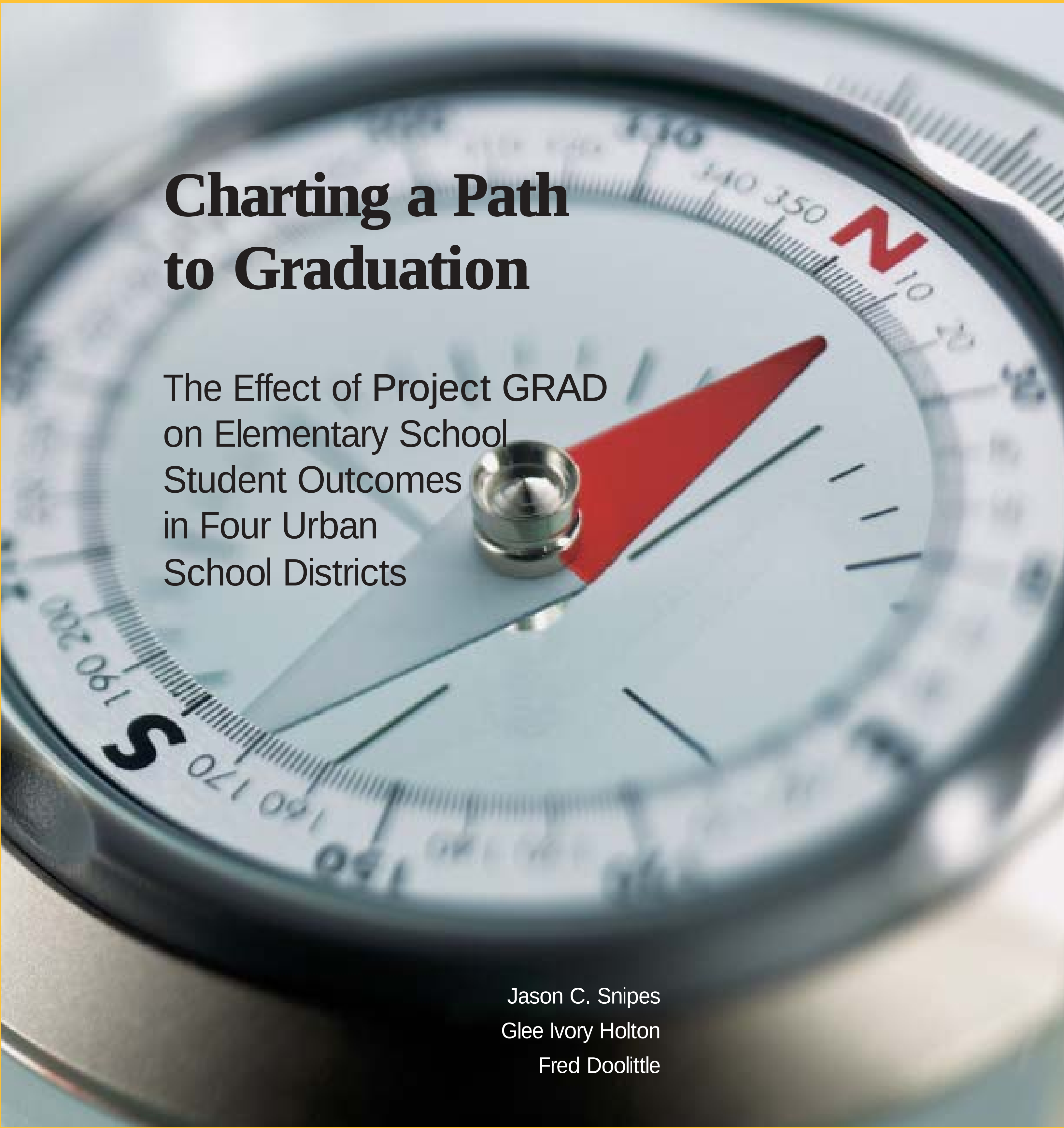


Charting a Path to Graduation

The Effect of Project GRAD on Elementary School Student Outcomes in Four Urban School Districts

Jason C. Snipes
Glee Ivory Holton
Fred Doolittle

July 2006

mdrc

BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

Charting a Path to Graduation

The Effect of Project GRAD on Elementary School Student Outcomes in Four Urban School Districts

Jason C. Snipes
Glee Ivory Holton
Fred Doolittle

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

July 2006

Primary funding for this report was provided by the Ford Foundation, and supplemental funding was provided by the Bill & Melinda Gates Foundation, Lucent Technologies Foundation, Project GRAD USA, and The Grable Foundation.

Dissemination of MDRC publications is supported by the following funders that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: Alcoa Foundation, The Ambrose Monell Foundation, The Atlantic Philanthropies, Bristol-Myers Squibb Foundation, Open Society Institute, and The Starr Foundation. In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, The Sandler Family Supporting Foundation, and The Stupski Family Fund, as well as other individual contributors.

For information about MDRC and copies of our publications, see our Web site: www.mdrc.org.

Overview

Project Graduation Really Achieves Dreams (GRAD) is an ambitious education reform initiative designed to improve academic achievement, high school graduation rates, and rates of college attendance for low-income students. It is an unusual reform model in that it intervenes throughout an entire "feeder pattern" of elementary and middle schools that send students into each Project GRAD high school. This report presents results of MDRC's multiyear evaluation of the effects of Project GRAD on student achievement at elementary schools in six feeder patterns, encompassing a total of 52 schools in four cities: Houston, Texas (the original site); Atlanta, Georgia; Columbus, Ohio; and Newark, New Jersey. A companion report examines Project GRAD's effects at the high school level in three urban school districts.

In elementary schools, Project GRAD implements reading and math curricula, with enhanced professional development for teachers. In addition, each elementary school builds support in the community for school improvement and college attendance, implements a classroom management program, provides students with access to needed social services, and receives special support from local Project GRAD organizations. At the high school level, Project GRAD's model assumes that better-prepared students would come from the feeder schools, would benefit from special academic counseling and summer academic enrichment in high school, and would qualify for a scholarship to attend college, which is the "cornerstone" of the Project GRAD reform.

The key findings of this report are:

- Scores on *state achievement tests* at Project GRAD elementary schools in Houston and Atlanta improved in the years following implementation of the initiative. However, in an environment of strong state and local focus on state achievement tests, scores improved by similar amounts at comparison schools in these same districts.

- Project GRAD produced statistically significant positive effects on elementary students' scores on *national achievement tests* in Houston and Newark; that is, while comparison schools experienced a decline in scores on these tests, Project GRAD schools saw scores remain constant or increase.

- In Columbus, the implementation of Project GRAD was initially weaker than in the other sites, and this appears to have lowered test scores — both absolutely and relative to comparison schools — in the early years of the initiative.

Contents

List of Boxes, Tables, and Figures

Box

Table

Table

Figure

Figure

Preface

In the past decade, school districts around the country have sought to improve struggling urban high schools, where high dropout rates, poor student achievement, and low rates of graduation and college-going remain all too prevalent. In a field crowded with reform initiatives, Project Graduation Really Achieves Dreams (GRAD) stands out as particularly ambitious, focusing as it does on improving conditions for high school students *before* they even reach high school.

First implemented in one high school in Houston, Texas, in the early 1990s, Project GRAD has evolved into a comprehensive reform model that intervenes throughout an entire "feeder pattern" of elementary and middle schools that send students into each Project GRAD high school. The developers of Project GRAD understand that high schools *inherit* struggling students, making it essential to improve both elementary and secondary schools in order to increase the rates of high school graduation, college-going, and college graduation. Project GRAD combines curricular reforms in the lower grades and the opportunity to qualify for a college scholarship in high school with a classroom management program, access to social services, and efforts to promote parental and community involvement at all grade levels.

With principal support from the Ford Foundation, MDRC has conducted an evaluation of Project GRAD in Houston and several expansion districts. This report focuses on student achievement at elementary schools in six feeder patterns, encompassing a total of 52 schools across four districts: Houston, Texas; Atlanta, Georgia; Columbus, Ohio; and Newark, New Jersey. A companion report offers findings on high school student achievement in Houston, Atlanta, and Columbus.

Taken together, the findings from both reports highlight the challenges that urban school districts face in significantly improving the academic performance of high school students. While Project GRAD schools made some significant gains in elementary test scores and for students at the flagship high school in Houston, in other areas they did not outpace the comparison schools, many of which were also engaged in local and districtwide reforms. Project GRAD USA, the dynamic organization that oversees the initiative, has already begun to refine its comprehensive approach to respond to some of the lessons suggested by this unusually rigorous evaluation — namely, the "leaky" nature of many feeder systems, the challenges of bringing reforms to scale, and the importance of taking action at the classroom level to get academic gains. Both the national organization and the local Project GRAD sites deserve credit for submitting their model to a rigorous comparative study.

This evaluation of Project GRAD is only one of several studies that MDRC has recently conducted of high school and district reform efforts, including Career Academies, First Things First, and Talent Development. For more information about MDRC's education research and to download all MDRC reports, please visit our Web site at www.mdrc.org.

Gordon Berlin
President

Acknowledgments

The national evaluation of Project Graduation Really Achieves Dreams (GRAD) benefited greatly from the extraordinary cooperation of the principals, teachers, and students in the Project GRAD sites. They provided invaluable information during site visits and interviews and through teacher and student surveys and focus groups. We are deeply grateful for the assistance of the superintendents, the district administrators, and other school staff who participated in interviews and shared with us important background material. Our work was greatly facilitated by each district's research department that prepared student record data and helped our team build a substantive database as the foundation for the impact study.

We are very appreciative of the assistance provided by the executive directors, board members, and staff of the local Project GRAD organizations and by the staff and leadership of Project GRAD USA. They facilitated site visits, participated in extensive interviews, and included MDRC in training sessions and special meetings, briefings, and conferences. They shared a wealth of background information and helped us keep abreast of new developments. We are especially grateful to Robert Rivera, President and Chief Executive Officer, for his assistance throughout the process, and to James L. Ketelsen for his support and helpful suggestions.

This report also benefited from the assistance of the Project GRAD component developers and their staff, both from the national and the local organizations, and — most important — the assistance of the coordinators, facilitators, and managers at the school-building level. We offer a special thank you to Kwame Opuni for sharing information on Houston and for his insights on the Project GRAD model and scaling-up effort in Houston.

Howard Bloom, MDRC's Chief Social Scientist, provided a technical framework for the evaluation as well as ongoing guidance for the impact analysis. From the inception of the Project GRAD evaluation, Robert Ivry, MDRC's Senior Vice President for Development and External Affairs, has played a major role in shaping the study, and this report benefited from his insights and guidance. We also thank Judith Gueron, former MDRC president, who provided feedback in earlier phases of the project.

Several MDRC staff members played major roles in acquiring and analyzing data for the evaluation, and we thank each of them — particularly, Alison Rebeck Black, Laura Sztejnberg, Amy Karwan, and Rasika Kulkarni — for their valuable contributions. We also thank our team of on-site researchers who, during the early phase of the evaluation, helped gather information on the implementation activities in sites and helped inform cross-site issues: W. Monty Whitney, Jill Lynch, Karen Edwards, Thomas Smith, and Erica Walker.

We are particularly grateful for the guidance provided by our external review committee: Richard J. Murnane, Charles M. Payne, Ronald Ferguson, Melissa Roderick, and Dan O'Brien.

We are most appreciative of the support, advice, and useful suggestions from program officers at the Ford Foundation, Lucent Technologies Foundation, Bill & Melinda Gates Foundation, and The Grable Foundation; we gratefully acknowledge the contributions of Alison Bernstein, Janice Petrovich, Cyrus Driver, the late Richard Curcio, Steven Zwerling, David Ford, and Susan Brownlee.

At MDRC, Vivian Mateo provided research assistance, prepared the tables and figures, and coordinated the production of the report. Mona Grant prepared sections of the report and served as the archivist for implementation data and other source documents. Gordon Berlin, James Kemple, Corinne Herlihy, Amy Rosenberg, Rob Ivry, Charles Michalopoulos, and John Hutchins reviewed drafts of the report. Robert Weber edited the report, and Stephanie Cowell prepared it for publication.

The Authors

Executive Summary

Project Graduation Really Achieves Dreams (GRAD) is an ambitious education reform initiative designed to improve academic achievement, high school graduation rates, and rates of college attendance for low-income students. It is an unusual reform model in that it intervenes throughout an entire "feeder pattern" of elementary and middle schools that send students into each Project GRAD high school. The initiative recognizes that high schools inherit problems that have arisen earlier in the education pipeline, making it essential to improve both elementary and secondary schools in order to increase the rates of high school graduation, college-going, and college graduation.

Project GRAD schools at all levels build support in the community for school improvement and college attendance, implement a classroom management program, provide students with access to needed social services, and receive special support from local Project GRAD organizations. To help students arrive in middle and high school better prepared academically, Project GRAD elementary schools implement specific reading and math curricula, with enhanced professional development for teachers. At the high school level, Project GRAD's model assumes that better-prepared students would come from the Project GRAD feeder schools, would benefit from special academic counseling and summer academic enrichment in high school, and would qualify for a scholarship to attend college, which is the "cornerstone" of the Project GRAD reform.

Given Project GRAD's emphasis on early intervention, understanding the program's effects on elementary student achievement is a key step in evaluating its overall effectiveness. This report describes the effects of Project GRAD on student achievement at elementary schools in six feeder patterns, encompassing a total of 52 schools across four districts: Houston, Texas (the original site); Atlanta, Georgia; Columbus, Ohio; and Newark, New Jersey. MDRC — a nonprofit, nonpartisan research organization — conducted a third-party evaluation to determine the effects of Project GRAD by comparing the changes in student outcomes at Project GRAD schools with changes at similar, non-Project GRAD schools in the same districts. (A companion report examines Project GRAD's effects at the high school level.)[1] In general, Project GRAD student outcomes are tracked from the implementation of the first components of the model at each site until the 2002-2003 school year. The key findings of this report are:

[1]Jason C. Snipes, Glee Ivory Holton, Fred Doolittle, and Laura Sztejnberg, *Striving for Student Success: The Effect of Project GRAD on High School Student Outcomes in Three Urban School Districts* (New York: MDRC, 2006).

- **Scores on *state achievement tests* at Project GRAD elementary schools in Houston and Atlanta improved in the years following implementation of the initiative. However, in an environment of strong state and local focus on state achievement tests, scores improved by similar amounts at comparison schools in these same districts.**

- **Project GRAD produced statistically significant positive effects on elementary students' scores on *national achievement tests* in Houston and Newark — that is, while comparison schools experienced a decline in scores on these tests, Project GRAD schools saw scores remain constant or increase.**

- **In Columbus, the implementation of Project GRAD was initially weaker than in the other sites, and this appears to have lowered test scores — both absolutely and relative to comparison schools — in the early years of the initiative.**

The remainder of the Executive Summary describes the Project GRAD model and how it was implemented in the school districts, explains how the evaluation was conducted, and summarizes the study's findings and explores their implications.

What Is Project GRAD and How Was It Implemented?

Project GRAD is unusual in recognizing the interconnection of educational issues at the elementary and secondary levels by working at the level of a feeder pattern — a high school and the associated elementary and middle schools that feed into it. Over time, Project GRAD has evolved from an effort to increase the rate of college-going among students at one Houston high school — by offering college scholarships — into a more comprehensive response to the educational problems that students at all levels face in scores of schools.

A complex, multilayered initiative, Project GRAD includes a set of core components for all the schools in a feeder pattern as well as components for the schools at each level, as described below.

Components at Project GRAD Elementary Schools

During the time covered by this study, Project GRAD had two curricular interventions at the elementary school level, as well as the components described below that seek to create an environment that is conducive to learning. (Currently, Project GRAD supports whatever reading and math curricula that participating districts adopt.)

- **Reading curriculum:** Most Project GRAD sites used Success for All (SFA), a nationally recognized reading program that focuses on the key elements of reading instruction during concentrated instructional time (90 minutes each day), with the goal of bringing students to grade-level reading by third grade.

- **Math curriculum:** Math Opportunities, Valuable Experiences, Innovative Teaching (MOVE IT™ Math) was Project GRAD's recommended math program. It offers elementary school teachers professional development and instructional materials organized around the National Council of Teachers of Mathematics (NCTM) Standards program, involves heavy use of manipulatives to address a wide variety of learning styles, emphasizes daily problem solving, and introduces algebra in the early grades.

- **Parental and community involvement:** Project GRAD seeks to engage parents and the community in the work of the schools, build awareness of the opportunity to attend college, and support the learning of students.

- **Social services and academic enrichment:** One of two programs — Communities In Schools (CIS) or the Campus Family Support (CFS) Plan (developed by Project GRAD) — bring additional social services, academic activities, and volunteers into Project GRAD schools to address issues that students and their families face and to build commitment to academic success.

- **Classroom management:** Programs developed by Consistency Management & Cooperative Discipline℠ (CMCD)℠ are designed to produce orderly classrooms focused on learning by promoting student responsibility and self-discipline and positive relationships among students, teachers, and other adults in the school.

Components at Project GRAD High Schools

At the high school level, Project GRAD includes the three components focused on parent and community involvement, social services and academic enrichment, and classroom management. In addition, Project GRAD high schools offer two components:

- **Project GRAD college scholarships** are provided to students who have a cumulative 2.5 grade point average, graduate within a four-year time period, complete a recommended college preparatory curriculum, and participate in two summer institutes. Scholarship amounts and criteria vary slightly by site but usually average $1,000 to $1,500 each year during the four years of college. Each Project GRAD high school has a scholarship coordinator who provides counseling, tutoring, and college admission preparation.

- **Summer institutes** provide an opportunity for qualifying Project GRAD students to experience a college campus-based program taught by college faculty and to enhance their academic skills.

Based on encouraging results in its first Houston feeder pattern, Project GRAD expanded to other feeder patterns within the district. In 1998, Newark, New Jersey, became the first site outside Houston to implement Project GRAD; Columbus, Ohio, and Atlanta, Georgia, followed soon thereafter. Currently, Project GRAD operates in five feeder patterns in Houston and in 12 school districts and 211 schools in eight states across the country, serving more than 131,000 students. To manage and support each Project GRAD initiative, local not-for-profit organizations were established in Houston and the expansion sites. Expansion within the Houston schools and to other school districts stretched the capacity of some program developers to support the model's components and prompted the development of a national organization in 2000 — Project GRAD USA — to sustain implementation efforts and to address implementation issues across sites.

Three important points should be noted about the implementation of Project GRAD in elementary schools in the four study sites:

- **Although the implementation process differed across sites, the feeder patterns of schools examined in this report generally implemented the core Project GRAD components and followed the approach set forth in the model.**

- **Local situations in the school districts meant that the strength of Project GRAD's implementation varied. Houston and Atlanta achieved the strongest implementation, followed by Newark and, finally, Columbus.**

- **Many of the comparison schools were also participating in reform initiatives, likely lessening the treatment contrast between Project GRAD and comparison schools, particularly in Houston and Atlanta.**

How Was the Evaluation Conducted?

The goal of this evaluation is to understand whether Project GRAD changed the academic achievement of children in the elementary schools it serves and, if so, how. The study focuses on test scores because they are the focus of policy attention and because other typical measures — like absence or expulsion rates — are already very low in Project GRAD elementary schools. The evaluation relied on the tests administered by the school districts, which included state achievement tests, national achievement tests, or — in Houston — both types of tests.

To estimate the program's effect on achievement, MDRC used an approach called "comparative interrupted time series analysis." The first step in estimating program impacts with this design is to compare the change at Project GRAD schools in a given student outcome after the school began implementing Project GRAD with the average outcome during a baseline period, before implementation. This estimate represents how student performance changed in the presence of Project GRAD but does not, by itself, provide a measure of the *effect* of Project GRAD. The next step is to measure the corresponding change during the same period for similar schools not implementing Project GRAD. This measurement provides an estimate of how student performance would most likely have changed at the Project GRAD schools in the absence of the reform. The *difference* between these two changes is an estimate of the *impact* of the reform — the effects that can be attributed to Project GRAD.

Project GRAD is typically implemented over several years as individual components of the model are put in place, so these findings reflect the initiative at a specific point in its history at each site. Being the first district to implement the model, Houston offers more years of follow-up data than the expansion sites, which were at an early stage in their operation of Project GRAD during the years covered by this report. Finally, student mobility into and out of schools is common in urban districts. While the findings presented here include all students at the Project GRAD and comparison schools, the findings are similar when the analysis focuses on "nonmobile" students who remained at these schools for multiple years.

How Did Project GRAD Affect Elementary Student Achievement?

- **In Houston and Atlanta, where Project GRAD implementation was strong, student scores on *state achievement tests* at the Project GRAD schools improved. During the same period, similar improvements on state tests also occurred at the comparison schools, which implemented other district- and school-level reforms (often focused on boosting scores on state tests).**

Achievement on the Texas state standards-based tests at Project GRAD Houston elementary schools improved substantially during the years following the initiative's implementation. However, comparison schools throughout the district made similar progress on these tests, suggesting that Project GRAD did not improve these outcomes beyond what would have happened without the program. The period from 1993 to 2003 was one of substantial progress in students' test scores across low-performing elementary schools in Houston. For example, over the eight available years of follow-up, average test scores for fourth-grade math (as measured by the Texas

Learning Index) at the Project GRAD schools in the Jefferson Davis feeder pattern rose from 63 to 82, while scores at the corresponding comparison schools rose from 61 to 81.

- **Scores on *national achievement tests* fell at comparison schools in Houston and Newark during the study period. Project GRAD frequently prevented or lessened a similar deterioration in performance on these tests, resulting in significant positive effects on elementary student achievement relative to national norms.**

In Houston, findings for the third grade demonstrate the pattern of effects. The Stanford Achievement Test (SAT-9) — a test comparing students to test-takers nationally — was first administered in Houston in 1998, several years after the initial implementation of Project GRAD in the Davis High School feeder pattern. Students' performance on the SAT-9 at the comparison schools used for Houston's three feeder patterns generally declined, whereas scores at the Project GRAD schools in two of the three feeder patterns generally remained relatively stable or fell by less than at the comparison schools. The net result is a consistent set of statistically significant positive effects on elementary-level SAT-9 achievement in both reading and math. For example, the analysis suggests that, in the absence of Project GRAD, third-grade SAT-9 math achievement throughout the Davis feeder pattern would have fallen to the 25th percentile; with Project GRAD, math achievement reached the 38th percentile.

In Newark during the six years prior to Project GRAD's implementation, test scores on the SAT-9 steadily declined, reflecting the district's turmoil. During the first two years of follow-up, scores at the Project GRAD schools stopped declining and improved substantially relative to the earlier trend line, whereas no similar break with prior negative trends occurred at the comparison schools. These effects were especially pronounced for several grades and subjects. For example, the analysis suggests that, in the first year of implementation, average third-grade math achievement at the Project GRAD schools reached the level of the 48th percentile instead of the 28th percentile — the level that was predicted, had the model not been implemented. Unfortunately, changes in testing in the Newark district prevented longer-term follow-up, so it is not possible to determine whether these positive, statistically significant, and substantial effects continued. It is important to note also that the positive effects in Newark began before the model's instructional components were even implemented, suggesting that the components relating to classroom discipline and social supports — by themselves — can have effects on academic performance.

- **In Columbus, trends in test scores reflect the site's inconsistent implementation of Project GRAD. Overall, there was little sustained improvement in test scores at either the Project GRAD or the comparison schools during this early follow-up period. There is some indication that**

Project GRAD may have produced small, negative impacts on some subjects and grades, most of which dissipated over time.

As has been found in other studies, difficult early implementation of complicated education reforms can temporarily result in stresses on schools and in unintended short-term effects on student outcomes. In some follow-up years and grades, scores at the Project GRAD Columbus schools appear to have fallen slightly below the baseline averages, while no similar declines from the baseline occurred at the comparison schools. By the third year of follow-up, these negative impacts had largely disappeared, except for declines in fifth-grade math.

What Are the Implications of These Findings?

In general, Project GRAD was able to operate in a variety of contexts that differed in terms of prior student achievement, local capacity, existing education reforms, and district staffing rules. The ambitiousness of the initiative's model and its expansion to sites other than Houston required the creation of a national organization (Project GRAD USA) that developed its own technical assistance capacity. The efforts to expand into additional feeder patterns in Houston and simultaneously into new cities sometimes stretched the capacity of the developers to support the model's reforms.

At the same time, the local context in each district where Project GRAD was attempted had important influences both on the success of the model's implementation and on its effects on student achievement. This leads to a few observations:

- **In settings that were already mounting reforms focused on improving state test scores — as in Atlanta and Houston — Project GRAD does not appear to have generated systematic improvements on state assessments that were greater than the improvements at the comparison schools.**

- **On the other hand, in both Houston and Newark, Project GRAD did reverse declining trends on national achievement tests. This suggests that Project GRAD has the potential to help schools improve — or, at least, to avoid deterioration in — the more general academic competencies measured by some national achievement tests.**

Although data limitations prevent a full examination of this theme, Project GRAD schools with reasonably good implementation appear to have achieved comparable improvements on state tests as similar local schools, while avoiding declines in scores on national tests. Some experts argue that a narrow focus on improving student performance on state standards-based tests can have an unintended deleterious effect on student achievement measured more broadly. Project GRAD's positive impact on national test scores may help address that concern.

- **Except in Houston, the Project GRAD programs were still relatively early in their life cycle when the data were collected. Many argue that it takes at least five years for education reforms to take hold and show results, which highlights the possibility that results in Atlanta and Columbus might still improve.**

This evaluation represents the experience of only four district sites and six feeder patterns. The expansion sites of Atlanta, Columbus, and Newark were the first of the new districts added to the Project GRAD network, which has since expanded to at least eight additional districts. Project GRAD's implementation process has undergone important revisions — many growing out of this early experience — that are not captured in the evaluation.

- **Project GRAD may be most useful in school districts where existing reform efforts may not yet be providing adequate support to improve elementary-level instruction — districts where the model's programmatic and structural elements may meet important needs.**

In some districts — even in low-performing districts that serve large proportions of economically disadvantaged students — ongoing reforms may be producing rising achievement scores, even though achievement levels may still be lower than desired. This is particularly likely to be the case for performance on state-mandated, standards-based assessments. In such settings, Project GRAD may not fill a gap in existing efforts to improve elementary-level instruction in ways that help meet the standards, and the initiative may compete with other reforms for attention and support. Even in these contexts, however, Project GRAD may improve (or at least prevent the erosion of) student performance on the more general skills that are not necessarily measured by state standards tests.

The key implication is to focus on districts that have low achievement and high levels of disadvantaged and minority students, where Project GRAD's emphasis on elementary-level instruction in reading and math and on classroom management and social service supports would represent a value-added difference over and above reforms that are already in place.

Chapter 1

Introduction

The Policy Context

In the current U.S. economy, a high school diploma and postsecondary education and training are all-important steps to success. In past decades, young people without the skills and credentials provided by a solid high school and postsecondary education could find opportunities in the labor market, but this option is becoming increasingly rare. Unfortunately, high school graduation rates have remained mostly unchanged over the past three decades, leaving the United States ranked as seventeenth in the world.[1] In addition, many high school graduates lack key skills needed to succeed in later life. The achievement gap between minority and non-minority students is of great concern. By the end of high school, African-American and Hispanic students, on average, have skills in both reading and mathematics that are the same as the skills of average white students in eighth grade.[2] Of adults age 25 to 29, less than 10 percent of Hispanics complete four or more years of college study, and only 17 percent of black young adults have completed a bachelor's degree.[3] Low rates of high school graduation coupled with low rates of college enrollment have serious ramifications for individuals and their families and for the country's workforce development.

High school reform efforts face a central paradox: The problems that plague most urban high schools are rooted in achievement difficulties that began much earlier than the students' arrival in the ninth grade. Educators, researchers, and policymakers point to weak preparation in basic subjects in the early grades, coupled with the lack of effective student discipline strategies and the absence of sufficient social service supports (especially at schools with high proportions of low-income students) as critical factors causing low rates of high school achievement and graduation. For at-risk youth in poor communities, low motivation, a weak sense of efficacy, and a belief that college is out of reach can compound these issues.

Project Graduation Really Achieves Dreams (GRAD) is a relatively new initiative that originated in Houston, Texas, in 1993. By 2004, it had expanded into 11 additional school districts and generated interest at the district, state, and federal levels.[4] This interest and growth

[1]National Center for Education Statistics (1999).

[2]Haycock (2001).

