, that students in this school had significantly higher first grade achievement scores than students in other schools. Perhaps even more important, the school also had a much lower rate of student turnover throughout the years of the study.

**Research Question Three. 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?**

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-19 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 but one of the effect sizes associated with a group that received Direct Instruction in first grade meet this threshold. The one exception involves vocabulary scores for schools in the HPODI group, where the effect size is .20. The effect size for comprehension for the Control schools is almost at the .25 threshold. However, this effect is half or less than half that found for the other schools. The effect sizes associated with the NIFDI-supported schools all far surpass the .25 criterion and are second in size only to those for the LPODI school, which had very low levels of turnover.

Table A-19: Effect Sizes of Change, First to Fifth Grade, by Group, Reading Achievement Norm Equivalent Scores

|                  | Vocabulary | Comprehension | Composite |
|------------------|------------|---------------|-----------|
| Control Schools  | -0.04      | 0.24          | 0.14      |
| All DI Schools   | 0.37       | 0.47          | 0.47      |
| NIFDI Supported  | 0.49       | 0.49          | 0.55      |
| Other DI Schools | 0.25       | 0.46          | 0.4       |
| HPODI Schools    | 0.20       | 0.46          | 0.36      |
| LPODI Schools    | 0.49       | 0.62          | 0.67      |

Note: All effect sizes are based on norm equivalent scores. The computations for the effect sizes for "all DI schools" and the "Control Schools" are based on the data in Table A-8 associated with Research Question 1. The computations for the other groups are based on the descriptive statistics given in Table A-14.



## References

Dunlap, William P., Jose M. Cortina, Joel B. Vaslow, and Michael J. Burke. 1996. "Meta-analysis of experiments with matched groups or repeated measures designs." *Psychological Methods* 1: 170-177.

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.

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