[3]Center on Educational Policy (2004).

[4]Project GRAD's expansion has been supported by a federal line item that provided $20 million in 2003. The Ford Foundation has provided over $50 million to the initiative. Each Project GRAD community raises local dollars, and school districts contribute by paying a percentage of the costs of specific components. The combined annual budget for the initiative across 12 sites in 2004 was about $75 million, which includes the

(continued)

reflects the fact that Project GRAD's core features are designed to respond directly to problems that plague urban school districts and that it offers a multidimensional approach to school reform. Moreover, Project GRAD has positioned itself as an initiative specifically developed for the most troubled schools in a district. It is an ambitious effort that works above the individual school level to focus simultaneously on all the students in kindergarten[5] through grade 12 in a set of elementary and middle schools and an associated high school (that is, all the students in a feeder pattern). Specifically, Project GRAD recognizes that high schools inherit problems that have arisen earlier in the education pipeline and that it is essential to effect change at all school levels (elementary, middle, and high school) in order to build a strong base of community support to advocate for change and successfully reform high schools.

What Is Project GRAD?

Project GRAD is an education reform initiative designed to improve student outcomes and close the academic achievement gap between low-income and minority students and their more advantaged counterparts. The mission of Project GRAD is to ensure a quality public school education for all children in economically disadvantaged communities, so that high school graduation rates increase and students are prepared to enter and graduate from college. Specifically, Project GRAD has two goals: (1) to ensure that 80 percent of all entering ninth-graders in its high schools graduate and (2) to ensure that 50 percent of Project GRAD high school graduates go on to college.[6] Toward these ends, Project GRAD focuses on grades kindergarten through 12 (K-12), including the elementary and middle schools that feed into a high school, and — in addition to a college scholarship offer — it combines a number of mutually reinforcing reforms that are designed to increase reading and math achievement, improve classroom behavior, encourage parental involvement and community support for school reform, provide social service supports for students and their families, and reduce dropout rates and increase graduation rates among at-risk high school students. Given the scholarship and the follow-up that is provided to students who receive it, Project GRAD essentially becomes a K-16 initiative.

Project GRAD originated in Houston, Texas, in 1993, emerging from an extended effort by its founder — James Ketelsen, the CEO of Tenneco — and other business leaders, who were attempting to increase the rate of college-going in a local high school serving low-income students, by offering scholarships to students who qualified. When this offer did not generate a significant increase in high school graduation and college enrollment, Ketelsen expanded the effort to include the associated middle and elementary schools that fed into the high school and

administrative costs of the national coordinating organization, Project GRAD USA. The present evaluation does not include a cost analysis.

[5]In some sites, Project GRAD also operates in pre-K classes.

[6]See Project GRAD USA (2004a, 2004b).

combined additional reforms intended to increase the chances that students could take advantage of the opportunity provided by a college scholarship. Based on encouraging results in the first feeder pattern, Project GRAD expanded to other feeder patterns within the Houston district. In 1998, Newark, New Jersey, became the first site outside Houston to implement Project GRAD. Currently, Project GRAD operates in five feeder patterns in Houston and in 12 school districts and 211 schools in eight states across the country, serving more than 131,000 students.[7]

Project GRAD's Approach to Improving Student Outcomes

As with many reform strategies, Project GRAD's principles were crafted in fairly broad terms at the outset and then were refined on the basis of early operational experiences and lessons. What has remained constant is a central premise that all students can be effective learners, regardless of their backgrounds, if appropriate and timely programmatic interventions are infused in the primary grades and if the appropriate supports are provided at the secondary level. Project GRAD is not based on one specific educational philosophy or pedagogical approach; rather, it is a collection of program components and complementary operational strategies backed by research and/or experiential support. As such, the initial choice of components was more pragmatic than theory-driven, compared with other reform approaches, and Project GRAD has continued to develop and refine both the program's individual components and its general approach.

Figure 1.1 illustrates the Project GRAD approach. Though the basic structure was initially presented as a set of five components, additional elements have been added over time, and Project GRAD's founders emphasize the importance of operating the initiative as an integrated set of programs and supports — structural components and program components that enhance each other and that collectively create a consistent educational experience for Project GRAD students and schools.

The Scope of the Project GRAD Effort

A defining characteristic of the Project GRAD model is its recognition that many of the challenges that undermine students' success in high school and beyond begin at a much earlier point in their educational careers. Therefore, Project GRAD breaks from the school-by-school approach used by many reforms and instead is implemented throughout a "feeder pattern" — that is, throughout a cluster of elementary and middle schools that feed into a particular high school where the Project GRAD program and scholarship guarantee are put into place. In so

[7]At the time of this report, Project GRAD sites include Akron, OH; Atlanta, GA; Brownsville, TX; Cincinnati, OH; Columbus, OH; Houston, TX; Kenai Peninsula, AK; Knoxville, TN; Long Island, NY; Lorain, OH; Los Angeles, CA; and Newark, NJ.

The Project GRAD Evaluation

Figure 1.1

Project GRAD's Approach to Improving School Functioning and Student Outcomes

Scope of Effort

All Students in Grades K-12 with Linkage to Postsecondary

Selected High School and "Feeder" Elementary and Middle Schools

Operational Strategies and Specific Components

Creating a School and Classroom Environment Conducive to Learning (Consistency Management & Cooperative Discipline℠ and Communities In Schools or Campus Family Support at all levels and the scholarship offer for qualifying high school graduates)

Strengthening Opportunities to Learn: Literacy and Mathematics (Success for All and MOVE IT™ Math in the early grades and summer institutes for high school students)[a]

Enhancing School-Level Capacity for Program Management and Implementation (Project GRAD local not-for-profit and Project GRAD USA)

Intended Effects on School Functioning and Student Outcomes

Improved School Climate and Classroom Instruction → **Higher Student Achievement in Elementary and Middle Grades** → **Higher High School Achievement** → **Better Long-Term Student Outcomes**

Improved School Climate and Classroom Instruction ↓
- Fewer student discipline referrals and suspensions and increased time on task
- Improved attitudes toward school and expectations for student achievement
- Improved student attendance

Higher Student Achievement in Elementary and Middle Grades ↓
- Higher test scores
- Larger percentage of students who meet state and district standards

Higher High School Achievement ↓
- Higher test scores
- Increased on-time promotion to the next grade
- Higher high school graduation rates

Better Long-Term Student Outcomes ↓
- Higher college enrollment rates
- Higher college graduation rates

NOTE: [a]Success for All and MOVE IT™ Math are the Project GRAD preferred components.

doing, Project GRAD seeks to implement a set of reforms that follow students in a feeder pattern from the elementary level through the end of high school. Simultaneously, Project GRAD offers support for students at all grade levels, even if they have not had the benefit of this initiative in their early school years. Project GRAD does not require the realignment of staff or new staff. It is implemented with the existing student body in the schools and with the schools' current administrators and teachers, in an effort to "work with the existing assets."[8]

Project GRAD's components, academic support, classroom management, social services, parental involvement, and other activities throughout the feeder pattern are intended to motivate students, to increase their engagement and achievement, to provide important skills and tools for teachers, and ultimately to help build community expectations. The Project GRAD approach as an improvement strategy for middle and high schools is influenced by the extent to which real feeder patterns exist where it operates. Since the majority of its efforts to improve academic instruction come directly in the early grades (and mostly at the elementary school level), much of its force at the middle and high school levels rests on a growing percentage of students' arriving at Project GRAD high schools with better academic preparation.

Operational Strategies

As detailed in Box 1.1, Project GRAD employs three operational strategies: (1) creating a school and classroom environment that is conducive to learning, (2) strengthening opportunities to learn, and (3) enhancing school-level capacity for program management and implementation. The initiative draws on specific program components and on the coordination and quality assurance provided by the local and national Project GRAD organizations, described in the next section. In short, Project GRAD combines a scholarship guarantee and a set of interrelated initiatives selected to improve the focus and motivation of high school students with a set of curricular, instructional, and behavioral reforms designed to improve the preparation of students arriving at Project GRAD high schools.

Key Components of Project GRAD

One tenet of the Project GRAD approach is a commitment to work with existing students, teachers, and administrators in the schools that are part of the feeder pattern selected for program implementation. As a result, districts that choose Project GRAD are not expected to

[8]See Project GRAD USA (2004a, 2004b).

Box 1.1

Project GRAD's Operational Strategies

- **Creating a school and classroom environment that is conducive to learning.** Project GRAD is typically introduced in schools that have discipline problems, weak student engagement, and a low sense of efficacy and pessimism about the possibility of academic success. Project GRAD feeder patterns are frequently described as containing the lowest-performing high schools and many of the most troubled elementary schools. In response, Project GRAD seeks to address students' social service needs, increase parental involvement, strengthen school-community linkages (through *Communities In Schools [CIS]* or *Campus Family Support [CFS]*), promote student self-discipline and engagement (through *Consistency Management & Cooperative Discipline [CMCD℠])*, and create excitement about the possibility of going to college (through a scholarship offer and summer institutes on college campuses and special Project GRAD activities such as the *Walk for Success*). The annual Walk for Success — which includes visits to the homes of students in the feeder pattern by school staff and community representatives — is sponsored each year by the local Project GRAD office to build awareness of the scholarship offer and the goals of the initiative. Both CIS and CMCD℠ are offered at all grade levels. Though only high school students who meet Project GRAD's requirements can participate in the summer institutes and receive a scholarship, the Project GRAD scholarship is also publicized in elementary and middle schools, to help improve students' (and teachers') motivation and create a sense of possibility for real change.

- **Strengthening opportunities to learn.** Low student achievement is often one reason that a district or group of schools adopts Project GRAD. An early goal of the initiative is to strengthen instruction in the lower grades, to help improve the basic academic skills of students and to prepare them for the secondary grades. Key Project GRAD components focus on literacy (typically, *Success for All [SFA]*) and mathematics (preferably, *MOVE IT™ Math*) in grades 1 through 6, to provide students with the skills they need to succeed in other subjects and in later grades. SFA has begun a middle school program that was pilot-tested in Project GRAD sites, and MOVE IT™ Math has expanded to the middle school grades.* The use of these middle school curricular components varies from site to site. The summer institutes for high school students (some of which are residential in specific sites) also provide an opportunity to address specific skills gaps among students and to enrich the educational offerings in the Project GRAD schools.

(continued)

Box 1.1 (continued)

- **Enhancing school-level capacity for program management and implementation.** In each Project GRAD site, a *local nonprofit organization* is formed to support the implementation of the program through the feeder pattern(s) and to coordinate with the local district. This local organization works with staff at the district and school-building level to implement the Project GRAD components effectively, and it plays a key role in building local support for the initiative and school reform efforts, by developing strategic partnerships with other institutions and by fundraising. Although the degree to which the local organizations are connected to the school district varies, a partnership with the district is an important element in the Project GRAD implementation strategy, even though the local organizations are viewed largely as independent catalysts supporting improvement in the Project GRAD schools. The *national Project GRAD organization* — Project GRAD USA — monitors progress, provides technical assistance, coordinates component service delivery, develops new components and approaches, and provides funding to the local not-for-profit organizations. Project GRAD USA also regularly convenes the Project GRAD network and provides fundraising support.

---

[*]When districts do not implement SFA or MOVE IT™ Math in Project GRAD feeder patterns, Project GRAD endeavors to support or supplement the district's curriculum choice. For example, support might include reading consultants or a reading manager for each school, along with special teacher training and materials.

select specific teachers or to reassign principals or teachers or other staff to implement the program.[9] Though Project GRAD is an evolving reform, its key components include:

- **The Project GRAD college scholarship.** The program's offer of a college scholarship is often referred to as the cornerstone of Project GRAD because it most directly represents the ultimate goal: increasing student enrollment and success in college. Through this component, Project GRAD seeks to

---

[9]Project GRAD describes its structural components as including the feeder system, the local Project GRAD organization and Project GRAD USA, community involvement and collaboration, and working with the district's existing assets (Project GRAD USA, 2004b).

raise the academic expectations of students in kindergarten through grade 12 (K-12) by providing a financial incentive and college awareness and preparation activities. Scholarship amounts and criteria vary slightly by site but usually average $1,000 to $1,500 each year during the four years of college. Project GRAD scholars are high school youth who sign a contract to document their intent to meet the scholarship criteria.[10] The program places a scholarship coordinator in each Project GRAD high school to help encourage students to pursue the scholarship and college access. The scholarship coordinators provide counseling, tutoring, and information about eligibility for the Project GRAD scholarship and other sources of financial support for college. They help students prepare for college entrance examinations, and they monitor students' progress in meeting graduation requirements. Scholarship coordinators build college awareness and guide scholars through the college selection and admission process.

- **Summer institutes.** Typically offering four to six hours per day of instruction and related activities for four to six weeks, the summer institutes provide an opportunity for Project GRAD scholars to experience a college-campus-based program taught by college faculty and to enhance their academic skills. The institutes vary by site but typically include reading, writing, math, science, and enrichment and as needed remedial activities. (See Box 1.2.)

- **Classroom management.** *Consistency Management & Cooperative Discipline℠ (CMCD℠)* is a classroom management and discipline program that seeks to build student support for classroom management, responsibility, and self-discipline by promoting cooperative learning and positive working relationships among students, teachers, and other adults in the school. CMCD℠ moves beyond the traditional concept of student discipline as a set of adult responses to a student's negative behavior to offer a comprehensive classroom management program that is intended to build a shared sense of responsibility among students and adults in the building. Usually, CMCD℠ is put in place in all grades at the Project GRAD schools.[11]

---

[10]Project GRAD's college scholarships are for students who have a cumulative 2.5 grade point average, graduate within a four-year period, complete a recommended college preparatory curriculum, and participate in two summer institutes designed to expose them to an enhanced curriculum while introducing them to college. Students who receive the scholarship also get follow-up support from Project GRAD while they are in college, expanding Project GRAD's reach to K-16.

[11]Project GRAD USA now has an agreement with CMCD℠ that allows the national organization to manage this component at each site after two and a half years of program implementation.

Box 1.2

The Summer Institutes

During their four years of high school, Project GRAD students are expected to attend two four-week summer institutes on local college campuses. These sessions are planned by the Project GRAD staff in conjunction with the college and usually are taught in part by college staff. Project GRAD students get real-life perspectives on college as they go back and forth to classes, eat in the cafeteria, and have access to other amenities that make up the college experience. Some sites provide stipends or make special arrangements for transportation so that students can attend the institutes. Some sites also have residential programs whereby the high school students live in the college dorms during their summer experience.

- **Social services and parental involvement.** *Communities In Schools (CIS)* is a national program that brings additional support (that is, volunteers, social services, and academic enrichment and support activities) directly into schools. CIS is implemented in all grades at Project GRAD schools. In some sites, each school has a full-time CIS staff person. The CIS component seeks to enhance social and academic support services available through the school and to provide targeted assistance to students who have problems outside school that may affect their classroom performance. The range of activities may include guidance counseling, community outreach, and family case management. Project GRAD views parental involvement as an important strategy to help build family aspirations for academic success, meeting graduation requirements, and going to college. Project GRAD USA has developed the *Campus Family Support (CFS) Plan,* which contains similar elements for school districts that have no CIS program. Coordinated by CIS or the CFS staff, Project GRAD conducts an annual Walk for Success, during which students' homes are visited by the Project GRAD staff, district staff (including principals and teachers), and community leaders, who encourage parents to connect with their child's school and who explain Project GRAD and the scholarship offer. (See Box 1.3.)

- **A research-based literacy program.** The elementary schools in the feeder pattern put in place a research-based literacy program. Most, but not all, Project GRAD sites use *Success for All (SFA),* a nationally recognized reading program that promotes comprehensive restructuring of school resources to

Box 1.3

The Walk for Success

During a time period when students' homes are seldom visited by school officials unless there is a problem, Project GRAD has initiated a unique strategy to bridge home and school — and ultimately to engage parents, community leaders, and other key stakeholders in school reform.

Project GRAD's annual Walk for Success in the fall is a community outreach effort in which volunteer principals, teachers, other school-level staff, district administrators, representatives from the business community, and key community leaders visit the homes of students attending Project GRAD schools. It is not unusual for the volunteers to also include college and university partners, parents, and older students in Project GRAD schools or for the total number of participants to include several thousand people. Typically, teams of two go door to door visiting the homes of ninth-graders on a designated Saturday and share background information about Project GRAD, encourage the students and their families to sign the scholarship contact, and conduct a "needs survey" to see whether the families require other assistance. The contract specifies the eligibility requirements for the Project GRAD scholarship and describes the services and activities that the initiative provides. In Houston, volunteers include Spanish-speaking individuals, and the information is provided in both English and Spanish.

Project GRAD volunteers in some sites also visit the families of students in the early elementary grades (K-2) and in the sixth grade. Parents of elementary and middle school students are encouraged to sign a pledge that helps encourage their children to pursue academic achievement and college. Some sites hold Walk for Success rallies in central locations and showcase high-profile individuals from a variety of fields (noted authors, athletes, political leaders, and so on) who serve as motivational speakers and role models for students, encouraging them to prepare for college and to be optimistic about their future.

The Walk for Success helps break down the barriers that many parents feel exists between schools and home, particularly when parents may have unhappy memories of their own school experiences. In essence, Project GRAD brings the school to parents' doorsteps. The Walk for Success is an important constituency-building activity for Project GRAD, providing a way for an array of stakeholders to become involved in the program and to help support school improvement efforts.

provide concentrated instructional time (90 minutes each day) for reading, to bring students to grade level in this subject area by the third grade.

- **A research-based math curriculum.** The elementary schools in the feeder pattern also put in place a research-based math curriculum. Project GRAD's preferred math component is *Math Opportunities, Valuable Experiences, Innovative Teaching (MOVE IT™ Math)*, which is organized around the National Council of Teachers of Mathematics (NCTM) Standards. It is a K-8 professional development program that advocates math instruction based on the use of manipulatives to address a wide variety of learning styles.[12] (MOVE IT™ Math emphasizes daily problem solving and introduces algebra in the early grades.)

Intended Effects on Student Outcomes

The Project GRAD components are hypothesized to improve the school climate and classroom instruction as shown in the bottom panel of Figure 1.1. The expectation is that the program will result in reduced discipline referrals and suspensions, improved attitudes toward school, higher expectations, greater time spent on task, and improved teaching and learning. These effects, in turn, are expected to lead to improved student achievement, as indicated by higher average achievement test scores and greater numbers of students performing at grade level, particularly in the lower grades. These improvements in achievement are also expected to reinforce the cycle of improved school climate and instruction.

The momentum created by improved student achievement in the lower grades is then hypothesized to provide a foundation for increased achievement at the middle and high school levels, as well as prepare students to perform better in existing courses and create a demand for higher-level courses. Over time, this is expected to generate increases in both the offerings and the completion of more demanding courses. This better high school preparation is in part motivated by the offer of the Project GRAD scholarships, which are able to offer an immediate reward: the chance to earn a scholarship to go to college. Finally, this improved student achievement then produces increases in the number of students completing high school, going on to college, and receiving a degree. The Project GRAD planners envision that success in one feeder pattern will lead to adoption of the initiative in other low-performing feeder patterns, up to the point where Project GRAD reaches a critical mass in the district and, as a result, develops "staying power," enabling the initiative to play an important role in driving change throughout the district.

[12]Project GRAD USA has licensed the right to serve as the national implementation manager of MOVE IT™ Math.

The Project GRAD Implementation Effort

It is critically important to understand the scope and complexity of the implementation task that Project GRAD undertakes. As soon as the initiative is announced, program staff begin working at once in 8 to 10 schools in a district and with whatever administrative unit or units the district has in place to manage the set of schools. The Project GRAD components are typically phased in over two or three school years, and several of them (CMCD℠, SFA, and MOVE IT™ Math) require staff development and training prior to implementation. Most components require support personnel (managers, coordinators, and facilitators) to help build the capacity needed for teachers and principals to implement them effectively. SFA and MOVE IT™ Math require program materials and special classroom artifacts, and CMCD℠ also requires special artifacts. Finally, the network of component staff must be coordinated by the local Project GRAD office in and across the schools in the feeder pattern.

As a national initiative, Project GRAD USA coordinates 12 districts, including rural sites in Alaska, and anticipates that several additional districts will join in the near future. Currently, 211 schools are part of the initiative in a network that includes 21 high schools, 34 middle schools, and 149 elementary schools. (Alaska's Kenai Peninsula has 7 schools that contain both elementary and secondary grades.) Project GRAD Ohio is a special statewide effort in which four of the state's largest districts are implementing the program.

The Project GRAD Evaluation

This report grows out of MDRC's independent third-party evaluation of the Project GRAD national expansion sites. It discusses the model's approach to school capacity-building and the specific components that are used in the initiative; summarizes Project GRAD's implementation in Houston — the initial site — and in three expansion districts (Atlanta, Columbus, and Newark); and presents what has been learned thus far about the impacts of this major school reform effort on students' academic performance at the elementary school level.[13] A companion report analyzes the early effects on high school outcomes in selected sites.[14]

Key Questions Addressed in This Report

The key research questions addressed in this report are:

---

[13]The initial expansion sites were Atlanta, GA; Columbus, OH; Los Angeles, CA; Nashville, TN; and Newark, NJ. Los Angeles is not included in this analysis because the district's data system does not provide the historical information that is required by the research methodology in order to assess program effects. Nashville is no longer a Project GRAD site.

[14]See Snipes, Holton, Doolittle, and Sztejnberg (2006).

- What was the local context in which Project GRAD was implemented, and what challenges were faced in its implementation?

- What are the impacts, or effects, of Project GRAD in Houston, the program's original site, on elementary students' outcomes, over and above the changes that would have been observed in the absence of the intervention?

- What is the impact of Project GRAD on elementary students' achievement in the expansion sites through the first several years of implementation?

- How do Project GRAD's impacts relate to the quality of program implementation and the nature of the context in which implementation occurred?

The Approach for Assessing Program Effects

In each site, Project GRAD was implemented in feeder patterns of elementary and middle schools that fed into a particular high school. Given this structure, it was not feasible to randomly assign schools or students to program and control groups. Therefore, this analysis is based on the application of an interrupted time series (ITS) framework to the available data in each site. Described more fully in Chapter 3, this approach rests on three basic comparisons:

- A comparison of post-program achievement at Project GRAD schools and pre-program achievement at those same schools to determine whether there were improvements over time

- A comparison of baseline and follow-up achievement patterns at a set of similar schools from the same district, to determine whether these schools — without Project GRAD — also experienced improvements over time

- A comparison of the changes from baseline achievement patterns at both the program schools and the comparison schools, to determine whether the improvements at the Project GRAD schools are greater than at the comparison schools

As a framework for the discussion of Project GRAD's results, this report covers three important areas. First, it discusses the context in the school districts where Project GRAD implemented its reforms. Second, it tracks the implementation process and describes how the initiative's components were put in place, to better understand the degree to which Project GRAD was in place in the sites over the time period studied. Finally, it addresses whether Project GRAD was able to generate improvement outcomes greater than those achieved by a similar set of comparison schools from the same district.

The Evaluation as a "Differential" Impact Study

This report focuses on the elementary school level and examines impacts on student achievement in reading and math. It is a *differential impact study*, meaning that it compares changes in student outcomes at Project GRAD's elementary schools with changes in student outcomes at comparison schools that may be implementing other reforms or that are subject to other school improvement efforts undertaken by the district.

Project GRAD proves to be attractive to districts that are interested in a reform agenda, and — in an era of accountability and emphasis on low-performing schools — it is seldom the only approach being utilized. For example, almost every school in most urban districts is doing something to improve student achievement, and districts are increasingly playing a direct role in instructional change and in improvement efforts at the building level. In some cases, districts adopted Project GRAD as one strategy for some schools in a broader, districtwide improvement plan. As a result, when Project GRAD schools are compared with similar schools, this typically means that the comparison schools are implementing other school-level reform strategies and approaches and are subject to districtwide improvement efforts. In the proliferation of comprehensive school reform models — and most recently with the advent of No Child Left Behind legislation — low-performing schools that have not been targeted for some type of school improvement are rare. These school-level reform efforts are summarized in Chapter 2, which discusses Project GRAD's implementation in each site.

In addition, during the period of Project GRAD's implementation, there has been a heightened climate of oversight and accountability because of federal and state initiatives. To varying degrees across the study sites, specific mandates for required improvement in a range of student outcomes, including benchmarks and timetables, were set for all schools in the districts. As a result, student performance outcomes were routinely reviewed by the district, principals, and model developers and program implementers. Further, over time, districts increasingly became involved in making curricular and instructional decisions at the elementary school level, which affected both Project GRAD and the comparison schools. These district initiatives led Project GRAD to become somewhat more flexible in the choice of curricular components. Formal and informal networks also supported efforts of schools implementing other non-GRAD reforms, and for some this also included coordination and technical support. Specific strategies that focused on improving instruction and teacher development emerged; although different from Project GRAD in their content, they were similar in their intent. In some instances, districts provided schools with other types of supports that were not available to Project GRAD schools, and, conversely, Project GRAD schools received support from districts over and above what their non-Project GRAD counterparts received.

Thus, in this evaluation, schools adopting Project GRAD faced a specific, ambitious, but policy-relevant standard: Impacts are defined as making more progress than comparison group schools — which were also making considerable efforts to improve test scores. To make the analysis clear and to document the progress that did occur, findings are presented in three stages:

1. Were there improvements at the Project GRAD schools over time (from the baseline period into the follow-up period)?

2. What was occurring at similar comparison schools over the same period?

3. How do changes from the baseline to the follow-up period at Project GRAD schools differ from changes at the comparison schools?

In summary, in studying an education reform effort like Project GRAD, both the strength of its implementation and the other local contextual features affecting similar schools can influence the impacts. Chapter 2, in discussing the implementation of Project GRAD, presents a framework that describes the experiences in Houston and in the three expansion sites as illustrations of different combinations of implementation strength and local context. This perspective can help readers to interpret the impact findings presented in the later chapters.

Topics Addressed in This Report

The remainder of this report is organized as follows:

- Chapter 2 provides an overview of the Project GRAD model's evolution in Houston, the flagship site. It also describes the implementation experience in the expansion sites and presents key lessons that have emerged from that experience.

- Chapter 3 describes the basic approach to the report's impact analysis.

- Chapter 4 discusses program impacts in Houston's elementary grades, and Chapter 5 covers impacts in the expansion sites.

- Chapter 6 summarizes the findings across the Project GRAD sites and places them in the broader context of school reform and research on comprehensive school reform.

Chapter 2

The Implementation of Project GRAD

This chapter provides an overview of the implementation of Project Graduation Really Achieves Dreams (GRAD) in Houston — the original site — and in the expansion sites covered in this evaluation. In interpreting findings from the evaluation, it is important to understand the timing and strength of Project GRAD implementation and the context in which the program was operating. This helps set expectations about when impacts might be expected to emerge, and it illustrates the nature of the comparisons with non-Project GRAD schools undertaken in each site.

The Evolution of the Project GRAD Model: The Houston Story

Project GRAD Houston developed from a business-supported scholarship offer that began at Jefferson Davis High School in Houston, Texas, in 1988. The business-school partnership provided Davis students with university-based summer institutes, mentoring, tutoring, social services, leadership development and summer jobs, and college scholarships. However, high school graduation and college enrollment rates remained lower than desired, and anticipated improvements did not occur. James Ketelsen — the former CEO of Tenneco and an active leader in the effort to reform Jefferson Davis High School — concluded that the supports at the high school level were offered too late to compensate for weak academic preparation in the earlier grades. In collaboration with the Houston Independent School District (HISD), Ketelsen sought promising reforms intended to enhance learning at the elementary and middle school levels and to enable more students over time to take advantage of the college awareness and scholarship opportunities in high school.

Table 2.1 illustrates the initiation and expansion of Project GRAD in the Houston feeder patterns. The Davis feeder pattern consists of a high school, one middle school, and six elementary schools. Project GRAD was pilot-tested in this feeder pattern in the 1993-1994 school year, beginning with the implementation of the classroom management component, Consistency Management & Cooperative Discipline℠ (CMCD℠).[1] However, not all the elementary schools began implementation that year.[2] Project GRAD implementation began at Jack Yates High School in the 1996-1997 school year, with the implementation of MOVE IT™ Math. In the 1999-2000 school year, the program began expanding into the schools that were

[1]See Chapter 1 for a full description of the Project GRAD components.

[2]Given that not all elementary schools implemented CMCD℠ in the 1993-1994 school year, this analysis considers 1994-1995 as the first year of Project GRAD.

Table 2.1

Project GRAD Implementation in Houston, Texas

Key Components for Davis Feeder Pattern	**1993-1994**	**1994-1995**	**1995-1996**	**1996-1997**
Consistency Management & Cooperative Discipline℠→	Implemented in elementary schools[a]		Implemented in middle schools	Implemented in Jefferson Davis High School
MOVE IT™ Math→		Implemented in elementary schools	Implemented in middle schools	
Success for All→			Implemented in elementary schools and middle schools	
Communities In Schools→		Implemented in middle schools and Jefferson Davis High School[b]	Implemented in elementary schools	

Key Components for Yates Feeder Pattern[c]	**1996-1997**	**1997-1998**	**1998-1999**	**1999-2000**
Consistency Management & Cooperative Discipline℠→		Implemented in elementary schools	Implemented in middle schools	
MOVE IT™ Math→	Implemented in elementary schools	Implemented in middle schools		
Success for All→		Implemented in elementary schools		
Communities In Schools→		Implemented in elementary schools		Implemented in middle schools[d]

(continued)

Table 2.1 (continued)

Key Components for Wheatley Feeder Pattern	1999-2000	2000-2001	2001-2002	2002-2003
Consistency Management & Cooperative Discipline℠→	Implemented in elementary schools[e] Implemented in middle schools[e]			Implemented in Wheatley High School
MOVE IT™ Math→	Implemented in elementary schools[f] Implemented in middle schools			
Success for All→	Implemented in elementary schools[g]		Implemented in middle schools[g]	
Communities In Schools→	Implemented in elementary schools[h] Implemented in middle schools and Wheatley High School			

SOURCES: Opuni (2005); Opuni and Ochoa (2002a).

NOTES:

[a]Ryan Jefferson and Lamar Elementary Schools started in 1993-1994. Four other Davis elementary schools began in 1994-1995.

[b]Communities In Schools (CIS) in the Davis feeder pattern operated prior to Project GRAD. Funding support from Project GRAD began in 1995-1996.

[c]CIS at Yates High School predated Project GRAD.

[d]Cullen Middle School began CIS in 1999-2000. Ryan Middle School began CIS before Project GRAD.

[e]Several elementary and middle schools began Consistency Management & Cooperative Discipline℠ (CMCD℠) in 1999-2000. Other elementary and middle schools began CMCD℠ in later years.

[f]Several elementary schools began MOVE IT™ Math in 1999-2000. Other elementary schools began MOVE IT™ Math in 2000-2001.

[g]Success for All (SFA) began in several elementary schools in 1999-2000. Other elementary schools began SFA in later years. SFA began in middle schools in 2001-2002 and in other middle schools in later years.

[h]CIS began in several elementary schools in 1999-2000. Other elementary schools began CIS in later years.

part of the feeder pattern for Phillis Wheatley High School with CMCD℠, MOVE IT™ Math, and Communities In Schools (CIS), the social services and parental involvement component.[3]

Project GRAD Houston, the initial site, benefited from having some familiarity with the core program components prior to their inclusion in the model and in having geographic proximity with three key component developers, creating a direct feedback loop regarding implementation challenges and needed refinements. CMCD℠ was developed at the University of Houston, was operated by an independent contractor, and initially was pilot-tested in the Houston schools. MOVE IT™ Math — the first curricular component that was put in place at the elementary schools that feed into Davis High School — was initially an independent program, but Project GRAD became the program operator in response to the development demands of the scaling-up process. CIS, largely viewed as a dropout prevention program, was an established service provider in the Houston schools and already operated at the first Project GRAD high school; coincident with Project GRAD's implementation, CIS was refined and strengthened. In contrast, Success for All (SFA) existed as a stand-alone reading program at some schools prior to its integration into the Project GRAD model and was operated by an independent developer with a national infrastructure.

Project GRAD's design and early experience were heavily influenced by the particular challenges faced and opportunities discovered in Houston. This unusual history influenced how the approach evolved, was presented to and perceived by potential expansion sites, and was replicated and scaled up. The Houston district context also shaped both the program and the organizational structure that emerged in the subsequent expansion effort.

The Houston setting in the late 1990s was conducive to Project GRAD's development and implementation. The district's priorities and strategic vision were compatible with those of Project GRAD, and the newly developed regional administrative structure and feeder patterns for student assignment facilitated its implementation. The district was willing to give the Project GRAD implementers significant flexibility in the day-to-day operation of the initiative. Further, Project GRAD's administration was eased by the flexibility in staffing and staff development days that were possible in a district without strong teacher and principal unions and without collective bargaining agreements covering these areas. As the district permitted Project GRAD to expand from its original feeder pattern to multiple high schools, Project GRAD's developers were able to refine the model and to add and eliminate elements during what was an extended pilot-test phase.[4] Thus, when Project GRAD expanded to other districts, its developers were

---

[3]The Project GRAD scholarship for Wheatley High School began earlier, in 1997-1998. In addition, two other Houston feeder patterns (the Reagan and Sam Houston patterns) began implementing Project GRAD later but, given the shorter implementation period, are not included in this analysis.

[4]When Project GRAD was launched in Houston, it also included or pilot-tested other services that were not continued in the later feeder patterns or sites, and the initiative built on existing initiatives in HISD. To reinforce (continued)

optimistic that the structure of the reform was established, and they were ready to specify its architecture for additional sites.

Organizational Evolution as Project GRAD Expands

In the late 1990s, reports of the improvements in academic achievement at the Project GRAD Houston schools attracted the attention of national funders and other urban districts seeking a solution to poor academic performance and low rates of high school graduation and college enrollment. These early trends were especially noteworthy because they represented the academic performance of students who had received only part of the full Project GRAD program — that is, they had not been exposed to the program in elementary or middle school. Such encouraging news coming out of Project GRAD's schools stimulated interest in the expansion effort.

As the initiative expanded, Project GRAD continued to rely on the core service providers that had participated in implementing key elements of the reform in Houston. For SFA, Project GRAD's national expansion could be accommodated within an existing national organization; for CMCD℠ and MOVE IT™ Math, however, the expansion required rapid development of new organizational structures and expanded delivery systems in order to work outside Houston. Both components lacked trainers and other resources needed to operate in several sites at once. Furthermore, Project GRAD's reliance on CIS (which is a national association of local affiliates) for key services presented a need for local organizational development in sites that did not have a local CIS program or redevelopment in sites where the existing CIS program was not immediately able to take on the role envisioned by the Project GRAD context. Although CIS had previously operated in Newark, Atlanta, and Columbus, the program in each of these sites had to refocus the services provided to Project GRAD schools in order to be consistent with the approach used in Houston.

The move to sites beyond Houston surfaced a range of typical expansion issues: quality control of the service delivery system of developers and contractors, coordination among and across components, the relationship of the original site to the growing expansion network of independent not-for-profit organizations, the need for implementation monitoring, and establishing cooperative agreements (programmatic and financial) with school districts.[5]

---

the leadership skills of principals, who played a central role in implementing the initiative, Project GRAD supported their participation in leadership training at the Harvard School of Education. At one point in the early years of Project GRAD in Houston, the University of Chicago's middle school math program was used at the middle school level, but this aspect of Project GRAD was not part of the expansion program. Project GRAD Houston also was able to offer summer programs to introduce incoming sixth- and ninth-graders to the new schools and to ease their transition into middle and high school.

[5]For a discussion of the implementation progress and challenges, see Holland (2005).

As the expansion sites began to operate and worked with the original Project GRAD staff in Houston, an inevitable tension arose between the new sites' desire to adapt Project GRAD to local circumstances and the program developers' desire for fidelity to the original model. During the early phase of the expansion period, new sites were required to implement four of the five Project GRAD components, and some sites were resistant to this approach. Over time, Project GRAD has moved away from this requirement. It is important to note this tension because those seeking to expand school reforms (and other types of social program innovation) often must decide whether they wish to *replicate* the original vision of the reform in new settings or *adapt* its core model to make it more appropriate to specific conditions in new settings.[6]

These issues and the continued growth of the network of sites led to the creation of a new organization, Project GRAD USA, which became operational in 2000. This national organization focused on implementation issues across the several Project GRAD districts, and it supported the local organizations in spearheading implementation efforts in individual school districts. Project GRAD USA faced numerous challenges. As new sites were added, it had to develop an infrastructure to support expansion and guide implementation. It had to package several of the model's components to be more user-friendly, and it had to monitor the quality of the components as they operated in new sites. And when district mandates and the initiative's strategies were not aligned or were in conflict, Project GRAD USA had to serve as broker between the sites and the school districts. Where unions operated in the districts, Project GRAD USA and local Project GRAD had to learn ways of working within the context of union agreements. Most important, the national organization had to help the local organizations position Project GRAD so that teachers and principals saw it as an important strategy to meet their goals and improve scores on the required state assessments.[7] Finally, Project GRAD USA was created after several sites had begun implementation. As this organizational evolution occurred, the expansion sites received varying levels of support, depending on when they joined the Project GRAD network; Newark and Columbus grappled with these issues early on, while Atlanta — a newer site — received more support from Project GRAD USA.

The Expansion Sites Included in the Analysis

Project GRAD's selection of expansion districts and schools has generally begun with the identification of districts facing serious educational challenges and a specific high school that is perceived to be particularly troubled, which will serve as the home for the Project GRAD scholars

[6]At times, the expansion sites expressed frustration with Project GRAD's requirements about the use of specific preferred components that they saw as unable to fully meet local needs. At varying points, some sites argued for the option of local choice for one or more of Project GRAD's central components.

[7]In fact, eventually Project GRAD moved away from specifying specific curricula for math and reading and positioned the initiative as a support mechanism for the district's chosen curricula.

and scholarships. All the expansion sites were facing serious academic challenges, and the feeders that were identified as appropriate for Project GRAD served low-income students, most of whom were in minority racial or ethnic groups. Across all the new sites, academic achievement in the new Project GRAD feeders was below the overall level of performance in the local district and tended to decline relative to state or national norms as students advanced through the grades. Further, in the sites where achievement was measured for several years prior to the start of Project GRAD, using comparable tests, trends in achievement were either declining or at best stable. Thus, Project GRAD succeeded in expanding into districts and feeders that met its criteria and were in need of a new education initiative to improve academic outcomes for students.

The school districts included in this report represent a substantial expansion and adaptation of the Houston-based Project GRAD model to new feeder patterns of schools in very different settings within a relatively brief time period. Although each of the districts is largely minority, urban, and poor, they differ from Houston in important ways, lending support to the idea that Project GRAD can take root in many different settings. For example, contextual differences from Houston include how (and by whom) the initiative was first introduced, the presence and strength of unions, different state and local accountability systems, and variations on the feeder pattern concept — all of which affected the process of implementation. The diversity of Project GRAD sites is an important finding in assessing the general applicability of the initiative.

Overview of the Expansion Sites

Tables 2.2, 2.3, and 2.4 in the following sections detail the implementation timetables for the three expansion sites in this study: Newark, Columbus, and Atlanta. Three key themes emerge from the information in the exhibits. First, the sites began implementing Project GRAD at different times, so the length of follow-up available for an analysis of student outcomes varies. Second, the model's components were not implemented in the same order across the sites. In each site, the launch of Project GRAD began with an announcement of the scholarships, but, beyond that, there were considerable differences in how the program unfolded. Choices were driven by local perceptions of most pressing needs and by component developers' capacities to take on new sites at specific times. Third, there was some variation across the sites in the specific curricular approaches implemented. While Project GRAD USA — particularly during the period covered by the evaluation — preferred sites to adopt the SFA reading program and MOVE IT™ Math, local pressure and desires sometimes dictated other choices. This variety is illustrated by the following site profiles, which present how the initiative began in each district, the district's context, the role of various organizations, and the accomplishments as Project GRAD was implemented.

Newark, New Jersey

Project GRAD Newark was the first expansion site and began implementation in the spring of 1998.[8] The feeder pattern consisted of seven elementary schools, a middle school, and a high school. The first Project GRAD high school was frequently described as the district's worst comprehensive high school, with low student and faculty engagement, serious discipline issues, an emerging gang problem, and low student achievement and graduation rates.[9]

The Newark Public School District was the subject of a state takeover in 1995, resulting from concerns about weak educational and financial management, low student engagement and attendance, discipline problems, poor academic achievement, and low graduation rates. Further, the district was part of a landmark series of state court cases that found that the system of educational finance had not provided sufficient financial resources to high-need districts to allow them to provide an adequate educational opportunity. Immediately prior to Project GRAD's implementation, each school in Newark was required to select one of five comprehensive school reform models as part of the remedy. The SFA reading program was one of these models, but Project GRAD was not. As a result, some schools had invested considerable time investigating the five options and had already expressed interest in a model other than SFA, and most were unfamiliar with Project GRAD.

The state-appointed superintendent strongly embraced Project GRAD as part of a broader reform strategy. Further, the initiative was supported by two foundations that publicly pledged to support the initiative for the first five years. A local not-for-profit board was created, which included the state-appointed superintendent and community representatives and several representatives of one of the key funders. In the early years of implementation, a former corporate executive led the implementation process, followed by the former principal of Malcolm X Shabazz High School. The early implementation activities in Newark predated the creation of Project GRAD USA as a not-for-profit organization and of a comprehensive new-site development or orientation plan.[10] A former principal from Project GRAD Houston was designated as the national implementation director, and her role was to provide guidance and coaching to the sites. Newark principals and other district administrators visited Houston to observe how the model operated and how the components provided training and staff development, but most of the implementation process was locally crafted, and negotiations with the developers and contractors who operated the Project GRAD components were done independently. During this

[8]Teacher training for CMCD℠, the classroom management component, began in the spring, and CMCD℠ was introduced into the classrooms in the fall of 1998.

[9]In 2000, the Newark Public School District and Project GRAD funders subsequently selected a second high school feeder pattern for further expansion, but this group of schools is not included in the MDRC evaluation.

[10]Subsequently, Project GRAD USA developed a comprehensive handbook and training program for new sites: *New Site Development Handbook* (Project GRAD USA, 2004c).

Table 2.2

Implementation Timetable for Project GRAD Newark

Core Elements of Project GRAD	Status in Newark, New Jersey
Implementation begins	1998-1999
Classroom management: Consistency Management and Cooperative Discipline℠ (CMCD℠)	Began in the fall of 1998 for grades K through 8 and in the fall of 1999 for high school.
Social service support: Communities In Schools (CIS) or Campus Family Support (CFS)	CIS began in the fall of 1998. In the fall of 2003, some schools switched to CFS.
First summer institute	1998
Reading: Success for All (SFA)	Began in the fall of 1999 for elementary schools. An SFA middle school pilot-test began in 1999 and was discontinued in the fall of 2002. A district-adopted reading program (for all schools) began in the fall of 2004, replacing SFA.
Math	Initially, district opted not to use MOVE IT™ Math and continued using district-developed math program. Replacement program, Math Wings, began in the fall of 2001, up to the spring of 2003, for grades K through 8. Replacement program, Everyday Math, began in the fall of 2003.
First graduating class	2000-2001 30 Project GRAD scholars

period, strong funding from the two anchor foundations provided special resources to address many implementation problems, but Project GRAD was so strongly identified with its anchor funders that the development of a broader local funding base was delayed.

The social services component, CIS, began in the fall of 1998,[11] and the first summer institute was held in the summer of 1999. SFA, the reading program, began in the fall of 1999 at the elementary and middle school levels (the middle school pilot-test was discontinued in 2003). The district opted to continue its own math program initially, but, in the fall of 2001, SFA's Math Wings program began in the Project GRAD schools. In the fall of 2003, the district switched to Everyday Math for all the schools in the district. By the end of its fourth year of

[11]Prior to this, a special CIS Academy operated as a "school within a school" in the high school, and it was phased out over time as Project GRAD was implemented. To service the Project GRAD schools, Communities In Schools of New Jersey, a statewide organization, had to create a Newark organization on an aggressive timetable. In 2003, some Newark schools switched to Campus Family Support (CFS).

implementation (spring of 2001), all five Project GRAD components were in operation.[12] The first graduating class in 2001 included 30 scholars.

Newark was Project GRAD's first exposure to a district with a strong teachers union, and it was the first time that the implementers had negotiated with a district other than Houston, and there were inevitable missteps. The announcement of the initiative and the scholarship offer came as a surprise to the teachers and many administrators, who would then be asked to implement it. Although teachers were asked to vote for the initiative, some felt that the expectation was that Project GRAD would eventually be implemented regardless of the vote. And, in the absence of a concerted information-sharing process, teachers did not fully understand the scope of the initiative and its components. In this new setting, ensuring that the relationships — conceptual and operational — among the individual components were understood became an issue. Teachers and principals reported that the Project GRAD components at the school level required greater specification and needed cross-component strategies for working together. Other funders were reluctant to embrace the initiative in part because, at its inception, it was largely identified with a major funder already and in part because there were concerns that it would disrupt existing programs in which the funders had made investments.

Nonetheless, in the early years, implementation activities proceeded largely as scheduled, and Project GRAD Newark made a concerted effort to communicate the goals of the initiative and to encourage school staff to buy in to the components. In later years, the press of competing district priorities, budget issues, and a weakened Project GRAD Newark board were major challenges.

During the early period of Project GRAD's implementation in Newark, district-level reform efforts that were spearheaded by the state's takeover largely focused on improving administration, personnel operations, and financial management.[13] Gradually the district began to focus more on instruction, leading to some disagreement about the appropriate math program to offer at the Project GRAD schools and an eventual choice — as discussed above — of programs other than MOVE IT™ Math and the transition to a district-developed reading program for all Newark schools.[14] Moreover, because of the statewide litigation discussed above, individual schools in Newark were often implementing some kind of whole-school reform. The individual schools that were the best comparison matches with the Project GRAD schools for the MDRC evaluation — based on student characteristics and trends in academic performance —

---

[12]The school district is now phasing out SFA (and other reading curricula used in non-Project GRAD schools) for a district-developed curriculum that will be implemented systemwide.

[13]See Community Training and Assistance Center (2000).

[14]With the 2004-2005 school year, Newark adopted a new reading program, the Harcourt Brace Trophies Series, and it was instituted districtwide. Project GRAD provides support to this program.

include several schools that operated whole-school reform models. At least two schools used SFA independent of any other Project GRAD components.[15]

Columbus, Ohio

Project GRAD Columbus began in the fall of 1999, a time frame in which two other sites, Los Angeles and Nashville (which are not part of this study), also began implementing the program. This rapid growth of Project GRAD placed added stress on the capacity of Project GRAD Houston and its component providers to support planning and local implementation. The feeder pattern in Columbus consisted of 12 schools: a high school, 4 middle schools, and 7 elementary schools.[16] Prior to becoming the Project GRAD high school, Linden-McKinley High School was nearly closed because of declining enrollments, weak student and faculty engagement, and low rates of student achievement and graduation. It was also a low-ranked school in the Ohio accountability system, and it operated in a deteriorating facility. The school is historically important to the local African-American community, which expressed strong opposition to its closure. In reaching a decision to keep the school open, the superintendent chose the Project GRAD initiative as a signal to the community that there would be a major break with the troubled past. Although technically the Project GRAD elementary and middle schools feed into the high school in the student assignment plan, all Columbus schools operate as "schools of choice," whereby parents and students may select schools throughout the district to attend.

Based on students' low scores, the school district in Columbus was identified by Ohio's statewide performance accountability system as a district in "academic emergency." This designation requires a district to develop a continuous improvement plan, and the district is subject to monitoring and intervention by the state's Department of Education and must adhere to annual progress guidelines. Columbus is also a district that for many years operated under a citywide school desegregation plan; in the wake of final court orders, there was considerable reassignment of students.

Project GRAD was brought to the Columbus school superintendent by a local philanthropist who became the chair of the Project GRAD Columbus board and a key voice in developing the local not-for-profit organization. Project GRAD Columbus struggled to establish an organizational framework. Initially, its office was headquartered in a middle school that, during one period, declined to participate in the initiative. Project GRAD Columbus at first relied on district staff to lead the organization, which may have given rise to unclear or competing priorities. Subsequently, the initiative did not always have an identity that was clear and distinct from

[15]In 2004, the Newark Public School District began to move away from the use of whole-school reforms as a key part of its reform approach, and many such reforms are no longer in use.

[16]One of the seven elementary schools, Hudson, closed in 2004.

Table 2.3

Implementation Timetable for Project GRAD Columbus

Core Elements of Project GRAD	Status in Columbus, Ohio
Implementation	1999-2000
Classroom management: Consistency Management and Cooperative Discipline℠ (CMCD℠)	Elementary schools began training in February of 2000. Three middle schools began in January 2001. One middle school began in the fall of 2003.
Social service support: Communities In Schools (CIS) or Campus Family Support (CFS)	High school began in 1999. Middle and Elementary schools began in 2000 and 2001.
First summer institute	2002
Reading	Grades K through 5 began Success for All (SFA) in the fall of 1999, but implementation varied widely. Several efforts were made to restart implementation thereafter. The middle school pilot-test began in the fall of 2000, but was discontinued in the spring of 2003. Project GRAD reading support for the district-adopted SFA replacement program, LACES, began in 2002-2003 with one consultant for SFA schools (6 of 7 elementary schools). By 2004-2005 Project GRAD completely supported LACES for grades K-5. Support for grades 6-8 began in 2005-2006.
Math	MOVE IT™ Math began in fall 2003 for grades K – 5 and fall 2004 for grades 6-8.
First graduating class	2002-2003 51 Project GRAD scholars

another not-for-profit organization, with which it shared key staff. Project GRAD USA was not fully in place when Project GRAD Columbus began, but it intensified its support and technical assistance as it developed over time. When the local organization floundered, Project GRAD USA placed a former staff member from Project GRAD Houston on-site for an extended period of time, to help the site reorganize and enhance its implementation of the program.

CMCD℠, the classroom management component, started in the Columbus schools in the 1999-2000 school year. Prior to Project GRAD, CIS operated in Columbus but did not provide school-level services; as a result, significant service adaptation was required. CIS began in the high schools in 1999, and it began in the elementary and middle schools in 2000. The SFA reading program began in the fall of 1999, but there was strong and persistent resistance from some teachers, who preferred another reading approach that was operating in the district. SFA was further constrained by the failure of Project GRAD Columbus to have in place a manager who was focused on the implementation of the reading component until well after the implementation activity had begun, and implementation was markedly uneven across schools. The first real summer institute was held in 2002, as the planners learned from an earlier unsuccessful effort. MOVE IT™ Math began in 2003 for grades K-5 and in 2004 for grades 6-8, after an important but difficult process whereby the component was aligned with the state's standards and staff development and training activities were refined. The math component operated in grades K-8. Implementation across schools has been inconsistent, however, and the scheduled expansion to a second feeder pattern was delayed.

In summary, the scaling-up process and the implementation pressure of several new cities — coupled with expansion in Houston — stretched the resources of Project GRAD's component developers, and at times this was reflected in the quality of the staff training that was provided to the expansion sites. Columbus, with a fragile organizational structure, was less able than Newark to resolve some of the problems in a timely fashion. Moreover, Columbus struggled to build momentum and gain buy-in at the individual school level. As a result, Columbus made several attempts to jump-start the initiative — in essence, to start over. As Project GRAD USA itself developed, its support became more strategic and more comprehensive, and at one juncture it placed an implementation expert on-site to help restructure the Columbus implementation strategies. The first class of Project GRAD scholars in Columbus graduated in the spring of 2003. Project GRAD Columbus is now extended to a second feeder pattern, the Marion-Franklin High School, which serves over 9,000 students. This feeder pattern is not part of the present evaluation.

Districtwide reform efforts developed more slowly in Columbus than in the other sites in this study, though the district did face strong pressures to improve schools because of a poor rating under the state's accountability system. At non-Project GRAD schools, several reading programs were in place; Comprehensive Literacy Reading: The Four Block Model and Direct Instruction operated in the elementary schools, as did SFA prior to Project GRAD.[17]

Atlanta, Georgia

Implementation of Project GRAD Atlanta began in 2000. The first feeder pattern consisted of 12 schools: a high school, 2 middle schools, and 9 elementary schools. The selected high school was the first African-American high school in the district and has a long history of service to the African-American community. Many noted public figures are alumni, but the school had experienced substantial declines in student achievement and attendance over the past decade. The schools in the feeder pattern have been the focus of other improvement efforts, including career academies at the high school and substantial investments in technology and facilities renovation

[17]In the 2003-2004 school year, a district-developed reading program was added as a districtwide curricular component, and other methods of reading instruction are being phased out.

Table 2.4

Implementation Timetable for Project GRAD Atlanta

Core Elements of Project GRAD	Status in Atlanta, Georgia
Implementation begins	2000-2001
Classroom management: Consistency Management and Cooperative Discipline℠ (CMCD℠)	Began in the fall of 2001 for elementary and middle schools, in the fall of 2002 for high schools.
Social service support: Communities In Schools (CIS) or Campus Family Support (CFS)	Began in the fall of 2000.
First summer institute	2001
Reading: Success for All (SFA)	Began in elementary and middle schools in the fall of 2000.
Math	MOVE IT™ Math began in the fall of 2002 for grades K through 8.
First graduating class	2003-2004 150 Project GRAD scholars

for schools serving the lower grades. Observers have noted that the introduction of Project GRAD helped lay the foundation for other reform efforts and fostered community support.

In 1999, the arrival of a new superintendent in Atlanta signaled a new period of reform for the school system. The superintendent — formerly the state-appointed superintendent in Newark, New Jersey — brought previous experience with the Project GRAD initiative to Atlanta and pushed for its establishment in the lowest-performing schools in the district. Project GRAD was positioned as an important part of the district's overall reform strategy. A board of directors was formed to create Project GRAD Atlanta, and it included the superintendent, school board members, funders, and representatives from Atlanta's business community and other educational institutions. The superintendent's executive assistant briefly led the initial planning activities while the board searched for an executive director. The board selected an executive director who had extensive experience in not-for-profit management and fundraising, and — to manage implementation across the schools — it chose a deputy director who had experience as a district-level administrator and as a school principal. The Project GRAD executive director attends the superintendent's cabinet meetings, which helped position the initiative in the district.

Project GRAD Atlanta made a strategic decision to implement the initiative at an accelerated pace and to rapidly institutionalize it within the district, with the expressed intent of allowing the district to gradually take over financial responsibility and oversight. SFA began in the fall of 2000 for elementary and middle schools. CIS coordinators were hired in the fall of 2000, and CMCD℠ began in the 2001-2002 school year at the elementary and middle school level and in the fall of 2002 for the high school. CIS already existed in Atlanta but did not serve the Project GRAD schools, and it needed revamping to operate as envisioned in the model. A second feeder pattern (which is not part of this study) began planning in 2001-2002.[18] MOVE IT™ Math was implemented in the fall of 2002 — after many discussions between the district and Project GRAD Atlanta and Project GRAD USA as to which math program would best meet the needs of students and align with state standards.

Atlanta can be considered the last of Project GRAD's first-generation expansion sites, and it was able to learn from and build on the implementation experiences of the other sites. Atlanta was also able to benefit from a network of community partners, including strong corporate sponsors, prominent funders, an organized volunteer network, and local colleges and universities that provided tutors and support to the summer institutes. Also, major fundraising efforts have been successful. Yet Project GRAD Atlanta faced serious challenges. The high rate of teacher turnover in Atlanta was a recognized problem, and so Project GRAD faced a district where teacher quality and staff development were ongoing issues. In addition, the accelerated rollout of the model's components stretched the capacity of school staff to learn multiple components within a tight time period and called for a concerted effort to put in place all the school-based positions required by various components. Nevertheless, Atlanta is frequently cited by Project GRAD USA as an example of strong implementation, and even the older sites have turned to Atlanta for implementation advice. The first class of Project GRAD Atlanta scholars graduated in May 2004.

During the early years of Project GRAD's implementation, Atlanta was a "reform-rich" setting at the elementary school level. The district's comparison schools for this study implemented such programs as New American Schools, America's Choice, Co-nect, Core Knowledge, Direct Instruction, Modern Red School House, and SFA.

Key Accomplishments in Implementing Project GRAD

Although Project GRAD's implementation included the range of typical problems that occur as new programs are introduced to schools, the initiative can be credited with several important accomplishments that helped undergird the implementation of specific components:

- Across the expansion districts, Project GRAD systematically trained substantial numbers of principals and teachers to implement the initiative's compo-

[18]In the fall of 2002, Project GRAD Atlanta expanded into the Carver and South Atlanta feeder systems.

nents and, as a result, provided new skills and ongoing support related to both instructional strategies and classroom management.

- Project GRAD had to contend with the realities of the urban teaching force, often facing high turnover and many emergency or alternative certified staff who had less experience. The high teacher turnover rate resulted in extensive demand for continued staff development, to maintain school staffs that are able to fully implement the reform and to provide the needed ongoing support for inexperienced teachers.

- Through activities like the Walk for Success and outreach efforts that are part of specific Project GRAD components, the initiative engaged parents and community members in school improvement efforts in meaningful ways. For parents and key stakeholders, Project GRAD engagement strategies have helped build a constituency for school reform.

- Project GRAD helped school staff develop a focus on student achievement and data measurement as critical markers of student progress. Throughout the Project GRAD network, participating schools have placed a premium on outcome data and have emphasized the tracking of trends in achievement and other student outcomes. This was made possible to large degree by the partnerships with the school districts and by using information generated routinely by the initiative's components.

- Although the sites differ in this regard, in general, Project GRAD's feeder-pattern focus bolstered the vertical alignment of schools and facilitated new ways of communicating across school levels and from school to school.

- A recurring theme reported by principals is that the training they received through curricular components and the classroom management program helped hone their classroom observation skills and provided a needed template for instruction and management across classrooms in the same school.

- Project GRAD's scholarship offer became a galvanizing force for teachers and parents of students in lower grades, fostering a greater awareness of college-going requirements and the importance of higher education.

- In each site, Project GRAD scholars have now graduated, and each year other students sign scholarship contracts attesting to their commitment to meet the criteria. To date, at least 4,300 young people from Project GRAD schools nationwide have been able to meet the scholarship requirements and go on to college — one of Project GRAD's priority goals.

The Implementation Story's Implications for the Impact Analysis

Subsequent chapters discuss the impacts of Project GRAD on a set of key elementary student outcomes. The following observations provide a framework for that discussion.

The nature of the differential impact test varies across the study sites

The impact analysis seeks to isolate the effects of Project GRAD over and above the effects of other efforts to improve student outcomes that constituted the education program in the evaluation's sites. Thus, the analysis includes trends in student outcomes at the Project GRAD schools and trends at similar comparison schools that represent the education options that students might have received if they did not participate in Project GRAD. As discussed earlier in this report, the period of this evaluation was a time of many different efforts to improve student academic outcomes. Some occurred at the level of individual schools, as evidenced by the list of reform efforts at the comparison schools used in the impact analysis. Such school-level reform efforts do seem to constitute "business as usual" for schools during this period. In part, Project GRAD emerged as an effort to overcome the limitations of this school-based approach. In design and execution (throughout a feeder pattern, with strong local and national support), Project GRAD is intended to be distinctive and more effective than this school-level approach, creating a network of principals, teachers, and schools that are focused on a specific school improvement agenda.

A second source of pressure and support for improvement occurred at the district level, partly in response to state and (more recently) federal pressures. All the schools and districts included in the analysis faced academic challenges and pressures to improve; thus, they were interested in Project GRAD as one of a set of potential reforms. But it appears that the way that the district push for improvement played out varied somewhat across the sites and over time. As discussed in Chapter 4, it appears that, among the sites in this study, the Houston school district played the most active role in directly supporting instructional improvement across all schools in the district. During the school years included in this analysis, district-level activity that was directly related to instructional improvement was intense, but it seems to have been inconsistent in Newark and to have been less pronounced in Columbus. In Newark, many of the key district efforts in the early years of Project GRAD dealt with basic issues of administration and financial management. The circumstances in Atlanta are probably closer to those of Houston than those of Columbus and Newark. This characterization is reinforced by overall district trends in student achievement, as discussed in Chapter 4.

Thus, in each site, the comparison schools are crucial to the analysis in that they represent the likely trends in student outcomes in the absence of Project GRAD. As discussed above, the forces producing these trends vary across the sites.

In addition, the implementation of Project GRAD varied in strength across the sites. In Houston, Newark (during the period covered in this analysis), and Atlanta, implementation of Project GRAD was generally strong, acknowledging the difficulties of working in large urban school districts and low-performing schools. In Newark, although implementation was markedly stronger in the early phases, it slowed in later years as the initiative increasingly faced competing district priorities and was buffeted by political and financial issues. In Columbus, however, implementation was noticeably weaker during the school years included in this analysis.

Both the local context and the local implementation of Project GRAD can influence impacts, and this analysis suggests that the four sites represent three illustrative cases:

- Houston and Atlanta had more district support for instructional improvement and strong Project GRAD implementation.
- Newark had less district support for instructional improvement (in part because of unclear and underspecified priorities) and strong Project GRAD implementation (during the period of analysis).
- Columbus had less district support for instructional improvement and weaker Project GRAD implementation.

Results must be interpreted in light of the stage of program implementation in each site

By offering this analysis at this point in the initiative's timetable, the evaluation is depicting Project GRAD's effects partway through the effort to put the program fully into place in some of the expansion sites: The model's components were coming on line and were being strengthened, and students were working their way through the new Project GRAD education approach. Even in expansion sites where all the components are in place, they have not yet had much time to work together as a combined initiative and may not yet have developed their full momentum. This fact is important in interpreting the findings, but it does not suggest that the analysis should be delayed until the program is fully in place in all sites. Project GRAD — with its gradual rollout of components and its focus on improvements in the early grades to position students to take advantage of the college scholarships — stands in contrast to other school reforms that are pushed more quickly and more directly to tackle the educational problems in high schools. Project GRAD's approach is based on the argument that an investment of effort over time to develop students who have better skills will provide substantial payoffs in the long run. So the findings here should be seen as part of an effort to track the early stages of this theory of action and its early results.

Feeder patterns vary in the extent to which they are "self-contained," complicating the challenge that Project GRAD faces

This analysis treats Project GRAD as an intervention that operates at the level of schools and feeder patterns, and it tracks effects on the academic performance of students in the Project GRAD schools over time. The initiative's theory of action presumes that feeder patterns are sufficiently self-contained so that improvements at the elementary schools — feeding into middle and high schools — can change educational outcomes at those later schools. But high mobility among students in urban school districts can undermine this assumption. For example, among Newark's first-graders in the early 1990s who were attending schools that later became Project GRAD schools, only 10 percent to 15 percent were attending the designated Project GRAD high school in the ninth grade, and only about 50 percent to 60 percent of ninth-graders in this school had attended a "Project GRAD" school earlier. Columbus, which provides many school-choice options, presents an even more daunting challenge for the implementation of Project GRAD.

To the extent that feeder patterns are not self-contained, many students leave Project GRAD elementary schools prior to the transition to Project GRAD secondary schools, and significant numbers of students in those secondary schools did not come from Project GRAD elementary schools; in such cases, the reform faces a challenge in implementation. Acknowledging this, as a supplement to the basic focus on Project GRAD as a school and feeder-pattern reform, the analysis will make note of findings at key points for students who have remained at Project GRAD schools over an extended time. The analysis for "nonmobile" students is presented in Appendix C.[19]

[19]In general, however, narrowing the sample to nonmobile students does not change the basic conclusions of the analysis.

Chapter 3

The Evaluation's Data Sources and Analytic Approach

The typical urban school district often has a multiplicity of reforms in place. Comprehensive school reform models are, or at least have been, the most common response to student achievement challenges in urban schools and school systems.[1] Even where comprehensive school reforms are not undertaken or are not supported by the central office, school districts often undertake efforts to improve teaching and learning across their schools. Preceding chapters indicate that the urban school systems that implemented Project Graduation Really Achieves Dreams (GRAD) were no exception. During the time that Project GRAD was being implemented, most of these districts were attempting to implement some other form of systemic reform. Moreover, it was often the case that non-Project GRAD schools serving similar populations were undertaking a variety of comprehensive school reform efforts.

Therefore, the central question for Project GRAD — as is generally the case for any school reform effort — is really one of *differential* impact. In other words, the analysis does *not* ask "What is the effect of Project GRAD on student outcomes versus doing nothing?" Rather, the evaluation focuses on identifying the effect of Project GRAD versus the available alternatives in the school systems where it is being implemented. Given that there are few — if any — struggling urban schools that are not undertaking *any* type of reform, a key question for Project GRAD is whether the program appears to have had an effect *over and above whatever reforms would have been implemented in the absence of the program.* In other words, is Project GRAD more effective than other approaches in improving student outcomes? Is it less effective? Or has it about the same effect on outcomes as the mix of reforms that would have occurred without it?

This report presents estimated program effects in Houston, Texas — Project GRAD's flagship site — and in three of its early expansion sites: Newark, New Jersey; Columbus, Ohio; and Atlanta, Georgia. Within these sites, the analysis focuses on six different sets of schools, or "feeder patterns," in which Project GRAD was implemented. This chapter presents the evaluation's key outcome measures and data sources and the approach on which the analysis is based.

Key Outcome Measures and Data Sources

Project GRAD's elementary reforms emerged in large part out of the sense that the lack of academic preparation among students entering high school remained a significant impediment to improving graduation rates and success in college and that, in order to address academic

[1]See, for example, Bifulco, Duncombe, and Yinger (2005).

preparation, the program had to begin as early as possible in the educational pipeline. This suggests that, at the elementary level, the key questions regarding Project GRAD's effectiveness center on the program's effects on reading and math skills, as measured by standardized tests.[2]

In order to address such questions, the evaluation relies on the individual student records databases obtained from each of the districts in this study. MDRC collected data for every student in every school in each of the four districts in the study, from several years before the program was implemented in each district through the spring of 2003. The resulting database includes 608 schools across four districts and a total of 44 school years. This database includes information on individual students' performance on the assessments administered in each district as well as data regarding students' race/ethnicity, enrollment status, and (sometimes) their eligibility for free or reduced-price lunch and their status as an English language learner. Importantly, the database links students over time and also links them to the particular schools they attended.

Table 3.1 describes the key elementary achievement outcome measures in each district and the number of baseline and follow-up years of data available for the elementary schools in each Project GRAD feeder pattern in the study. For the majority of outcomes included in the study, the database contains at least three years of baseline data. There are a few exceptions to this rule, however, and the implications of this are discussed later in the chapter.

The Analytic Approach

The most challenging aspect of this analysis is reliably determining what would have happened in the absence of the program. In the literature of program evaluation, this is often referred to as the "counterfactual." The most reliable technique for establishing the counterfactual is the random assignment of subjects to a treatment group that has access to the program or to a control group that does not. Alternatively, as in the case of a comprehensive school reform, one may randomly assign entire schools to an experimental group that has access to the program or to a nonexperimental group that does not. In either case, because membership in these groups occurs at random, this approach can establish for certain that there are no systematic differences in measured or unmeasured characteristics of the individuals who have access to the program and of those who do not. Therefore, the control group's outcomes represent a reliable estimate of the outcomes

[2]Student attendance rates, another common educational outcome measure, are generally high at the elementary school level, and expulsions and suspensions are generally rare and strongly influenced by policy changes, such as the institution of "zero tolerance" rules. Student retention is also rare and strongly influenced by policy changes, such as the ending of "social promotion." Thus, attention is focused here on elementary student achievement as measured by performance on standardized tests. Appendix D does present findings on promotion rates.

Table 3.1

Overview of Tests Used in Project GRAD Sites

Site	Test in Place	Grades	Type of Test	Baseline Years[a]	Follow-Up Years[b]	Special Issues
Houston, TX						
Davis	Texas Assessment of Academic Skills (TAAS)	3, 4, 5	Criterion-referenced[c]	1	8	TAAS was administered until 2002.
	Stanford Achievement Test (SAT-9)	3, 4, 5	Norm-referenced[d]	1	4	SAT-9 administration started in 1997-1998.
Yates	Texas Assessment of Academic Skills (TAAS)	3, 4, 5	Criterion-referenced	3	6	TAAS was administered until 2002.
	Stanford Achievement Test (SAT-9)	3, 4, 5	Norm-referenced[d]	1	4	SAT-9 administration started in 1997-1998.
Wheatley	Texas Assessment of Academic Skills (TAAS)	3, 4, 5	Criterion-referenced	3	3	TAAS was administered until 2002.
	Stanford Achievement Test (SAT-9)	3, 4, 5	Norm-referenced[d]	1	4	SAT-9 administration started in 1997-1998.
Newark, NJ	Stanford Achievement Test (SAT-8/9)	2, 3	Norm-referenced[d]	6	2	The SAT-8 test was administered until 1996. In 1997, the new version, SAT-9, began administration.
Columbus, OH	Ohio Proficiency Test (OPT)	4	Criterion-referenced	3	4	
	Metropolitan Achievement Test (MAT)	2, 3, 5	Norm-referenced	3	4	
Atlanta, GA	Criterion-Referenced Competency Test (CRCT)	4	Criterion-referenced	1	3	

NOTES:
[a]Baseline years are the years prior to Project GRAD implementation.
[b]Follow-up years are the years following the point of implementation and forward.
[c]Criterion-referenced tests are used to measure absolute performance and assess whether students have met standards.
[d]Norm-referenced tests provide a student's scores relative to a group.

that would have been observed among treatment group individuals in the absence of the program. Any differences between the two groups can be reliably attributed to the program.

In the case of Project GRAD, the feeder-pattern design of the intervention and the fact that the evaluation started several years after the program was implemented combine to make the random assignment of students or schools impossible. An alternative approach is needed in order to generate reliable estimates of what would have happened in the absence of the program. In this analysis, the evaluation relies on several different applications of the interrupted time series (ITS) approach with comparison groups.[3] This approach uses the prior history of achievement in the program schools as the best predictor of the performance levels that would have been observed in the absence of the program. The basic logic of the approach is that — absent any intervention or change in the school's student body — the best predictor of future educational outcomes in a given school is the history of student outcomes in that same school. The approach therefore rests on two comparisons, discussed below.

> **Comparison 1 — Deviation from the Baseline Average:** The difference between baseline achievement patterns at Project GRAD schools and actual achievement in the years following program implementation, that is, the *deviation from the baseline achievement pattern*

Figure 3.1 illustrates a hypothetical example of this comparison as applied to academic achievement test scores in a fictional set of program schools. The figure plots student achievement over a three-year baseline period and through two years of follow-up after program implementation. The asterisks on the left-hand side of the graph represent average achievement scores among, for example, third-grade students across all program schools in the analysis. The solid line running through these points represents average test scores across the baseline years.

The right-hand side of Figure 3.1 illustrates average achievement patterns during the two-year follow-up period. The dashed line projects the baseline averages into the follow-up period and serves as the "benchmark" against which postimplementation outcomes can be measured. The points that are marked "D_1" and "D_2" represent hypothetical outcomes in each follow-up year. The distance between these points and the baseline mean — that is, the deviation from the baseline average — represents the first component of the impact estimate.

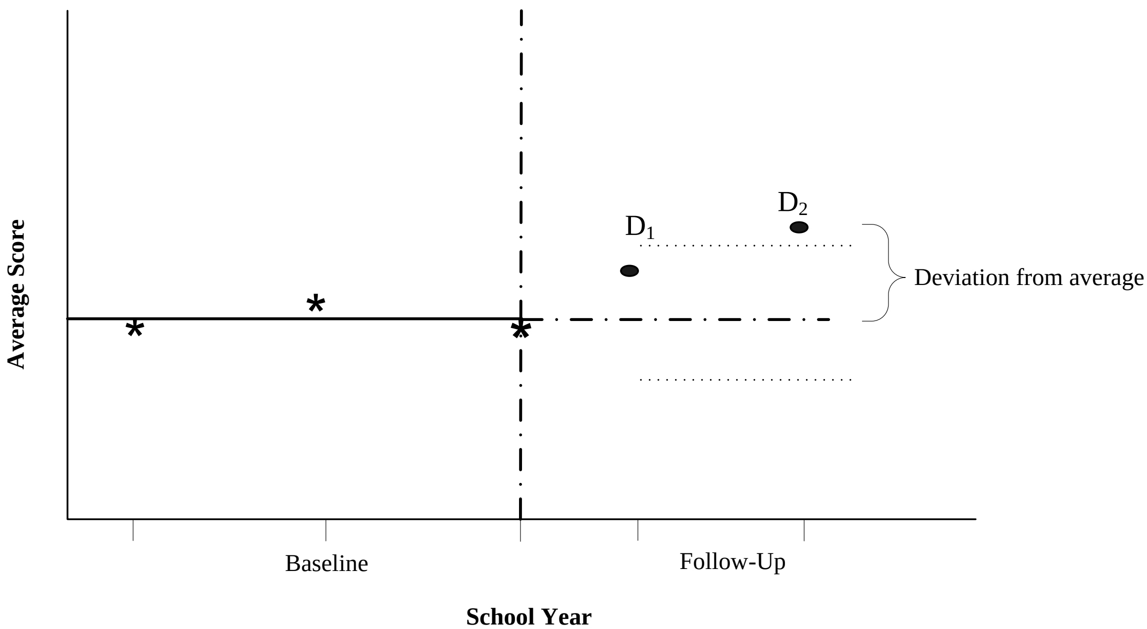
The dotted lines above and below the projected baseline represent a 95 percent "confidence interval," that is, the margin of error around the projection. To the extent that the average outcome value in any follow-up year falls outside the margin of error for the predicted trend, the estimated deviation from baseline for that year is statistically significant.

[3]Bloom (2003); Snipes (2003); Kemple, Herlihy, and Smith (2005).

The Project GRAD Evaluation

Figure 3.1

An Illustration of Interrupted Time Series Analysis: Deviation from the Baseline Average



For example, in the context of Figure 3.1, D1 indicates the actual outcome in the first follow-up year. It is somewhat higher than would be expected based on the trend before the program was implemented. However, D1 lies within the confidence interval for the predicted trend in achievement. Therefore, one cannot be sure whether this difference reflects real change or is simply a result of random fluctuation in test scores. In the second follow-up year, the difference between the average achievement in the program schools and the baseline average (indicated by "D2") is larger, and the average outcome is outside the margin of error. In this case, one can say with at least 95 percent confidence that the deviation from the baseline average in the second year of the program did not occur by chance.

"Local History." While improvement over baseline achievement patterns in a set of program schools is suggestive, it does not necessarily represent a program effect. In particular, an important threat to the validity of this basic (interrupted time series) comparison is the potential for local events other than the program to affect student achievement. Looking at the program schools alone, it is impossible to determine how much of the observed change from the baseline achievement patterns results from the program in question and how much is generated by changes in other local circumstances. If, during the same period as the particular program, major changes that are unrelated to the program are implemented (such as district-level initiatives, new state standards, or changes in curriculum), these changes could drive the observed differences, rather than the program itself. Inasmuch as reform is the rule rather than the exception in urban school districts, it seems particularly important to account for this phenomenon in the evaluation of Project GRAD.

The primary mechanism for addressing this issue is to add comparison schools from the same local context to the analysis. To the extent that the deviations from baseline patterns at Project GRAD schools can be compared with the average deviation from baseline at a set of comparison schools from the same district, one can rule out the possibility that the impact estimates are biased as a result of systemwide events that coincide with Project GRAD's implementation. In particular, instead of relying solely on the deviation from the baseline, the final impact estimates are based on the following comparison.

> **Comparison 2 — Difference in the Deviation from the Baseline Average:** The *difference* between the deviation from the baseline patterns at Project GRAD schools and the deviation from the baseline patterns at a set of carefully selected comparison schools from the same district[4]

Figure 3.2 illustrates this approach. The upper panel of the figure illustrates the average deviation from the baseline at the original hypothetical set of program schools (as shown in Figure 3.1). The lower panel illustrates the deviation from the baseline at a hypothetical set of comparison schools. In the second year of follow-up, the upper panel shows a significant deviation from the baseline average at the program schools. On the other hand, among the comparison schools, systematic differences from the baseline mean do not appear. The impact estimate would be based on the *differences* between these two patterns (for example, $D_2 - D_2^{'}$). To the extent that the deviation from baseline at the program schools is statistically distinguishable from the apparently smaller deviation from the baseline at the comparison schools, one concludes that the program had a statistically significant impact on student performance. If, on the other hand, student achievement improves relative to the baseline by similar amounts at *both* the program schools and the control schools, one cannot conclude that the program had an impact on student achievement over and above the effects of whatever reforms were present at the comparison schools or in the district as a whole. In short, the approach is to:

1. Calculate the deviation from baseline at a set of schools implementing Project GRAD (as illustrated in Figure 3.1).

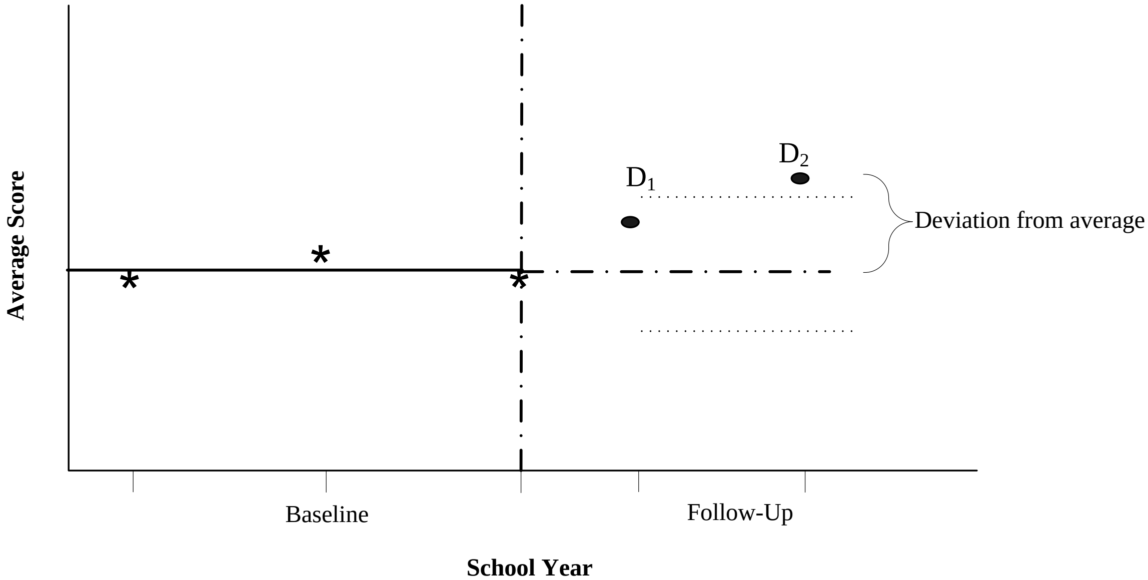
[4]For details, see Bloom (2003) and Snipes (2003).

The Project GRAD Evaluation

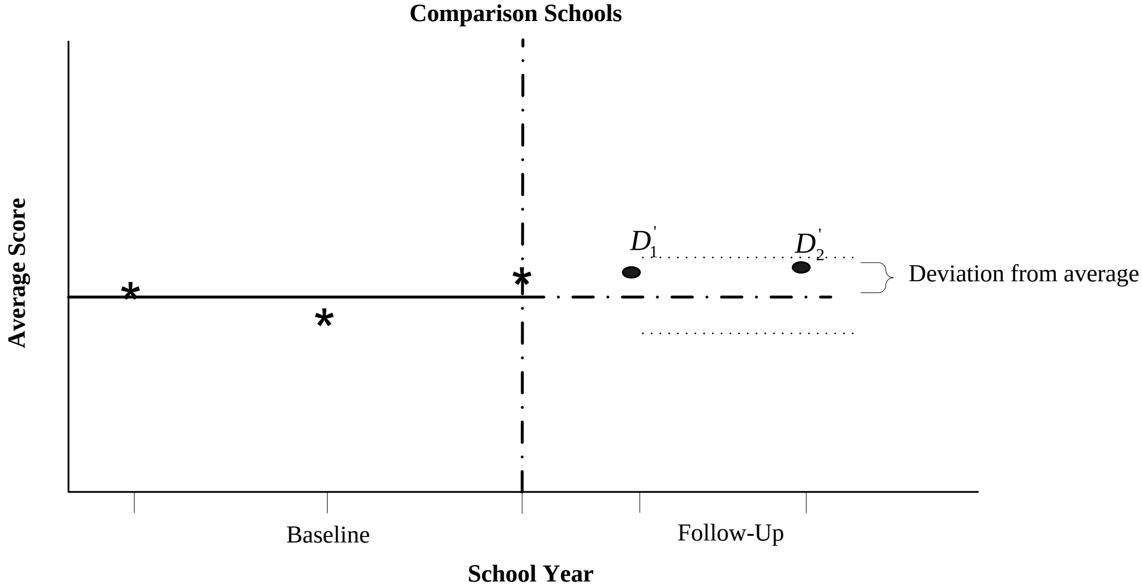
Figure 3.2

**An Illustration of Interrupted Time Series Analysis:
Difference in the Deviation from the Baseline Average**

Program Schools



Comparison Schools



2. Calculate the deviation from baseline at a set of carefully chosen comparison schools (as illustrated in the lower panel of Figure 3.2).

3. Compare the deviations from baseline at Project GRAD schools with the deviation from baseline at comparison schools from the same districts.

To the extent that the deviations from baseline at program and comparison schools systematically differ in a positive direction, one can conclude that Project GRAD had an effect on student achievement, *over and above what would have occurred in the absence of the program.* To the extent that the deviations do not differ, such a conclusion is not supported.

Choosing Comparison Schools

In effect, comparison schools are included in the analysis in order to provide an estimate of the progress that would have occurred at the program schools without Project GRAD. Therefore, the goal in choosing comparison schools was finding a set of schools from within the same district that, in the absence of any intervention, would be expected to perform similarly to the Project GRAD schools (and to be treated similarly by local stakeholders). Logic suggests that the most accurate predictor of future performance on any given outcome is usually previous performance on that same outcome. Therefore, inasmuch as this report focuses on Project GRAD's effect on elementary school achievement, prior academic achievement is the primary criterion by which comparison schools were selected.

It could be argued that, even with similar prior achievement patterns, schools that serve different student populations might be expected to evolve differently over time, particularly in response to local events or district policies. Moreover, it is important to ensure that, for purposes of face validity, one does not compare schools that simply appear to be serving different "types" of students than the Project GRAD schools, even if they exhibit similar achievement levels. Therefore, in addition to having similar prior achievement patterns, the comparison schools were also limited to those that served demographically similar student bodies.

Given these priorities, for each Project GRAD school, a set of comparison schools was selected from the same district, and these schools had to meet the following criteria:

- Average achievement in reading and math during the baseline period (typically the three years prior to Project GRAD's implementation) were each within 0.25 standard deviation of average achievement at the Project GRAD school.[5]

[5]In Project GRAD Newark, there appeared to be a trend in baseline test scores. As a result, an additional criterion was applied in selecting comparison schools: that the average change in test scores at each comparison (continued)

- The percentages of students in key racial/ethnic groups were within 20 percentage points of levels for the modal racial/ethnic groups at the Project GRAD school.[6]

These matching criteria were generally applied to the earliest grade being studied in each district. In most cases, this process resulted in a set of several comparison schools for each Project GRAD school in the study. More important, the process resulted in sets of comparison schools that, for each feeder pattern, were similar to the Project GRAD schools in terms of student demographics as well as baseline achievement. This is clear in Table 3.2, which shows the baseline achievement levels and racial/ethnic composition of the Project GRAD and comparison schools from each feeder pattern in the study.[7]

Accounting for Pre-Program Trends in Achievement

This chapter presents an interrupted time series model with comparison groups based on the difference between postprogram achievement and *average baseline achievement.* For most of the feeder patterns in this evaluation, there was little evidence of either positive or negative trends in student achievement prior to program implementation. Therefore, this approach seems appropriate. If there are meaningful positive or negative pre-program trends in student performance, however, it is important to account for these in the analysis. Otherwise, one cannot distinguish between a change in patterns of achievement and continuation of the growth or decline in student outcomes from the years prior to implementation. Where trends in baseline achievement exist, the interrupted time series model can be adapted to compare the difference between postprogram achievement levels and the baseline *trend.* In this case, postprogram achievement is compared with the levels of achievement based on projecting the pre-program trend into the follow-up period. In five of the six feeder patterns in this study, this was considered unnecessary. However, this "deviation from trend" model was applied in Newark, where, during the six years before program implementation, achievement scores at the set of schools that were eventually chosen for Project GRAD substantially declined. As such, rather than being based on the deviation from the baseline mean, the program effects in the Newark site are based on a deviation from baseline trends in achievement.[8]

---

school during the six years prior to Project GRAD's implementation had to be in the same direction as the program school with which it was being compared.

[6]For example, if the majority of students at the Project GRAD school were black, then comparison schools were limited to those at which the percentage of black students was within 20 percentage points of the percentage of black students at the Project GRAD school.

[7]As further evidence of the similarity of the Project GRAD and comparison schools, Appendix B presents a full description of student characteristics and baseline achievement levels at Project GRAD and comparison schools for each feeder pattern.

[8]This analysis is discussed in detail in Snipes (2003).

The Project GRAD Evaluation

Table 3.2

Baseline Averages and Demographic Characteristics of Project GRAD and Comparison Schools, by Feeder Pattern

Feeder Pattern	Project GRAD Schools	Comparison Schools	All Project GRAD and Comparison Schools in District
Houston, Davis			
3rd-grade math TAAS[a]	60.2	62.2	61.7
3rd-grade reading TAAS	70.6	70.1	70.3
Race/ethnicity			
Black	9.5	12.5	11.7
White	1.9	5.3	4.4
Hispanic	88.4	80.4	82.6
Houston, Yates			
3rd-grade math TAAS	66.1	67.4	67.1
3rd-grade reading TAAS	72.0	73.0	72.8
Race/ethnicity			
Black	85.7	80.1	81.6
White	1.8	2.2	2.1
Hispanic	11.7	16.1	14.9
Houston, Wheatley			
3rd-grade math TAAS	72.3	73.4	73.2
3rd-grade reading TAAS	77.2	78.3	78.2
Race/ethnicity			
Black	45.0	40.5	41.1
White	1.3	6.3	5.7
Hispanic	53.1	51.0	51.3
Newark			
3rd-grade math SAT[b]	43.6	42.2	48.7
3rd-grade reading SAT	37.9	37.4	43.3
Race/ethnicity			
Black	87.7	94.0	69.6
White	0.4	0.0	6.7
Hispanic	12.0	5.8	22.3
Columbus			
3rd-grade math MAT[c]	33.6	34.8	42.8
3rd-grade reading MAT	37.5	37.4	44.5
Race/ethnicity			
Black	85.6	83.1	53.4
White	10.7	13.3	41.2
Hispanic	0.7	0.8	1.1

(continued)

Table 3.2 (continued)

Feeder Pattern	Project GRAD Schools	Comparison Schools	All Project GRAD and Comparison Schools in District
<u>Atlanta</u>			
4th-grade math CRCT[d]	295.3	290.7	296.3
4th-grade reading CRCT	298.0	293.0	300.1
Race/ethnicity			
Black	97.7	98.1	87.9
White	1.5	0.5	7.5
Hispanic	0.6	1.2	3.1

SOURCES: MDRC calculations from individual student school records from the Houston Independent School District, Newark Public Schools, Columbus Public Schools, and Atlanta Public Schools.

NOTES:

[a]TAAS stands for Texas Assessment of Academic Skills. The average presented in this table is a Texas Learning Index (TLI) score; see Chapter 4.

[b]SAT stands for Stanford Achievement Test, versions 8 and 9. The average presented in this table is a percentile score.

[c]MAT stands for Metropolitan Achievement Test. The average presented in this table is a percentile score.

[d]CRCT stands for Criterion-Referenced Competency Test. The average presented in this table is a scale-score point.

Key Threats to Validity

By adding carefully chosen comparison schools from the same district, the methodology of using interrupted time series with comparison groups effectively eliminates the possibility that districtwide events might undermine the validity of impact estimates. However, several other threats to validity must be taken into account.

Test Changes and Availability of Baseline Data

In general, most of the comparative interrupted time series analysis in this report relies on at least three years of baseline data. That somewhat increases the confidence level that the analysis can ascertain whether or not there is a baseline trend to be accounted for and that it can develop a reasonable estimate of the baseline mean. As Table 3.1 illustrates, however, there are a couple of feeder patterns for which three years of baseline data are not available. In Houston, for example, the Davis High School feeder pattern has only one year of state data for the Texas

Assessment of Academic Skills (TAAS) prior to the implementation of Project GRAD. Ideally, one would prefer a longer trajectory of baseline data in order to ascertain whether a baseline average or a baseline trend model is more appropriate and to be certain that the baseline average is an accurate representation of school performance. The lack of data, however, makes this impossible. Yet the fact that there are multiple program and comparison schools in the analysis increases confidence in the baseline mean and reduces the likelihood that the findings are driven by random fluctuations in the data.

In addition to the state test, the Houston Independent School District (HISD) began annual administrations of the Stanford Achievement Test (SAT-9) to elementary school students throughout the district in 1998 — several years after the implementation of Project GRAD. The SAT measures students' performance relative to a nationally representative sample of students. Given the systematic gains on the TAAS during the period of this study, the focus on performance on the TAAS throughout the local school system and entire state, and the high performance levels observed in HISD near the end of the follow-up period, estimates of Project GRAD's impact on an alternative assessment (the SAT-9) may shed useful light on the effects of the program. In particular, the estimates might address questions regarding the presence of ceiling effects — that is, an inability to measure Project GRAD's impacts on achievement due to high scores on the TAAS throughout the school system. The estimates might also address the possibility of tradeoffs between progress on local assessments and progress on broader measures of student achievement, as well as the interaction of Project GRAD's reforms with this dynamic.

The fact that the test changes occurred after Project GRAD's initial implementation deprives the analysis of a pre-program baseline. As such, the analysis uses 1999 as a "pseudo-baseline" and estimates the effects of Project GRAD on SAT-9 improvements since 1999 in the three feeder patterns for this analysis.[9] This approach has two primary limitations: (1) these estimates would not reflect any impacts on skills measured by the SAT-9 occurring prior to 1999, and (2) the lack of multiple years of baseline data reduces the precision with which the pseudo-baseline is estimated.[10]

In Atlanta, the state test (the Criterion-Referenced Competency Test, or CRCT) was administered to all *fourth*-grade students in each year of the follow-up period. Unfortunately, the first year of test administration was only one year prior to program implementation. As such, it is difficult to be confident in this as a measure of baseline achievement. However, ab-

[9]In order to minimize the potential distortion associated with the initial year of administration of a new test, the 1998 data were not used in this analysis.

[10]For more on how the properties of interrupted time series approaches vary depending on data availability, see Quint, Bloom, Rebeck Black, and Stephens (2005).

sent a genuine treatment effect, there is no reason to think that this phenomenon would affect program and comparison schools differently.

Compositional Shifts

Another potential threat to the validity of this analytic approach is the possibility of systematic changes over time in the student populations served by the Project GRAD schools. Whatever their cause, such "compositional shifts" can generate changes in student performance that are unrelated to the program and that make it difficult to distinguish between program effects and changes in student performance that would have occurred whether or not the program had been in place. For example, suppose that Project GRAD schools succeed at retaining more low-performing students than they would have in the absence of the program. Without some way to correct for this phenomenon, it would drive down average achievement and would result in an underestimate of the effect of Project GRAD. On the other hand, suppose that well-informed parents — on hearing about the new program — made a concerted effort to enroll their children in the Project GRAD schools. This scenario could result in a shift in the average levels of academic readiness and parental support among the students who are typically served by Project GRAD schools and, consequently, could overestimate the effects of Project GRAD on student outcomes.

This issue is addressed by using individual students' background characteristics and prior achievement levels to control for potential changes in students' characteristics. As such, the estimates of annual achievement levels sometimes differ from the unadjusted numbers often reported by districts, states, and other stakeholders. Specifically, estimated program effects in this report control for such characteristics as race/ethnicity, eligibility for free or reduced-price lunch, and prior academic performance. Prior academic performance is measured two ways. An indicator variable based on each student's age controls for whether he or she appears to have repeated a grade in the past. When possible, estimates also adjust for students' achievement test scores in the prior grade. For example, when evaluating trends in third-grade achievement, the analysis adjusts for second-grade achievement data, if available. This "value-added" approach is analogous to evaluating the trends in achievement *gains* between second and third grade, and it controls for any measurable shifts in student composition that occur throughout the life of the program.

This approach affects the interpretation of the estimated program effects in the analysis. In each site, the estimated effects for the earliest grade examined — where prior achievement data are not available — reflect the *cumulative* program effects across all the lower grades. The program estimates for subsequent grades represent the incremental effects, or the added value of the program over and above the effects in prior grades. Together, these estimates provide a complete picture of program effects across the grades for which achievement data are available.

It is possible for cumulative effects through the highest elementary grade to exist but not to be reflected in these estimates. In order to explore this possibility, program effects were estimated with and without controlling for background characteristics and student performance. This did not have any effect on the overall pattern of estimates. In addition, subsets of students in the higher elementary grade were identified who had been at Project GRAD schools for their entire elementary school experiences. Estimated program effects among these students were also similar to those in the incremental estimates presented in the body of this report.

Selection Bias

In general, the interrupted time series approach — even with comparison schools — is vulnerable to biases generated by the process of selecting schools into the treatment sample. Because the selection of schools for Project GRAD is not random, the process whereby schools are selected into the program group can have important implications for treatment effects. To the extent that schools that are more or less likely to improve over time are systematically more or less likely to receive the treatment, comparative interrupted time series analysis could yield biased estimates of program effects.

If program participation is completely voluntary, it can be argued that — even among schools that are similar in terms of average baseline achievement and demographics — schools that would be systematically more or less likely to improve *even without the intervention in question* might be more likely to volunteer. Schools that have energetic and effective leaders, for instance, might be more likely to seek out new interventions and to build the staff consensus needed to adopt a new program. One might expect that such leaders could generate improvements in student achievement even in the absence of a genuine program effect. If program participation is not voluntary but is assigned by the district or other local stakeholders, selection can still be done in ways that could bias the analysis. Districts may make a concerted effort to ensure that reforms target schools that have sufficient capacity for successful implementation and progress. This could also result in overestimates of program effects. On the other hand, districts may make a concerted effort to direct these resources to schools that are in the most need of support. This may result in program schools that — even given similar student populations and average levels of achievement — have less capacity and are less likely to improve in the absence of an intervention. This could generate underestimates of program effects.

The fundamental problem is that — as long as program participation is not random — such phenomena cannot be distinguished with absolute certainty from genuine program effects. In the case of Project GRAD, however, elementary schools are selected on the basis of their membership in feeder patterns that send students to Project GRAD middle schools and high schools. Although the assignment of schools is not random, program developers and district personnel are restricted in their choices of elementary schools to those that feed into the Project

GRAD high schools being served by the program. Even if high schools and elementary schools were selected simultaneously, the presence of six to eleven elementary schools per feeder pattern suggests that stakeholders would not be able to freely choose elementary schools that are systematically more or less likely to improve on their own. Rather, if district personnel and program developers choose a high school, they are restricted to the set of elementary schools that are already associated with that high school. So while selection is not random, the process does *not* suggest the presence of any systematic selection issues for elementary school analysis.

Chapters 4 and 5 present the results of applying the analytic approach outlined in this chapter to estimating Project GRAD's effects in six different feeder patterns across four school districts.

Chapter 4

The Effects of Project GRAD on Elementary Student Achievement in Houston

A core concept underlying the model for Project Graduation Really Achieves Dreams (GRAD) is that progress in achievement at the elementary level is essential if students are to succeed in high school and college. Therefore, understanding whether the program improves elementary-level achievement is a fundamental question underlying the efficacy of Project GRAD. In order to address this question, the analysis presented in Chapters 4 and 5 focuses on whether Project GRAD improved students' achievement in elementary schools *over and above what would have been observed in the absence of the intervention.* In particular, changes in achievement at the Project GRAD schools are compared with changes in achievement at the set of carefully chosen comparison schools. This chapter focuses on the results in Houston, Texas — Project GRAD's flagship site. Chapter 5 presents estimated impacts in three of the early expansion sites: Newark, New Jersey; Columbus, Ohio; and Atlanta, Georgia. Chapter 6 discusses the implications of these results.

This chapter focuses on two measures of student achievement: (1) the Texas Assessment of Academic Skills (TAAS), a state test that was the focus of great state and local attention; and (2) the Stanford Achievement Test (SAT-9), a national test that is designed to measure a broader range of academic skills, is not sensitive to specific instructional approaches, and allows comparisons of Houston's students and the national pool of test-takers. Having these two different measures of student achievement is important because of the intense state and local focus on the TAAS and its reputation as a relatively easy test and because of concerns that emphasis on TAAS scores might undermine progress on broader measures of student achievement. Moreover, high overall performance levels on the TAAS that were observed near the end of the follow-up period could theoretically obscure program effects, should they exist. As such, estimated impacts on an alternative assessment could shed useful light on discussion of the effects of Project GRAD on elementary school achievement. The analysis presented in this chapter suggests several key findings:

- **In general, achievement on the Texas state standards-based assessments (the TAAS) at Project GRAD Houston schools improved substantially during the years following Project GRAD's implementation.**

- **Comparison schools throughout the district made similar progress on these tests, suggesting that Project GRAD did not improve these outcomes beyond what would have happened without the program.**

- **In the years following Project GRAD's implementation, student achievement on the national test (the SAT-9) declined at the district's comparison schools, which were similar to the schools implementing Project GRAD.**

- **At the same time, SAT-9 achievement at Project GRAD Houston's elementary schools generally remained stable. The differences suggest that — as measured by nationally "norm-referenced" tests — Project GRAD had a positive effect on elementary student achievement in Houston.**

As mentioned earlier, Project GRAD was implemented in several different feeder patterns in the Houston Independent School District (HISD). Specifically, Project GRAD began implementing the elementary components of the program in the Jefferson Davis High School feeder pattern in the 1993-1994[1] school year, in the Yates High School feeder pattern in the 1996-1997 school year, and in the Wheatley High School feeder pattern in the 1999-2000[2] school year. The following sections present the results of the analysis for each of these feeder patterns, beginning with the Davis feeder pattern, where Project GRAD has been in place the longest.

The Texas Assessment of Academic Skills (TAAS)

TAAS Results in the Davis Feeder Pattern

- **Elementary TAAS scores increased substantially after Project GRAD was implemented in the Davis feeder pattern.**

- **In most years, TAAS increases at Project GRAD schools were matched by similar increases at comparison schools from the district.**

[1]Given that not all elementary schools in the Davis feeder pattern implemented CMCD[SM] in the 1993-1994 school year, this analysis considers 1994-1995 as the first year of Project GRAD.

[2]Although scholarships were announced in 1997-1998, the first program components did not begin at Wheatley until 1999-2000. Project GRAD was also implemented in two other feeder patterns in Houston in subsequent years. However, due to the limited time elapsed for follow-up analysis, these schools are not included in this analysis.

The TAAS was administered to elementary school students in Houston (and in the rest of Texas) annually from the spring of 1992 through the spring of 2002.[3] The available data track the effects of Project GRAD in the Davis feeder pattern through the first eight years of implementation.

The primary TAAS outcome considered here is a score referred to as the "Texas Learning Index" (TLI).[4] Figure 4.1 illustrates the estimated program effects on TLI math scores across the Davis feeder pattern for grades 3 through 5. (Figures 4.1 to 4.6 are found on pages 59 to 70.) The shaded bars represent the average *change or deviation* from the baseline average in the TLI score at Project GRAD schools during each of eight years of implementation, controlling for any changes in students' demographic characteristics. The unshaded bars represent the average deviation from baseline in the TLI scores at the comparison schools over the same school years, also controlling for any changes in students' demographic characteristics.

The numbers above the bars represent the *difference* between the changes from the baseline average at the Project GRAD schools and the changes from the baseline average at the comparison schools in each follow-up year. This is the estimated program effect for each year. It represents the estimated impact of Project GRAD *over and above what would have been observed in the absence of the program.*

The estimates presented in Figure 4.1 show that, over the course of the follow-up period, average TLI math scores at the Project GRAD schools increased relative to the baseline average. However, scores also increased at the comparison schools. The results for the two groups of schools suggest that progress at Project GRAD schools exceeded progress at the comparison schools in some grades and years. Examining the follow-up period as a whole, however, shows that math test scores improved substantially for both sets of schools. The patterns do not generally suggest sustained positive effects for Project GRAD, *over and above the progress that occurred in similar schools throughout the district.*

---

[3]In 2003, the State of Texas began a new test, the Texas Assessment of Knowledge and Skills (TAKS). Although MDRC has obtained TAKS data for 2003, given the change in tests and test metrics, those data are not included in the analysis.

[4]The TLI is a continuous score that describes a student's performance on the TAAS. The TLI is provided for both the TAAS reading and mathematics tests. The raw score on the TAAS is simply the number answered correctly on the test, and since this raw score can be interpreted only in relation to the number of items on the test, the score is limited in use. The TLI makes it possible to relate student performance to a passing standard and to compare student performance from year to year. In each year, the raw scores are standardized into TLI scores relative to the state's passing standard of 70. For TLI frequency distributions for each grade and subject, see Texas Education Center (2005).

For example, the top panel of Figure 4.1 shows math findings for third-graders. In the baseline period, TLI math scores among third-graders at Project GRAD schools averaged 60.1 points. (Baseline averages for the two sets of schools are shown below the bars in each panel.) In the first year of the program, the average TLI math score for third-graders at the Project GRAD schools improved by 5.2 points (represented by the height of the shaded bar), to 65.3 points. Over the same follow-up years, the average TLI math score at the comparison schools fell from 63.1 points in the baseline period to 62.7 points in the first year of the program, a difference of -0.4 point. In other words, TLI math scores for third-grade students at Project GRAD schools improved by an estimated 5.7 points *more than* did the scores at the comparison schools. The asterisk on this number indicates that this difference is statistically significant; that is, it is unlikely to have occurred by chance.[5] This suggests that Project GRAD had a statistically significant impact of 5.7 points during its first year of implementation in the Davis feeder pattern.

As Figure 4.1 shows, however, this impact did not persist over time. Third-grade TLI math scores on the TAAS generally improved relative to the baseline average over the eight years after implementation, and, in most years, progress at Project GRAD and comparison schools was similar. In particular, it appears that — for the most part — the growth at the comparison schools during the eight-year follow-up is statistically indistinguishable from the growth at the program schools, as indicated by the lack of asterisks on most of the impact estimates. Interestingly, in the eighth year of the program, the growth in third-grade math test scores at Project GRAD schools again outpaced the growth at the comparison schools, suggesting a net impact of 8.7 points on the TLI. Given that this result appears only in the last year of follow-up, it is difficult to know what to make of it.

A similar pattern of growth in TLI math scores occurred in the fourth and fifth grades, as shown in the middle and bottom panels of Figure 4.1. Compared with the third grade, however, in these grades, there were even fewer systematic positive differences between the growth at the Project GRAD schools relative to the comparison schools.[6] Across the grades and during the follow-up period as a whole, the patterns do not generally suggest sustained positive effects

---

[5]In order to conserve space, the report uses exhibits like Figure 4.1 to report most of the impact estimates. Appendix A presents the estimated impacts in tabular format for all grades and sites in the evaluation. These impacts are translated into an "effect size" metric — that is, the estimated effect divided by the standard deviation of the outcome in question.

[6]It is worth noting that the program estimates for fourth and fifth grade control for achievement in the prior school year. As a result, they represent the net effect of Project GRAD over and above any effects in previous grades. Appendix C examines program effects without these controls and does not find substantial differences between the results.

on TAAS math scores for Project GRAD, *over and above the progress that occurred at comparison schools throughout the district.* In other words, there is little evidence that Project GRAD schools achieved any progress that would not have occurred without the program.

Figure 4.2 illustrates estimated effects on TLI reading scores on the TAAS at the elementary schools in the Davis feeder pattern, and it suggests a similar pattern. Average TLI reading scores among third- and fourth-graders at Project GRAD schools appear to have improved over the course of the follow-up period. Again, these improvements are generally matched by improvements at the comparison schools, suggesting no systematic effects from Project GRAD. Interestingly, as shown in the bottom panel of Figure 4.2, fifth-grade reading scores do not show a consistent pattern of improvement over time at either set of schools. While there are estimated positive effects for fifth-graders in several follow-up years, there are negative effects in at least one year. Given the manner in which the data fluctuate from year to year, it is hard to draw conclusions regarding progress in fifth-grade reading achievement in the Davis feeder pattern. One possibility is that Project GRAD began to generate effects on reading scores in the third and fourth years of the intervention but that, for some reason, neither test scores nor program effects were consistent after that point.

TAAS Results in the Yates and the Wheatley Feeder Patterns

- **As in the Davis feeder pattern, elementary school TAAS scores within the Yates and the Wheatley feeder patterns rose in the years following Project GRAD's implementation.**

- **Also as in the Davis feeder pattern, the progress at the Project GRAD schools in the Yates and the Wheatley feeder patterns was matched by progress at comparison schools throughout the district.**

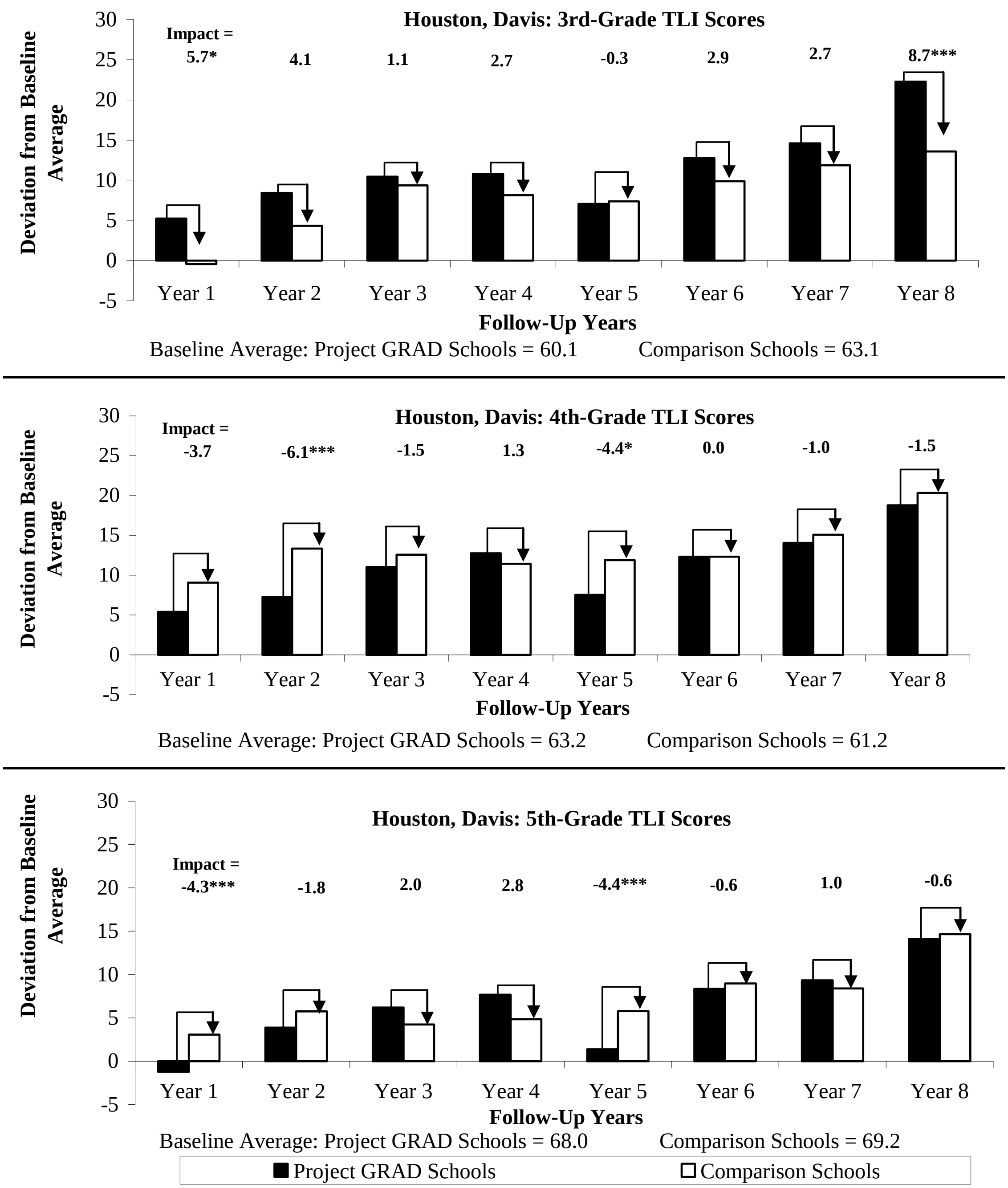
Project GRAD began implementation at the elementary schools feeding into Yates High School in the 1996-1997 school year, allowing for six years of follow-up through the spring of 2002. Figures 4.3 and 4.4 present estimated impacts on average TLI scores for math and reading, respectively, in the Yates feeder pattern. Though there are some fluctuations, Figure 4.3 shows that math scores at Project GRAD elementary schools generally improved after program implementation. It also shows that the test scores at the comparison schools experienced similar improvements during the follow-up period. While the impact estimates fluctuate from positive to negative, there seem to be no sustained effects by Project GRAD on students' math scores over and above the progress at the comparison schools.

Similarly, Figure 4.4 shows that Project GRAD schools experienced improved reading test scores over the six-year follow-up period. In most grades and years, this growth is statistically indistinguishable from improvements at the comparison schools. However, there do appear to be statistically significant negative effects on fourth-grade reading achievement in the first few years of implementation. While the Project GRAD schools appear to have "caught up" in the latter years of the program, this pattern should at least raise questions about the effectiveness of Project GRAD with respect to fourth-grade reading in the Yates feeder pattern. In the fifth grade, there appears to have been less improvement overall and no systematic pattern of difference at the Project GRAD schools versus the comparison schools.

Project GRAD began implementation in the elementary schools that feed students into Wheatley High School in the 1999-2000 school year. Figures 4.5 and 4.6 present findings for elementary achievement at these schools through 2002 — three years of program implementation. The estimates show positive effects for Project GRAD on fourth- and fifth-grade math achievement in the first year of implementation. However, these estimated impacts diminish in the second and third years of the program. For reading scores, the improvements at the Project GRAD schools exceeded improvements at the comparison schools only for fifth-graders in the third year of implementation. Otherwise, changes at the Project GRAD schools and the comparison schools are similar.

Figure 4.1

Impact Estimates Across All Project GRAD Elementary Schools, Davis Feeder Pattern, Math Scores, by Grade



(continued)

Figure 4.1 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

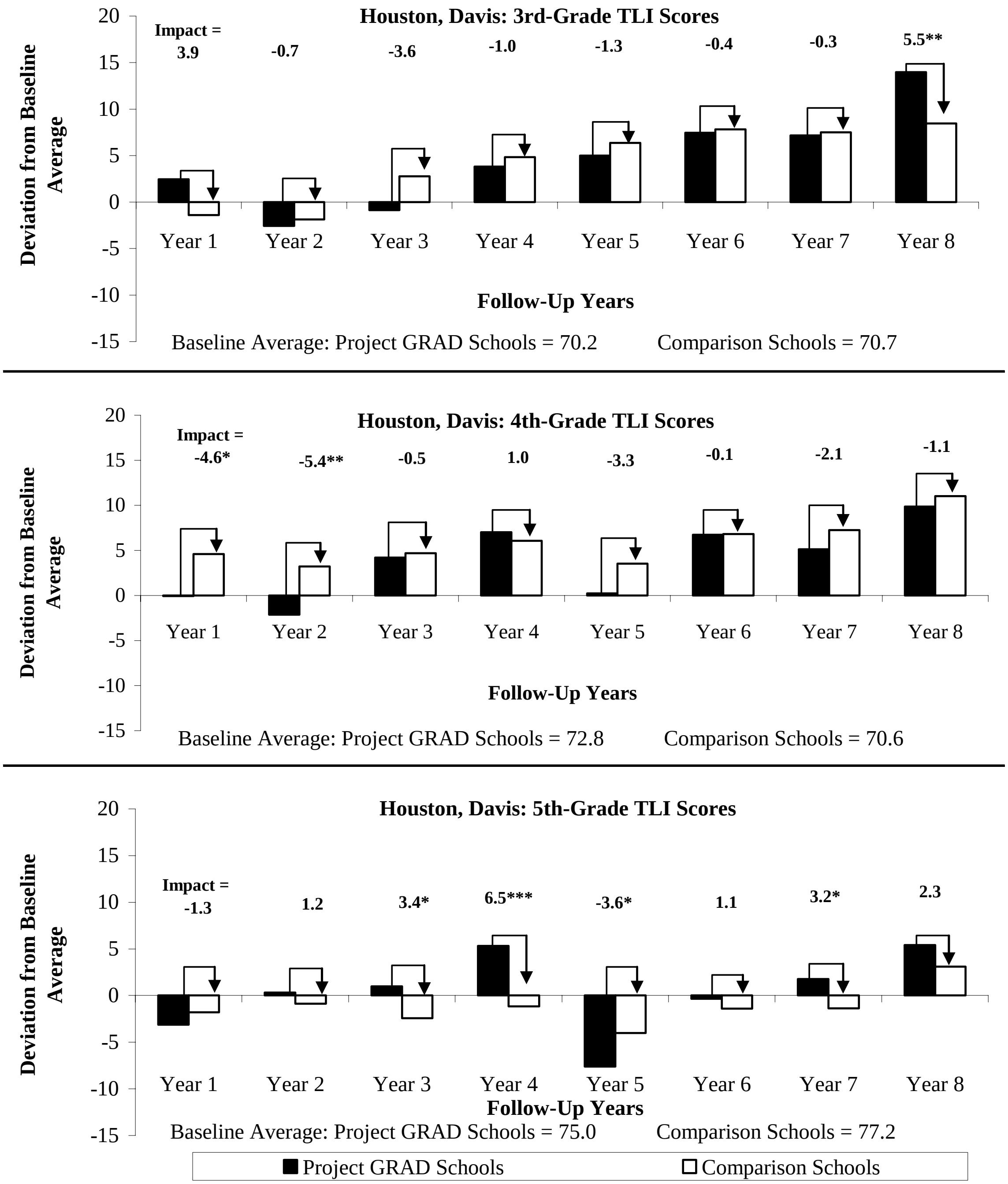
A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 8 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

The Project GRAD Evaluation

Figure 4.2

Impact Estimates Across All Project GRAD Elementary Schools, Davis Feeder Pattern, Reading Scores, by Grade



(continued)

Figure 4.2 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

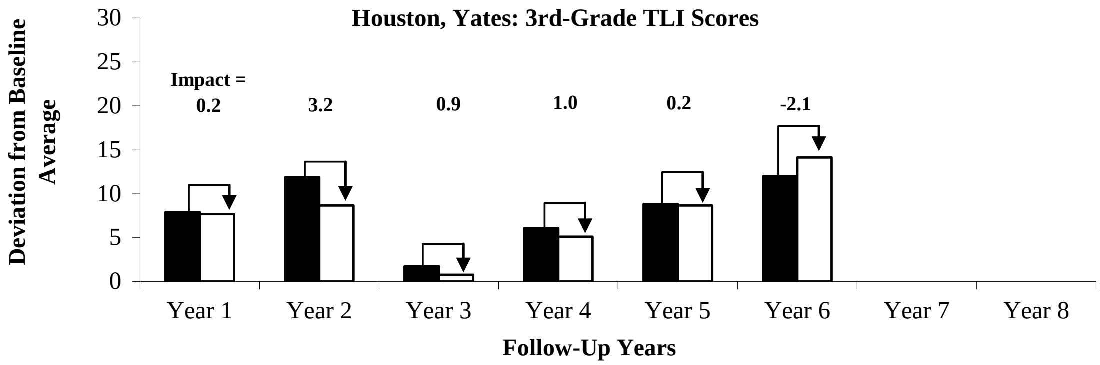
A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 8 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

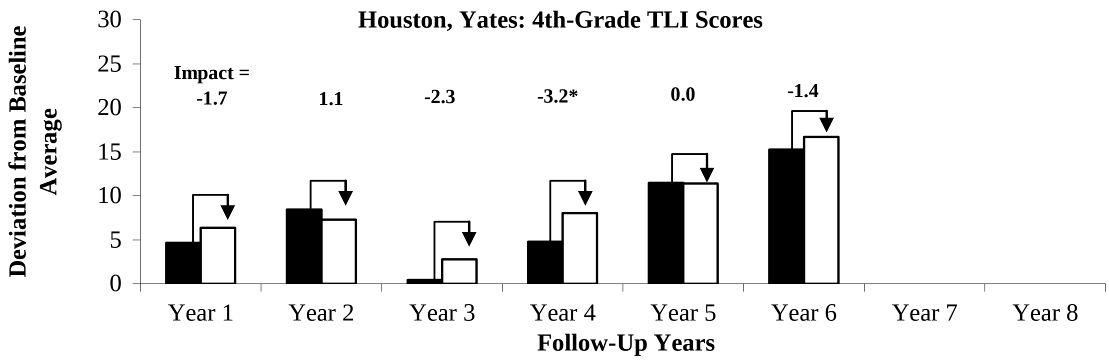
The Project GRAD Evaluation

Figure 4.3

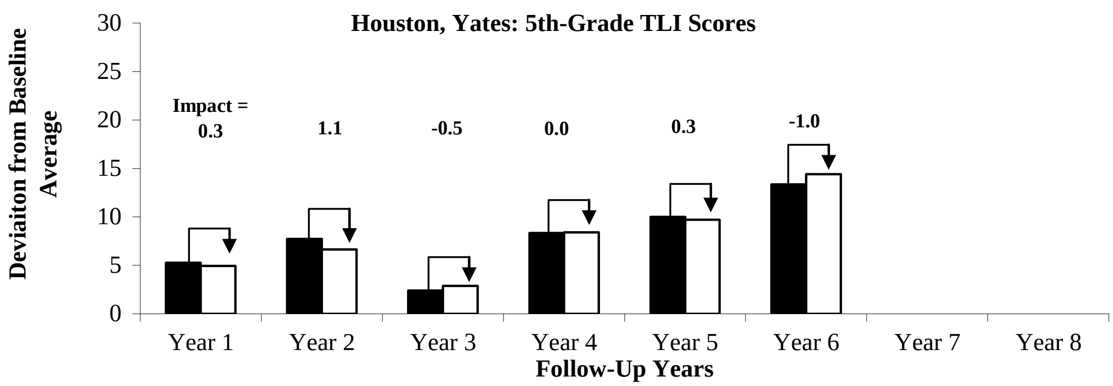
Impact Estimates Across All Project GRAD Elementary Schools, Yates Feeder Pattern, Math Scores, by Grade



Baseline Average: Project GRAD Schools = 66.2 Comparison Schools = 66.5



Baseline Average: Project GRAD Schools = 68.2 Comparison Schools = 67.3



Baseline Average: Project GRAD Schools = 69.4 Comparison Schools = 69.2

■ Project GRAD Schools □ Comparison Schools

(continued)

Figure 4.3 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

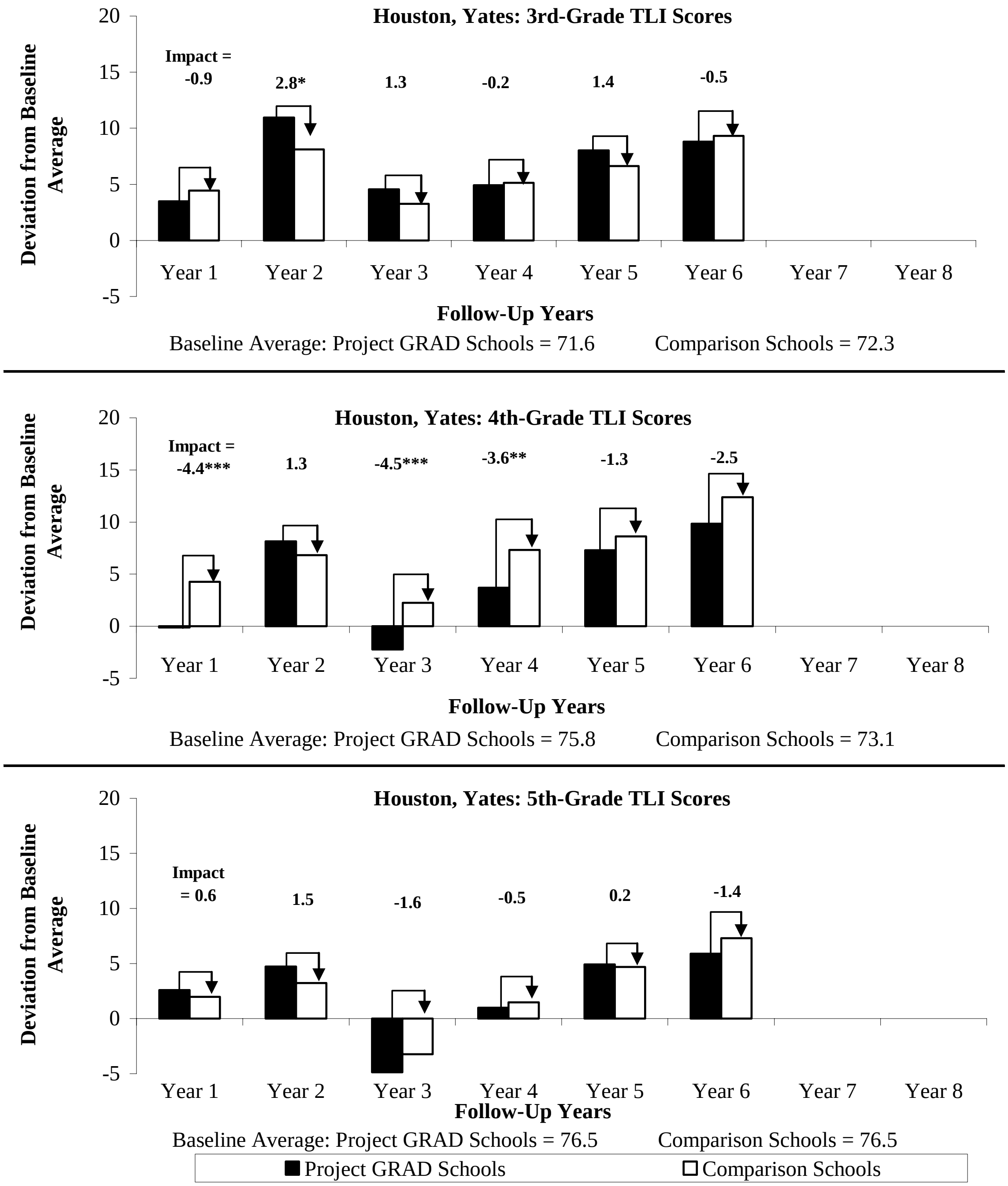
A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 4 and 15 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

The Project GRAD Evaluation

Figure 4.4

Impact Estimates Across All Project GRAD Elementary Schools, Yates Feeder Pattern, Reading Scores, by Grade



(continued)

Figure 4.4 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

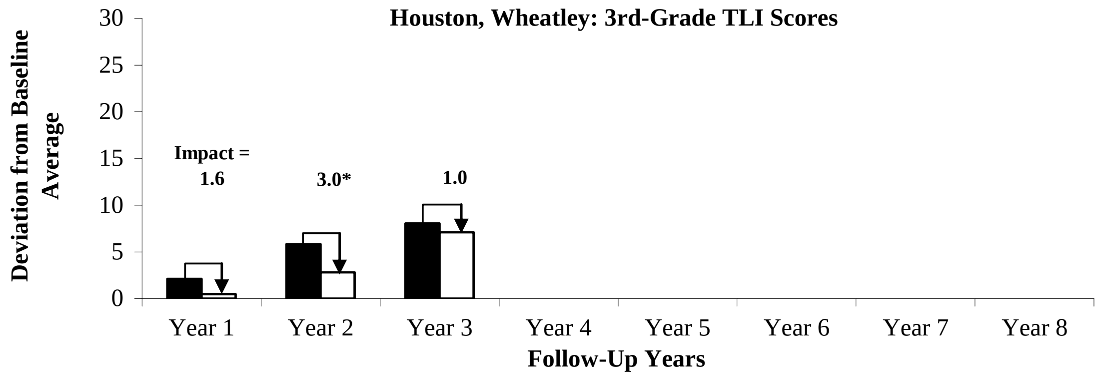
The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

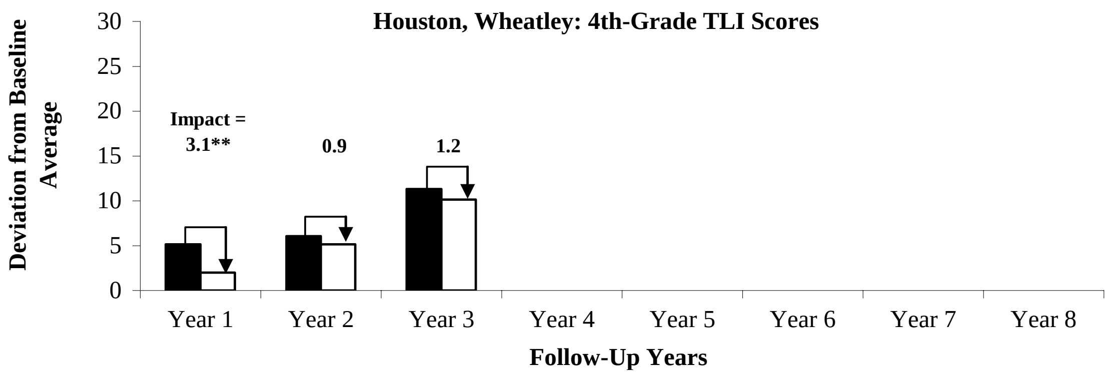
Clusters consist of a Project GRAD school matched with a group of between 4 and 15 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

Figure 4.5

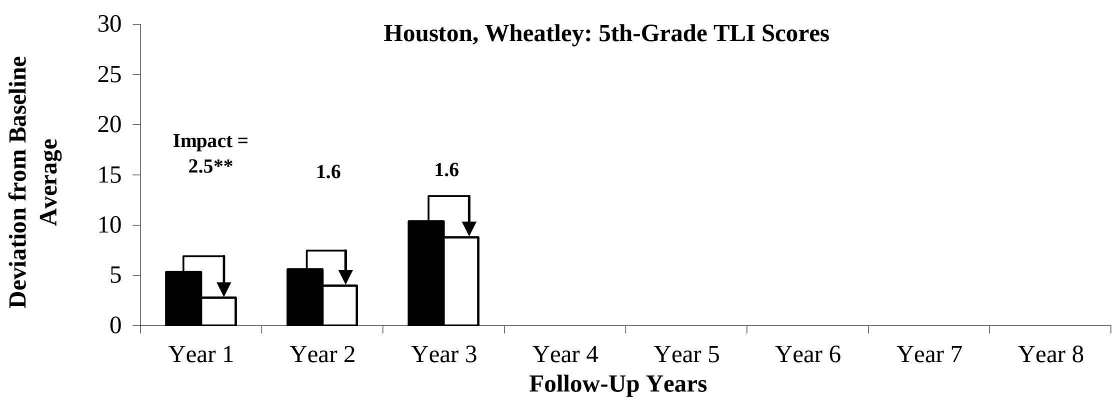
Impact Estimates Across All Project GRAD Elementary Schools, Wheatley Feeder Pattern, Math Scores, by Grade



Baseline Average: Project GRAD Schools = 72.1 Comparison Schools = 72.5



Baseline Average: Project GRAD Schools = 72.9 Comparison Schools = 73.8



Baseline Average: Project GRAD Schools = 76.0 Comparison Schools = 77.1

■ Project GRAD Schools □ Comparison Schools

(continued)

Figure 4.5 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

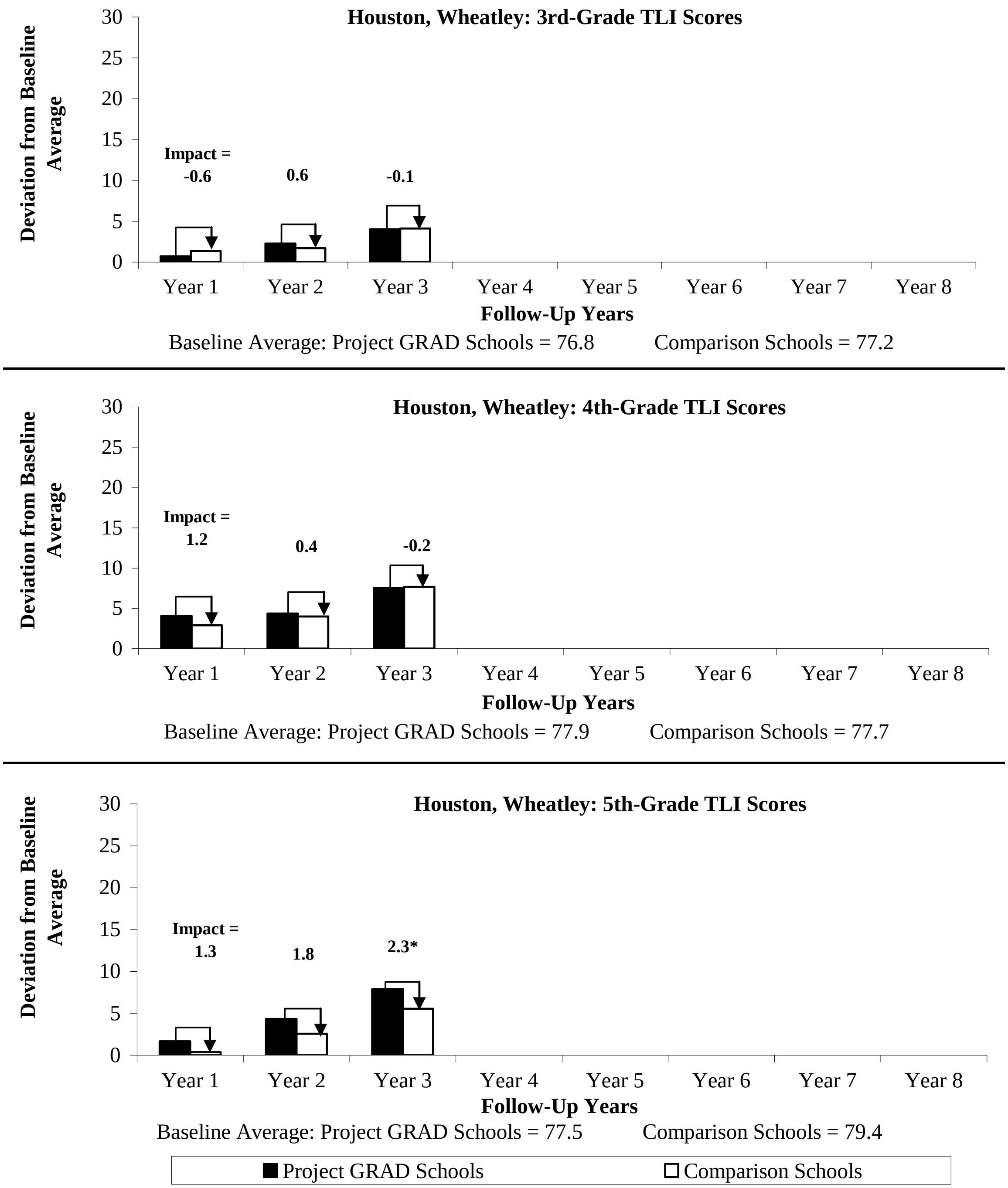
A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 2 and 19 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

The Project GRAD Evaluation

Figure 4.6

Impact Estimates Across All Project GRAD Elementary Schools, Wheatley Feeder Pattern, Reading Scores, by Grade



(continued)

Figure 4.6 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 2 and 19 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

The Stanford Achievement Test (SAT-9)

In the spring of 1998, HISD began administering the national Stanford Achievement Test (SAT-9) to students at various grade levels within the school system. Previous research has shown that student achievement patterns on state tests that are designed to measure student performance relative to a state standard often differ from achievement patterns on nationally norm-referenced achievement tests.[7] Moreover, given the strong state and local emphasis on improving TAAS scores and the resulting "rising tide" of TAAS scores in HISD as well as in Texas as a whole, it is possible for Project GRAD to have generated achievement impacts that would have been obscured by "ceiling effects" on the TAAS.[8] Therefore, supplementing the TAAS analysis with a second measure of student achievement that is not subject to these pressures and problems is an important addition to the analysis.

As mentioned in Chapter 3, the use of the SAT-9 to estimate impacts on student achievement is hampered by the fact that SAT-9 administration began after Project GRAD was implemented in the Davis and the Yates feeder patterns (but before implementation in the Wheatley pattern). As such, one cannot compare SAT-9 progress with a pre-program baseline. Nevertheless, changes in SAT-9 achievement can be compared with the average in the early years of implementation. Specifically, one can compare the differences between SAT-9 progress at Project GRAD schools and progress at the comparison schools in the years following the initiation of SAT-9 testing.

- **In the several years following implementation of SAT-9 testing, students' performance on the SAT-9 at the Project GRAD comparison schools actually declined.**

- **At the same time, SAT-9 achievement at the Project GRAD schools within the Davis and the Wheatley feeder patterns tended to remain relatively stable, to increase, or to fall by less than at the comparison schools.**

- **The difference between these patterns suggests that SAT-9 achievement at the Project GRAD schools in the Davis and Wheatley feeder patterns was significantly higher than it would have been in the absence of the program.**

---

[7]Koretz (2002); Klein, Hamilton, McCaffrey, and Strecher (2000).

[8]In other words, to the extent that many students were doing well on the TAAS and were answering most questions correctly, it is possible for a program to improve achievement in ways that would not be measured accurately by this test. To some extent, the use of TLI scores instead of the pass rate in analyzing TAAS results protects against this.

- **Interestingly, the results in the Yates feeder pattern do not suggest that Project GRAD improved elementary-level SAT-9 achievement at these schools.**

Figures 4.7 and 4.8 illustrate SAT-9 math and reading progress, respectively, in the Davis feeder pattern using scores from the spring of 1999 as a quasi-baseline.[9] (Figures 4.7 to 4.12 are found on pages 75 to 86.) As in earlier figures, the shaded bars show changes since the spring of 1999 at the Project GRAD schools, and the unshaded bars show changes at the comparison schools. The estimated impacts for each year are shown above each pair of bars. For example, the top panel of Figure 4.7 shows that average third-grade SAT-9 achievement in the Davis feeder pattern was relatively flat, starting at 47.1 normal curve equivalents (NCEs)[10] in the spring of 1999 and declining by 3.4 NCEs by the fourth year of the analysis (spring of 2003). However, SAT-9 achievement at the comparison schools fell noticeably after 1999. Over the same period of time, SAT-9 achievement at the comparison schools fell by 11.3 NCEs — from 52.0 NCEs in 1999 to 40.7 NCEs in 2003. Comparing these differences, there is an estimated significant, positive effect of 7.9 NCEs (or 0.4 standard deviation) on third-grade math achievement in the fourth year of the analysis. In other words, the evidence suggests that third-grade SAT-9 achievement is 0.4 standard deviation higher than it would have been without the program. Moreover, the effects are statistically significant in every year, and the magnitude of the effects grows over the course of the follow-up period. Though the estimated program effects are somewhat smaller, a similar pattern of estimated impacts on SAT-9 math scores can be seen in the fourth and fifth grades as well.

Figure 4.8 presents estimated effects on SAT-9 achievement in reading within the Davis feeder pattern. The estimates suggest significant positive effects on third-grade reading throughout the follow-up period. As is the case for math achievement, these effects are generated by the combination of relatively stable SAT-9 achievement at the Project GRAD schools compared with declining SAT-9 performance at the comparison schools. Differences in changes in SAT-9 scores between Program GRAD schools and comparison schools in fourth and fifth grades are generally small and are not statistically significant.

A somewhat different pattern can be observed in the estimated effects on SAT-9 achievement at the elementary schools in the Yates feeder pattern (Figures 4.9 and 4.10). In

---

[9]While SAT-9 testing began in 1998, it is often the case that the initial year of a particular test is a less reliable measure of student achievement than the following years. Therefore, 1999 was used as a pseudo-baseline year, and the SAT-9 analysis focuses on progress from that point on.

[10]Like percentile scores, NCEs describe students' achievement relative to the national sample against which the test was normed. NCEs can range from 0 to 100, with a mean of 50 across the population against which the test was normed. Because it is valid to perform arithmetic operations only on cardinal measures, the analysis reported here is based on NCEs rather than on percentile measures of student achievement.

general, these estimates suggest similar patterns of achievement at both sets of schools. For example, the top panel of Figure 4.9 indicates that third-grade SAT-9 math achievement at the Project GRAD schools averaged 50.9 NCEs in 1999. This fell to 45.3 NCEs by 2003 — a decline of 5.6 NCEs. At the same time, SAT-9 math scores at the comparison schools fell from 49.5 NCEs to 43.7 NCEs — a decline of 5.8 NCEs. In short, there is virtually no difference in the degree of progress on this outcome over time. The middle and bottom panels of Figure 4.9 illustrate the existence of similar patterns in fourth- and fifth-grade SAT-9 math scores. Figure 4.10 shows that SAT-9 reading achievement at both the Project GRAD and the comparison schools in the Yates feeder pattern were relatively stable since 1999, suggesting no program effects on this outcome.

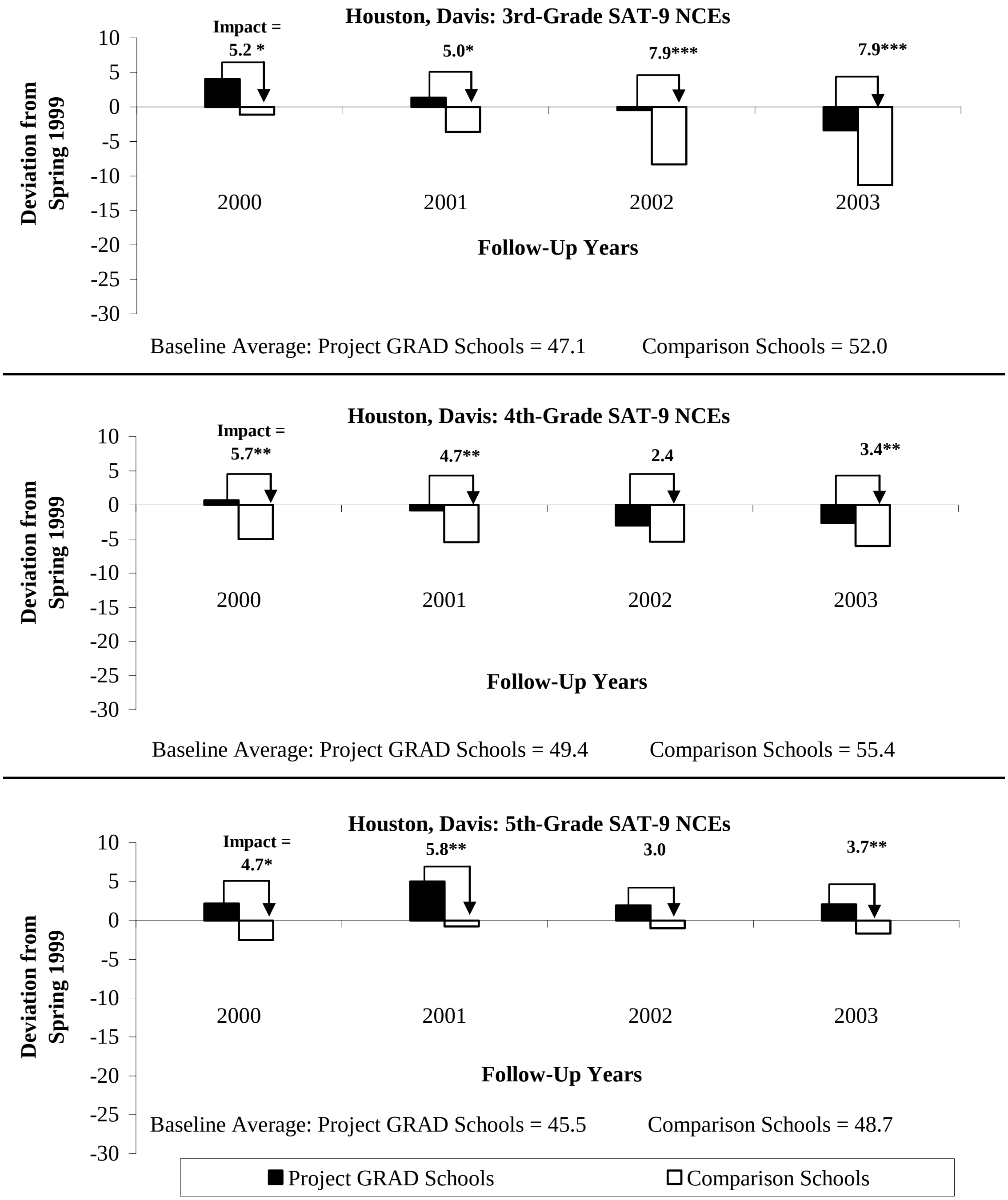
Figures 4.11 and 4.12 present estimated program effects on SAT-9 scores at the elementary schools within the Wheatley feeder pattern. As in the Davis feeder pattern, these estimates suggest that Project GRAD had a significant positive effect on achievement. For example, Figure 4.11 presents estimated effects on SAT-9 achievement in math. The top panel indicates that, in the first two years after 1999, average third-grade math achievement improved from 47.3 NCEs to 52.9 NCEs — an increase of 8.8 NCEs. At the same time, average third-grade math scores at the comparison schools declined from 51.2 NCEs to 47.2 NCEs. The difference suggests a positive effect of 9.9 NCEs, or 0.5 standard deviation. Interestingly, this effect appears to have dissipated by the end of the follow-up period. Figure 4.11 also shows effects on fourth- and fifth-grade math achievement, which are also consistently positive and statistically significant.

Figure 4.12 presents estimated effects on SAT-9 reading achievement at the elementary schools within the Wheatley feeder pattern. As is the case for third-grade math, the estimates reveal initial progress in third- and fourth-grade SAT-9 reading performance, which eventually fades out during the final years of the follow-up period. The effects on fifth-grade reading, however, are positive and statistically significant throughout the follow-up period.

The Project GRAD Evaluation

Figure 4.7

**Impact Estimates Across All Project GRAD Elementary Schools,
Davis Feeder Pattern, SAT-9 Math Scores, by Grade**



(continued)

Figure 4.7 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

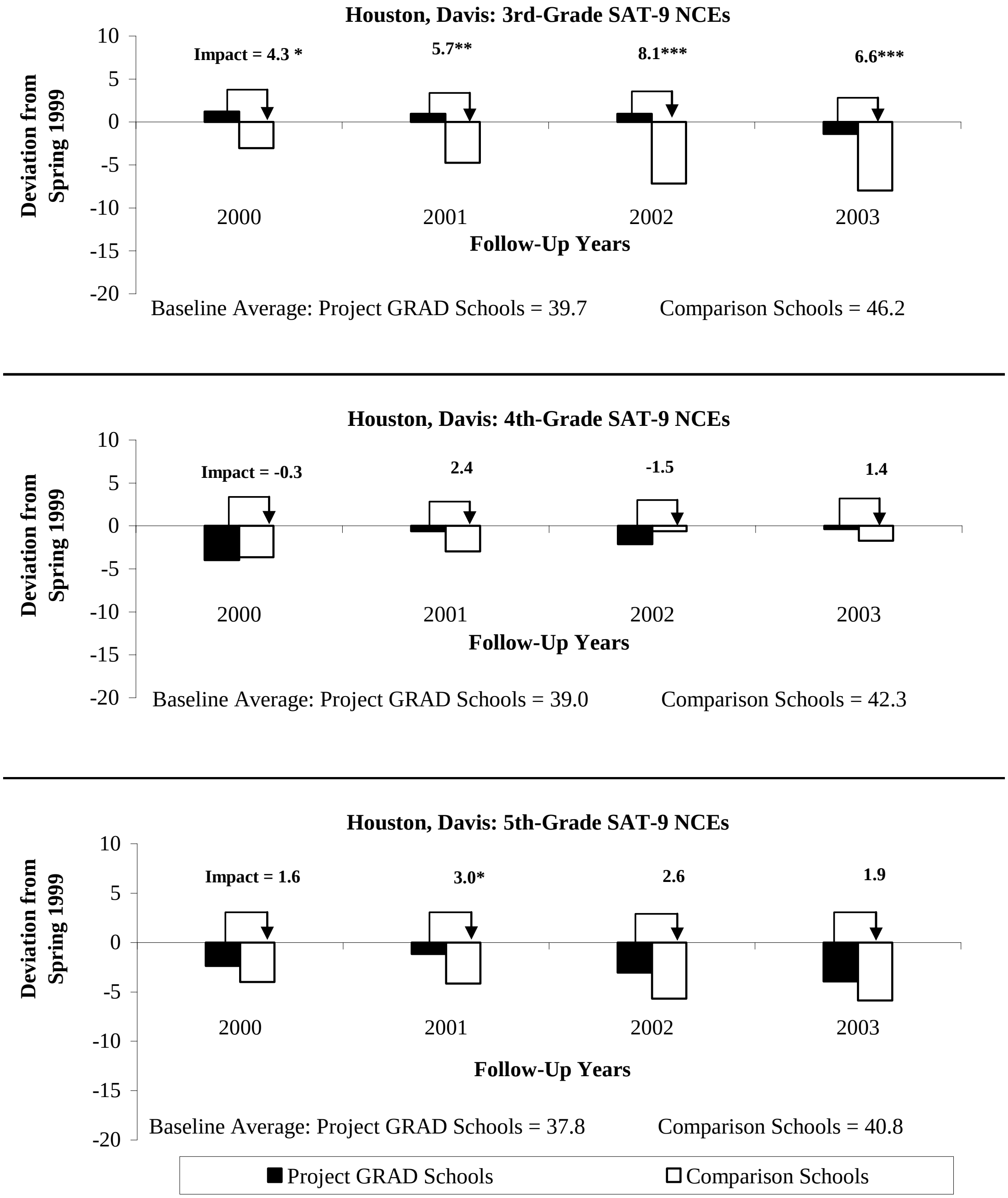
A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 8 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

The Project GRAD Evaluation

Figure 4.8

Impact Estimates Across All Project GRAD Elementary Schools, Davis Feeder Pattern, SAT-9 Reading Scores, by Grade



(continued)

Figure 4.8 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

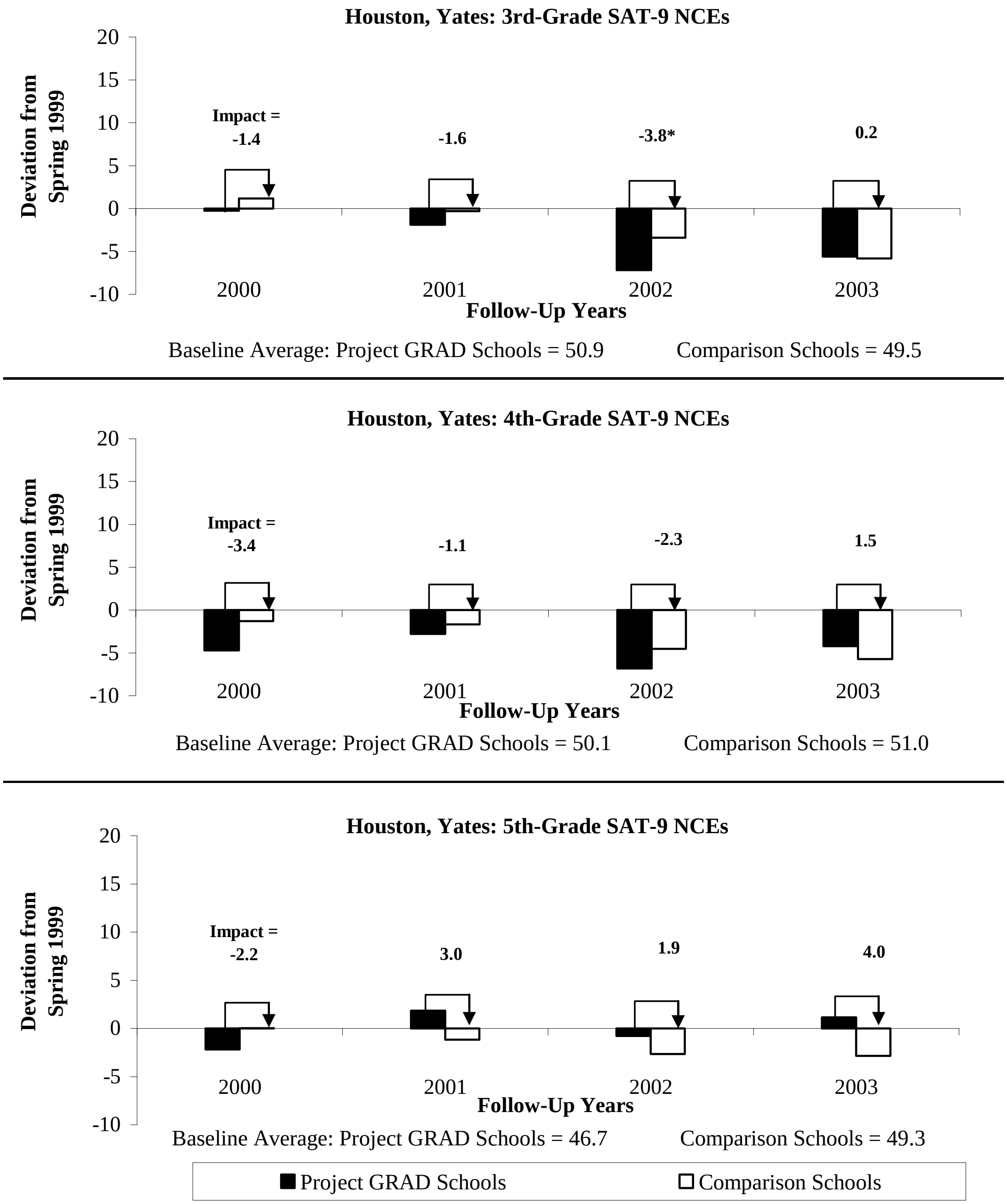
The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 8 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

Figure 4.9

Impact Estimates Across All Project GRAD Elementary Schools, Yates Feeder Pattern, SAT-9 Math Scores, by Grade



(continued)

Figure 4.9 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

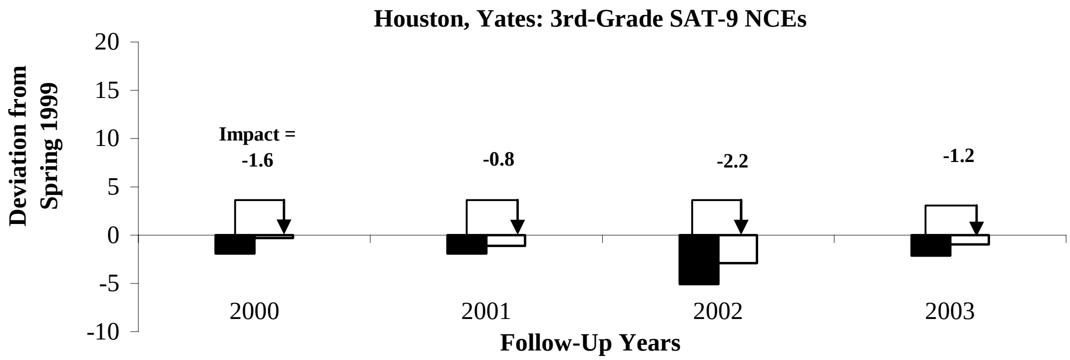
The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

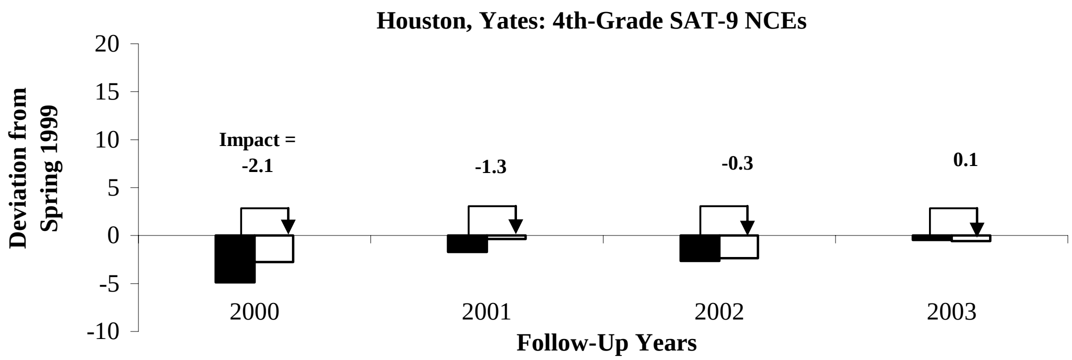
Clusters consist of a Project GRAD school matched with a group of between 4 and 15 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

Figure 4.10

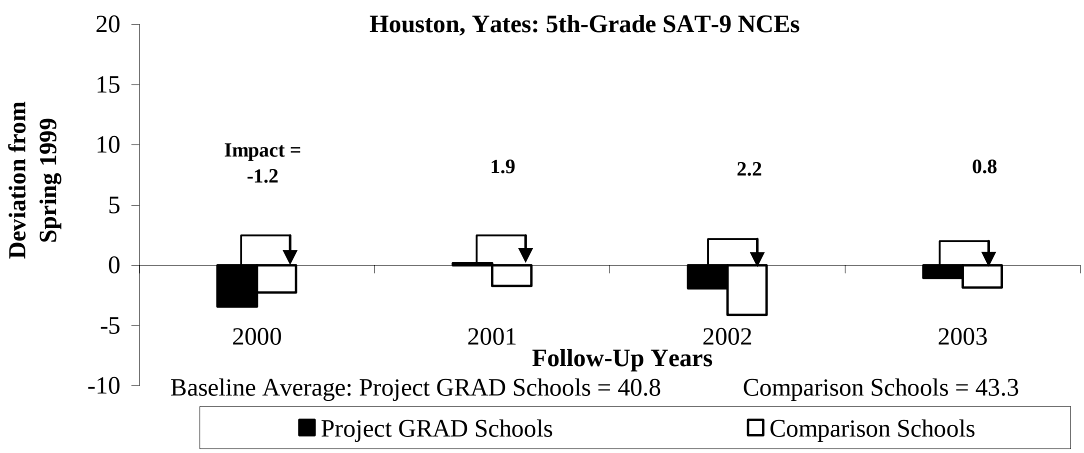
Impact Estimates Across All Project GRAD Elementary Schools, Yates Feeder Pattern, SAT-9 Reading Scores, by Grade



Baseline Average: Project GRAD Schools = 45.1 Comparison Schools = 44.2



Baseline Average: Project GRAD Schools = 42.4 Comparison Schools = 42.8



(continued)

Figure 4.10 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

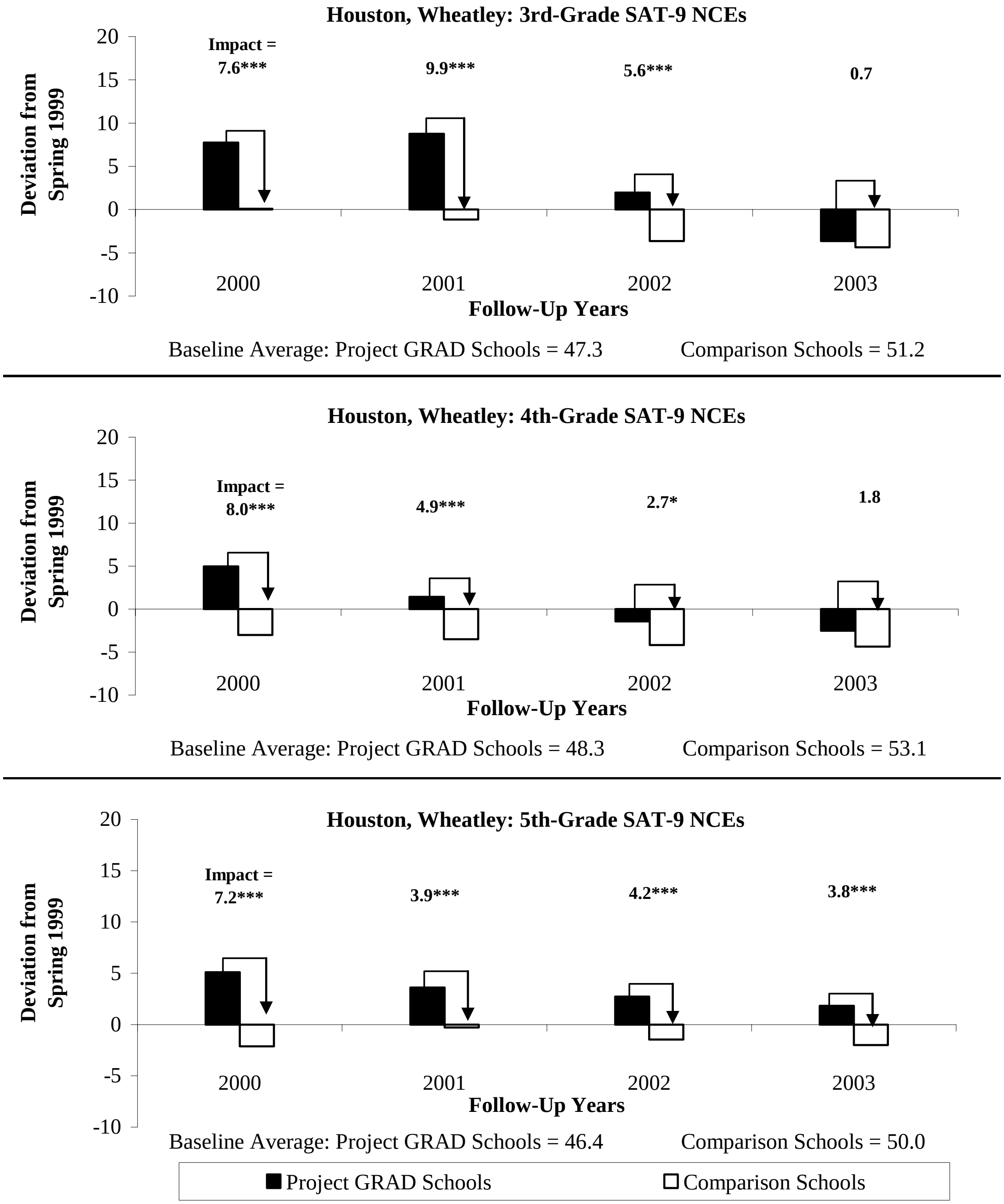
A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 4 and 15 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

The Project GRAD Evaluation

Figure 4.11

Impact Estimates Across All Project GRAD Elementary Schools, Wheatley Feeder Pattern, SAT-9 Math Scores, by Grade



(continued)

Figure 4.11 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

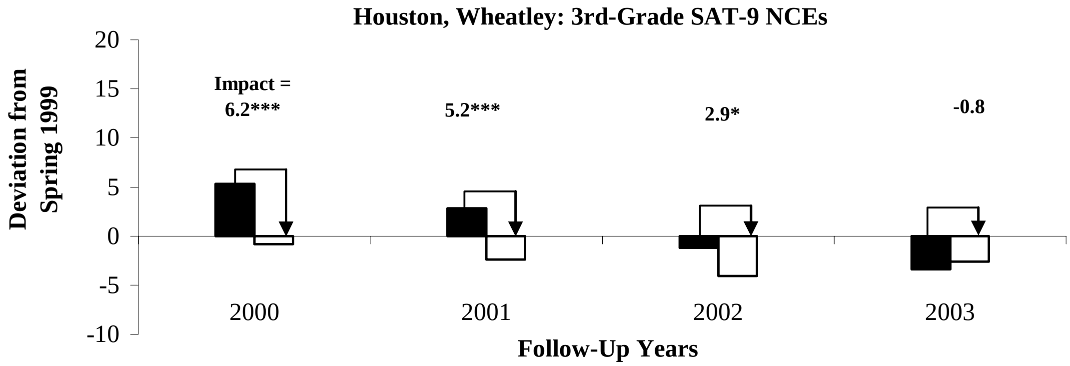
The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

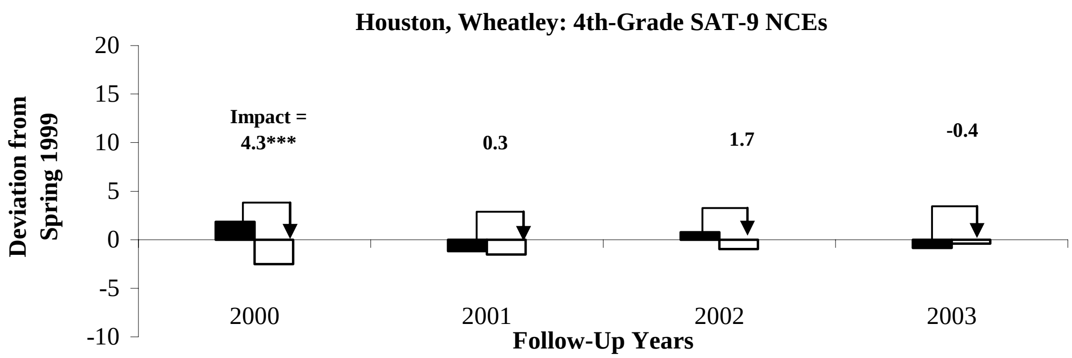
Clusters consist of a Project GRAD school matched with a group of between 2 and 19 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

Figure 4.12

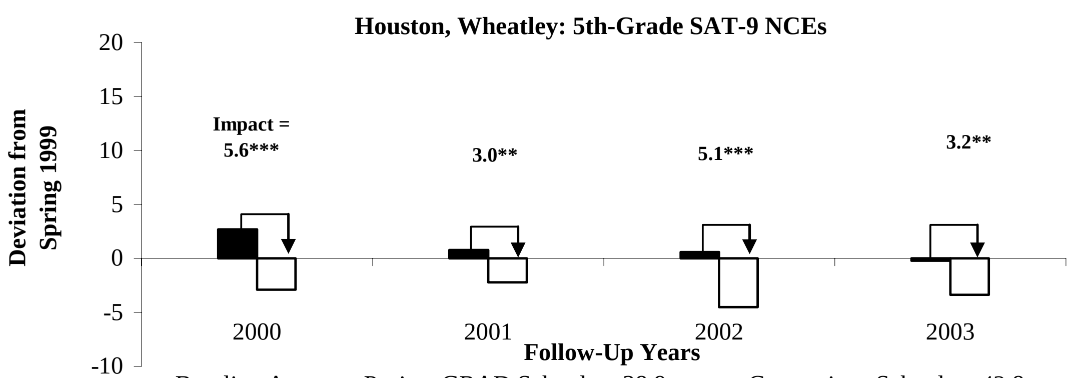
Impact Estimates Across All Project GRAD Elementary Schools, Wheatley Feeder Pattern, SAT-9 Reading Scores, by Grade



Baseline Average: Project GRAD Schools = 41.5 Comparison Schools = 44.9



Baseline Average: Project GRAD Schools = 39.8 Comparison Schools = 43.1



Baseline Average: Project GRAD Schools = 38.9 Comparison Schools = 42.8

■ Project GRAD Schools □ Comparison Schools

(continued)

Figure 4.12 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: Estimates are regression-adjusted for students' background characteristics and prior achievement.

Sample consists of students for whom administrative records exist between the 1990-1991 and 2002-2003 academic years.

The "deviation from the baseline" for each year was calculated as the difference between the baseline average and the average for the specified year.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 2 and 19 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

Implications of the Findings from Project GRAD Houston

The findings across the feeder patterns in Project GRAD Houston suggest substantial increases in TAAS scores at the Project GRAD elementary schools. Yet the comparison schools that did *not* participate in Project GRAD also experienced similar increases in these test scores. In general, the period from 1995 to 2003 was one of substantial progress in students' test scores across the low-performing schools in Houston. For example, third-grade reading achievement at the comparison schools used for the Davis feeder pattern improved by 0.61 standard deviation; the pass rates on this portion of the TAAS rose from 58 percent to 79 percent. The progress was even more substantial in math: Third-grade student achievement improved by 0.84 standard deviation, and the pass rates grew from 43 percent to 73 percent.

This ten-year period encompassed a phase of intense reform in the Houston Independent School District.[11] In fact, the available information does suggest that a substantial set of reforms aimed at improving student performance was in place at many low-performing schools during the period covered by this evaluation — among both the Project GRAD schools and the comparison schools.

Given the similarity in progress on the TAAS among both sets of schools in this reform-rich environment, it is not possible to separate out the specific contributions of Project GRAD to this overall progress. If Project GRAD was the principal improvement in the Project GRAD schools, and if other district-initiated reforms were focused on other schools (including the comparison schools), the similar progress would imply that Project GRAD and the other reforms were similarly effective in improving TAAS scores. In other words, while other school reform efforts may have produced progress at the comparison schools, Project GRAD may have been the vehicle driving the TAAS increases at the Project GRAD schools.

On the other hand, it is possible that the district- and state-initiated programs and policies that were in place at the comparison schools were *also* in place — at least to some extent — at the Project GRAD schools. Moreover, it is likely that the reforms present at the comparison schools are the best proxy for reforms that would have been implemented at the program schools in the absence of Project GRAD. However, it is not possible to determine (especially after the fact) exactly how much the district- and state-initiated efforts to improve student achievement affected Project GRAD schools. To the extent that this is the case, the findings reported here suggest that the TAAS improvements at the Project GRAD schools were driven by a combination of the program reform and other reform efforts. To the extent that the progress at these schools is similar to the progress at the schools where Project GRAD was not part of the reform mix, the marginal value of Project GRAD as a tool for improving TAAS performance is limited.

---

[11]McAdams (2000); Snipes, Doolittle, and Herlihy (2002).

Finally, some have questioned the extent to which the gains on the TAAS, in general, and in Houston, in particular, represent real improvements in student learning versus the effects of helping students improve their test-taking skills, combined with teachers' improved sense for and focus on the content of the TAAS.[12] To the extent that this is true, these results would suggest not the presence of two effective reforms but, rather, two sets of schools that were both affected by a regime of high-stakes accountability that was in place across the state.

The SAT-9 analysis presented here sheds light on this issue and suggests that — though Project GRAD did not have a differential impact on elementary students' TAAS scores — the program did have consistently positive, statistically significant, and substantively important effects on students' performance on the SAT-9. In many cases, this was a result of an erosion in SAT-9 performance that occurred at the comparison schools throughout Houston but that did *not* occur at the Project GRAD elementary schools.

This finding underscores the possibility, as has been raised in previous research, that the progress on the TAAS reflects improvements on different dimensions of skill than those measured by the SAT-9 and that the progress on the specific material germane to the TAAS may have come at the expense of the dimensions of academic skill measured by nationally norm-referenced tests such as the SAT-9. The findings also reflect the possibility that ceiling effects on the TAAS obscure meaningful variation across schools in student performance and school progress.

Perhaps just as important, these estimates strongly suggest that Project GRAD schools made progress on the TAAS at the same time that they avoided the apparent erosion in the skills measured by the SAT-9 that was occurring at comparison schools throughout the district. The resulting effects are not only statistically significant but also substantively important. The sizes of these effects range from 0.2 to 0.5 standard deviation. While there is no absolute standard for program effects, in the context of education reforms aimed at improving test score outcomes, these are relatively substantial.[13] For example, an improvement of 0.5 standard deviation translates into the difference between the 25th and the 43rd percentiles on the SAT-9.

In the context of a reform-rich environment focused on meeting state standards, Project GRAD had substantial positive effects on elementary students' achievement relative to

---

[12]See, for example, Klein, Hamilton, McCaffrey, and Strecher (2000); Haney (2000).

[13]Impact effect sizes are calculated by dividing the program effect by the standard deviation of the outcome in question. While no absolute standard exists as to what represents a large versus a small effect size, many researchers nevertheless rely on a rule of thumb that suggests that effect sizes of approximately 0.20 standard deviation or less be considered small, effect sizes of 0.50 be considered moderate, and effect sizes of 0.80 be considered large. A meta-analysis of treatment effectiveness studies by Lipsey (1990) also sheds light on these issues. This study found that, out of 102 studies — most of which were from education research — the bottom third of the distribution of impacts ranged from about 0 to 0.32, the middle third of impacts ranged from 0.33 to 0.50, and the top third of impacts ranged from 0.56 to 1.26.

national norms, in two out of three feeder patterns. Though it cannot be known for certain, these effects may reflect Project GRAD's emphasis on relatively well-articulated curricular and instructional reforms at the elementary level that emphasize fundamental academic skills rather than specific state competencies. In particular, it is possible that — even in the presence of a set of reforms focused on the specific content mandated by the state standards — the presence of Success for All and MOVE IT™ Math may have encouraged enough of a focus on core academic skills to support the current levels of achievement on the more general set of academic skills measured by the SAT-9.

Chapter 5

The Effects of Project GRAD on Elementary Student Achievement in the Early Expansion Sites

As discussed in detail in Chapter 2, Project Graduation Really Achieves Dreams (GRAD) has expanded into several different districts across the country. This chapter examines the program's effects in three early expansion sites: Newark, New Jersey; Columbus, Ohio; and Atlanta, Georgia.

The overall pattern of effects is somewhat mixed in these three sites. In Newark, Project GRAD appears, in two years of follow-up, to have reversed a substantial decline in second- and third-graders' performance on the Stanford Achievement Test (SAT-9), resulting in positive effects on achievement. In Columbus, where implementation faltered, student achievement scores at Project GRAD schools seem to have fallen behind the scores at comparison schools — at least initially. Finally, in Atlanta (a reform-rich environment like Houston), student achievement on state-mandated criterion-referenced tests at Project GRAD schools increased substantially in the several years following implementation — progress that was mirrored by the comparison schools.

Project GRAD Newark

- **Prior to the implementation of Project GRAD Newark, second- and third-grade test scores were declining substantially both at Project GRAD schools and at comparison schools throughout the district.**
- **In the first two years of follow-up after Project GRAD's implementation, student achievement at Project GRAD elementary schools improved substantially relative to the pre-program trend. No such improvement occurred at the comparison schools.**
- **This difference indicates that Project GRAD Newark had positive effects on elementary student achievement in the first two years of follow-up.**
- **The lack of follow-up data beyond two years of program implementation makes it impossible to identify the program's longer-term effects.**

Implementing the program at seven elementary schools beginning in the 1998-1999 school year, Project GRAD Newark was the first of the initiative's expansion sites. During the first year of the program, Communities In Schools (CIS), Consistency Management & Cooperative Discipline℠ (CMCD℠), and the college scholarship offer were put in place. In the second year, the Success for All (SFA) reading program began operation.[1] Unfortunately, the limited availability of data makes it impossible to analyze the Newark program's effects on elementary student achievement beyond the first two years of implementation.[2]

Over the six years prior to Project GRAD's implementation in Newark, average test scores at what became the Project GRAD schools were declining. For example, average third-grade reading scores on the Stanford Achievement Test (SAT)[3] at Project GRAD schools declined from approximately 47 normal curve equivalents (NCEs) to 36 NCEs. This is a change of about 0.56 standard deviation and is equivalent to a drop from the 44th to the 25th national percentile. In math, average third-grade achievement at Project GRAD schools declined from 53 NCEs to 41 NCEs — a decline of about 0.54 standard deviation, which is equivalent to dropping from the 56th to the 34th national percentile. Similar declines in reading and math achievement took place at the comparison schools in Newark.

These trends must be taken into account when calculating program effects. Therefore, in Newark, rather than comparing postprogram achievement with the baseline average, follow-up achievement is compared with the achievement levels that would be predicted by the baseline *trend* at each school. The average deviation from trend at the Project GRAD schools is then compared with the average deviation from trend at the comparison schools. As in Houston (see Chapter 4), the difference in the deviation from the baseline pattern at the program schools versus the comparison schools represents the estimated program effect.[4]

Figure 5.1 illustrates estimated impacts on second- and third-grade math in the Newark feeder pattern.[5] The graphs indicate that, at the program schools, achievement grew relative to the baseline trends but that, at the comparison schools, achievement fell below even already-negative baseline trends. The net result was substantial estimated program effects in both follow-up years. In particular, as shown in the lower panel of the figure, the estimated program effects indicate that third-grade math scores were approximately 9.2 NCEs, or 0.44 standard

[1]These program components are described in further detail in Chapter 1 of this report.

[2]In particular, second- and third-grade students were no longer tested after the 1999-2000 school year. Other changes in test administration also make it difficult to develop reliable impact estimates of Project GRAD's effects in any elementary grades other than the second and third.

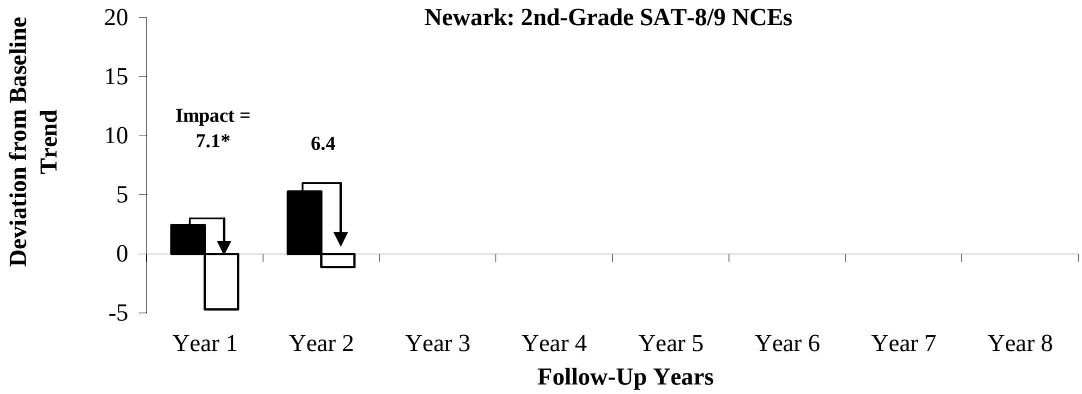
[3]Prior to the 1996-1997 school year, the Newark Public Schools administered Version 8 of the Stanford Achievement Test to elementary students throughout the district. In the spring of 1997, it began administering Version 9 of the same test.

[4]For greater detail, see Bloom (2003) and Snipes (2003).

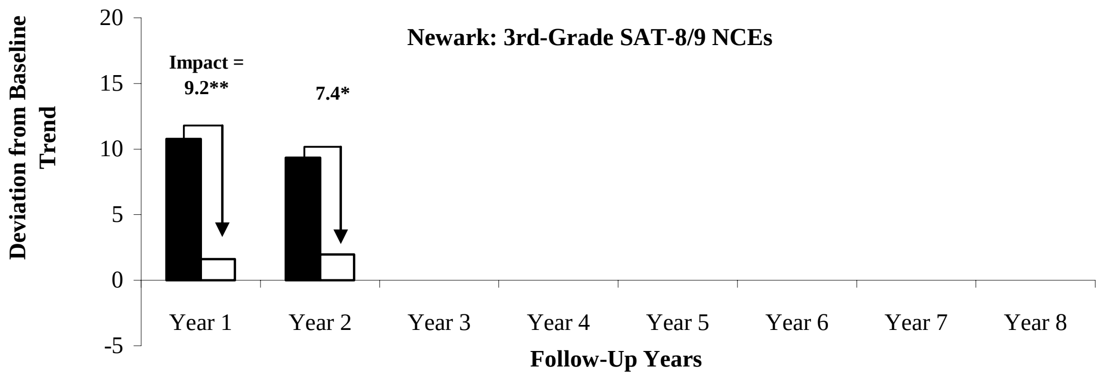
[5]Whereas the bars in the graphs for the other sites represent the deviations from the baseline average, the bars in this graph represent the deviation from the baseline trend. As in the other sites, the difference between the deviation from trend at the Project GRAD and at the comparison schools represents the estimated program impact.

Figure 5.1

Impact Estimates Across All Project GRAD Elementary Schools, Newark Feeder Pattern, Math Scores, by Grade



Baseline Average: Project GRAD Schools = 36.3 Comparison Schools = 41.1



Baseline Average: Project GRAD Schools = 38.4 Comparison Schools = 39.4

■ Project GRAD Schools □ Comparison Schools

(continued)

Figure 5.1 (continued)

SOURCE: MDRC calculations from individual student school records from Newark Public Schools.

NOTES: The "deviation from the trend" for Year 1 was calculated as the difference between the expected Year 1 achievement and the actual Year 1 achievement. The "deviation from the trend" for Year 2 was calculated as the difference between the expected Year 2 achievement and the actual Year 2 achievement.

The "impact" was calculated as the difference between the "deviation from the trend" for Project GRAD schools and the "deviation from the trend" for the comparison schools.

Estimates are regression-adjusted to account for background characteristics.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 1 and 7 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

deviation, higher than they would have been without the program, and this difference is statistically significant. In the second year of the program, third-grade math achievement was approximately 7.4 NCEs, or 0.35 standard deviation, higher than they would have been otherwise.

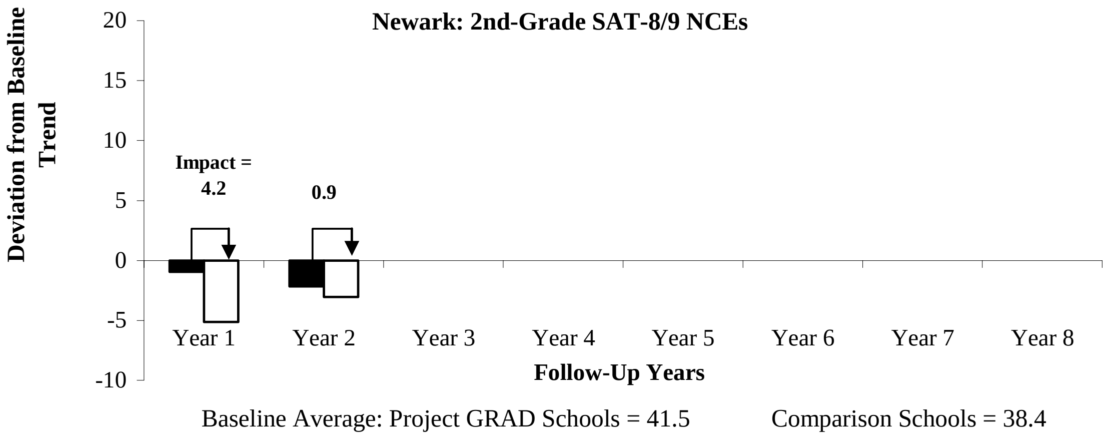
A similar pattern can be observed for the second grade, in the upper panel of Figure 5.1. In the first program year, math scores were 7.1 NCEs, or 0.32 standard deviation, above what they otherwise would have been predicted to be, and this difference is statistically significant. In the second year, this difference declined slightly and is no longer statistically significant. Interestingly, while not shown in Figure 5.1, the estimates indicate a positive, statistically significant impact on the percentage of second-graders who scored above the 25th percentile in math in both the first and the second years of the program.

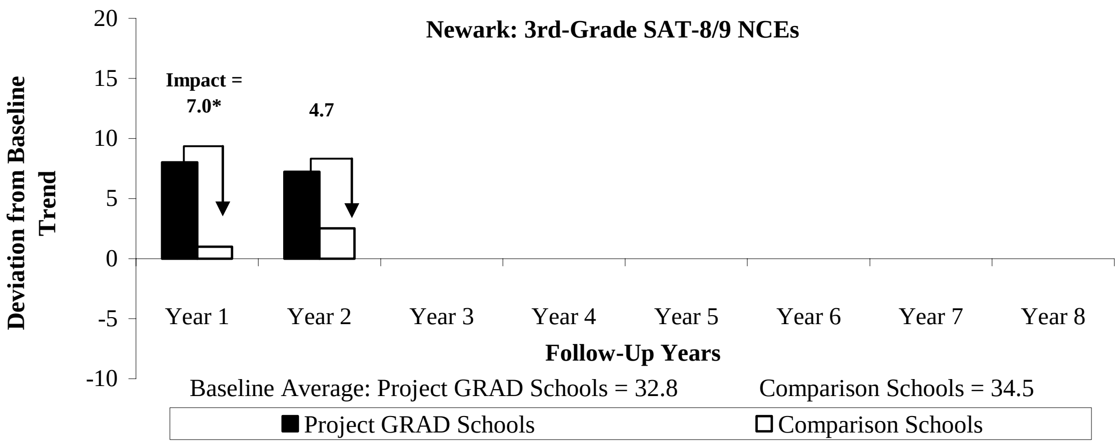
Figure 5.2 illustrates estimated program effects on second- and third-grade reading in the Newark feeder pattern. The figure does not suggest positive effects in the second grade. On the other hand, the estimates indicate that, in the first year of the program, Project GRAD Newark had a positive impact of approximately 7.1 NCEs, or 0.37 standard deviation, on third-grade reading scores. In the second year of the study, this effect appears to have diminished and is no longer statistically significant.

These estimates indicate that Project GRAD Newark had a substantial impact on early elementary achievement in the first two years of the study. These impacts represent meaningful differences between the math scores achieved at Project GRAD elementary schools and the levels

Figure 5.2

Impact Estimates Across All Project GRAD Elementary Schools, Newark Feeder Pattern, Reading Scores, by Grade





(continued)

Figure 5.2 (continued)

SOURCE: MDRC calculations from individual student school records from Newark Public Schools.

NOTES: The "deviation from the trend" for Year 1 was calculated as the difference between the expected Year 1 achievement and the actual Year 1 achievement. The "deviation from the trend" for Year 2 was calculated as the difference between the expected Year 2 achievement and the actual Year 2 achievement.

The "impact" was calculated as the difference between the "deviation from the trend" for Project GRAD schools and the "deviation from the trend" for the comparison schools.

Estimates are regression-adjusted to account for background characteristics.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 1 and 7 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

that would have occurred in the absence of the intervention. For example, in the spring of the 1999-2000 school year, average third-grade math scores across Project GRAD schools reached 45.6 NCEs, or the 48th percentile. In the absence of the intervention, the estimates imply that third-grade scores would have averaged 36.2 NCEs, or the 28th percentile. So while the math scores at Project GRAD Newark schools remained below the national average, these estimates indicate that their position relative to the rest of the nation would have been much lower in the absence of the program. Translated into effect sizes, the effects on math represent an impact of 0.44 standard deviation. While there is no absolute standard for program effects, in the context of education reforms aimed at improving test score outcomes, these are relatively substantial.[6]

These effects, however, diminished over the second year of the study, leaving at least two important questions. First, one must ask what was driving these program effects. In the first year of Project GRAD Newark, the program's curricular components were not yet implemented; the only components that were in place were classroom management and community involvement and support. Therefore, to the extent that effects occurred in the first year of the program in Newark, they must have been driven either by these noncurricular components or by some sort of "halo effect" associated with program implementation. Second, the district discontinued testing in

[6]Impact effect sizes are calculated by dividing the program effect by the standard deviation of the outcome in question. While no absolute standard exists as to what represents a large versus a small effect size, many researchers nevertheless rely on a rule of thumb that suggests that effect sizes of approximately 0.20 standard deviation or less be considered small, effect sizes of 0.50 be considered moderate, and effect sizes of 0.80 be considered large. A meta-analysis of treatment effectiveness studies by Lipsey (1990) also sheds light on these issues. This study found that, out of 102 studies — most of which were from education research — the bottom third of the distribution of impacts ranged from about 0 to 0.32, the middle third of impacts ranged from 0.33 to 0.50, and the top third of impacts ranged from 0.56 to 1.26.

the second and third grades after the second year of the program. As a result, it is impossible to ascertain whether these effects were sustained beyond the first two years of the program.

Project GRAD Columbus

- **Project GRAD Columbus does not appear to have had a positive effect on elementary student achievement.**
- **As is sometimes the case with new reforms, achievement at Project GRAD schools declined relative to the comparisons schools in the early years of the program.**
- **In most grades, the negative effects dissipated over the first several years of the study. In some cases, however, these effects persisted through the end of the follow-up period.**
- **This pattern is consistent with the presence of implementation challenges that are discussed in Chapter 2.**

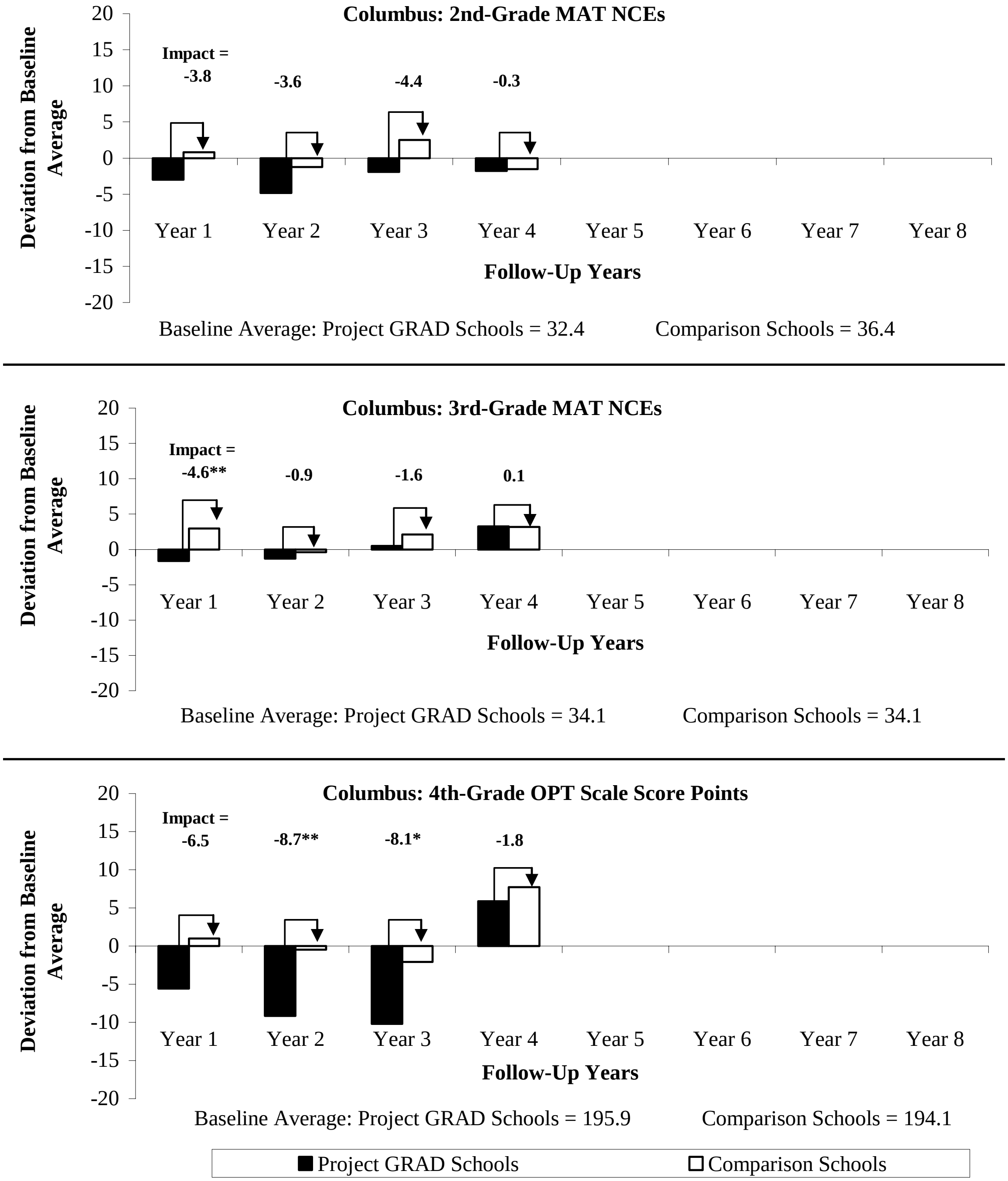
In the 1999-2000 school year, Project GRAD began implementation in a set of seven elementary schools in Columbus. Figures 5.3 and 5.4 present program effects through the first four years of implementation. Columbus Public Schools administered the Metropolitan Achievement Test (MAT) to second-, third-, and fifth-graders and the Ohio Proficiency Test (OPT) to fourth-graders. The MAT is a nationally norm-referenced test, while the OPT is a criterion-referenced test developed by the State of Ohio.

Figure 5.3 shows that, across several grades and years, average math achievement at Project GRAD Columbus schools in the follow-up period declined relative to the baseline average. In some years, this decline represented significantly less progress than was observed at the comparison schools. For example, in the first year of the program, average third-grade MAT math scores at Project GRAD schools dropped by 1.6 NCEs. At the same time, average MAT scores at the comparison schools increased by about 3 NCEs, for a net difference of negative 4.6 NCEs. This difference is statistically significant, and it implies that the presence of Project GRAD suppressed third-grade math scores in the first year of the program. These effects appear to have dissipated over the next three follow-up years, as the progress at Project GRAD schools caught up with that at the comparison schools. This is consistent with the type of disruption that can be associated with new programs. It is also consistent with the implementation findings reported in Chapter 2, which suggest that the conflicts associated with the attempts to implement Project GRAD Columbus may have been somewhat disruptive.

The Project GRAD Evaluation

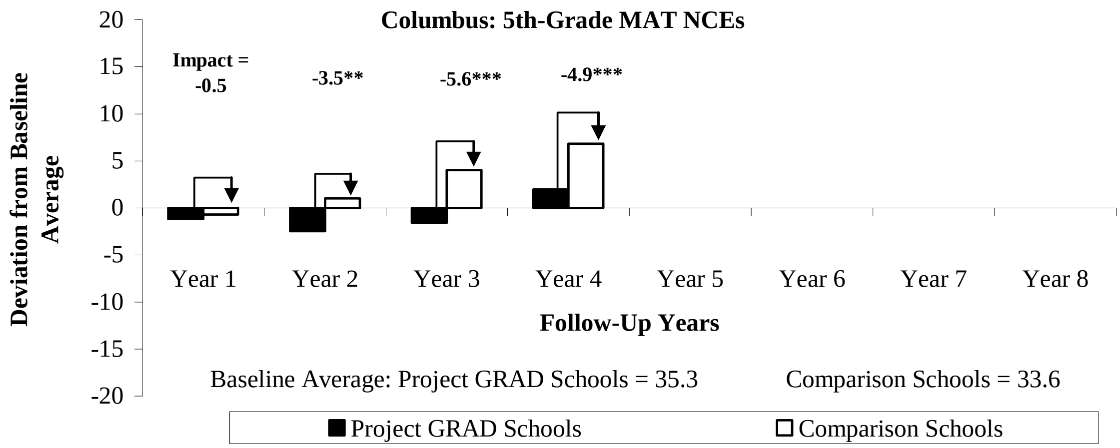
Figure 5.3

Impact Estimates Across All Project GRAD Elementary Schools, Columbus Feeder Pattern, Math Scores, by Grade



(continued)

Figure 5.3 (continued)



SOURCE: MDRC calculations from individual student school records from Columbus Public Schools.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average. The "deviation from the baseline" for Year 2 was calculated as the difference between the baseline average and the Year 2 average. The "deviation from the baseline" for Year 3 was calculated as the difference between the baseline average and the Year 3 average.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent. Homoscedasticity was assumed when summing degrees of freedom for significance-level calculations.

Clusters consist of a Project GRAD school matched with a group of between 6 and 7 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

Among fifth-graders, the negative effects on math seem to have persisted or even grown over the first four years of the study. There was little or no difference in the deviation from the baseline average of fifth-grade math scores in the first year of the program. However, from the second year on, as fifth-grade math scores improved at the comparison schools, achievement levels at the Project GRAD schools appear to have stagnated or declined slightly. Taken together, these two estimates imply a negative effect on fifth-grade math performance that did not necessarily dissipate over time.

The estimated effects of Project GRAD Columbus on reading scores are more consistent with the presence of a temporary disruption in progress. Figure 5.4 shows that, in the early years of the program, the progress in reading achievement at Project GRAD schools was generally weaker than at the comparison schools. In some grades, scores actually declined at both sets of schools. But whether reading scores were increasing or decreasing, the progress at the comparison schools generally exceeded progress at the Project GRAD schools. Among third-grade students, these differences in the first two years of implementation are statistically significant, though they dissipated over the next two years. The negative effects among fourth-grade students — though never statistically significant — declined over the follow-up period as well and even reversed in the third and fourth years of the study. Among fifth-graders, reading effects were negative throughout the follow-up period, but they are not statistically significant.

Project GRAD Atlanta

- **Project GRAD Atlanta exhibits a pattern like Houston's, in that average scores on the state achievement test rose in the three years following Project GRAD's implementation.**
- **Also as in Houston, however, this progress was matched by increased achievement at comparison schools in the district.**

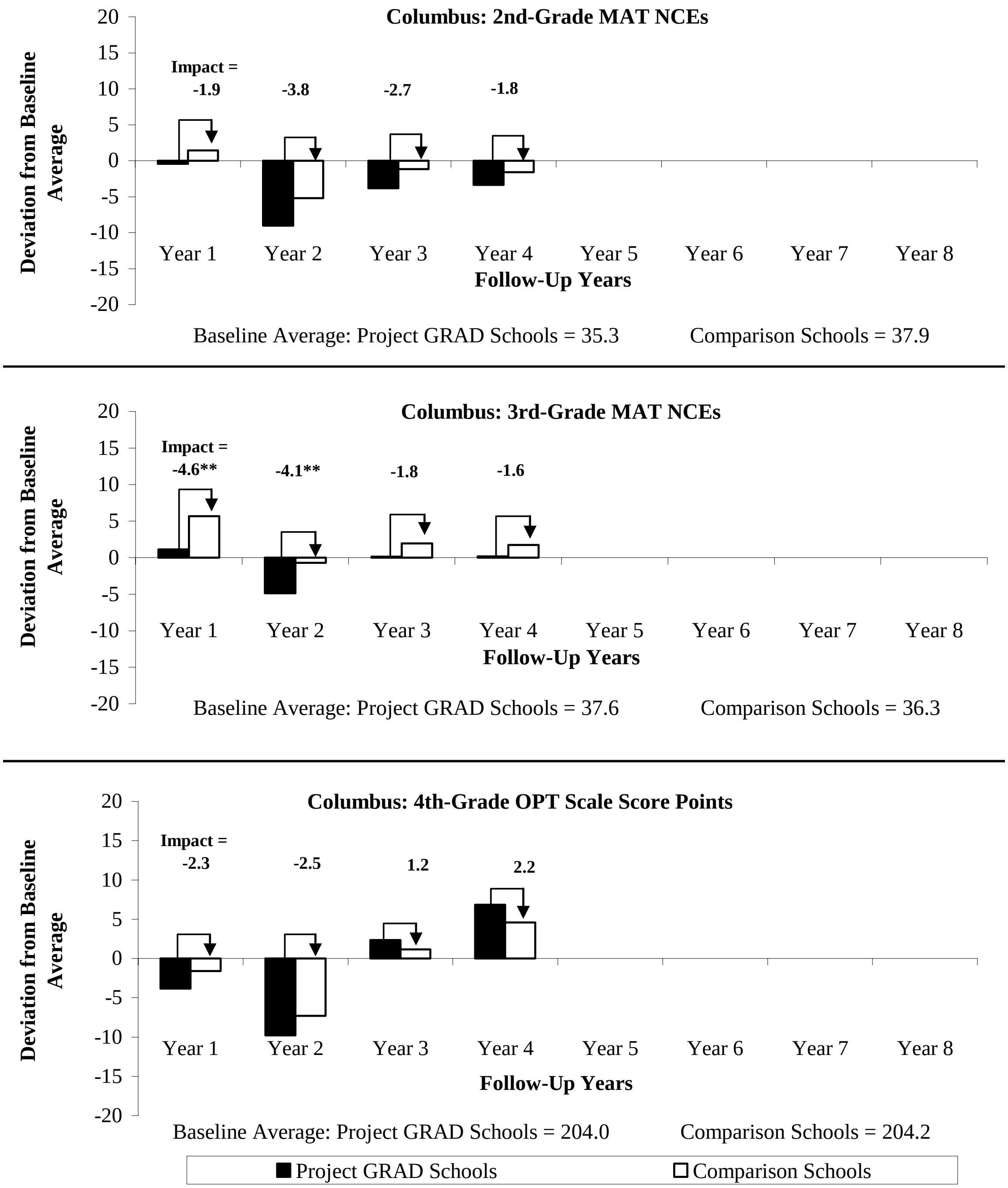
Project GRAD was implemented in nine elementary schools in Atlanta Public Schools in the 2000-2001 school year. The implementation discussion in Chapter 2 suggests some commonalities between Atlanta's program and Project GRAD Houston. In particular, both sites achieved relatively full implementation of the model. At the same time, both sites were implemented in the context of relatively "rich" reform environments with a focus on meeting state standards. The effects on achievement on the state standards tests also share some similarities, in that while they may reflect progress on the state assessments, they do not generally suggest systematic differences between the progress at Project GRAD elementary schools and the progress at comparison schools from throughout the same district.

Beginning in the spring of 2000, the Georgia Criterion-Referenced Competency Test (CRCT) was administered to fourth-graders in Atlanta Public Schools. The test has been administered in each of the three years since Project GRAD Atlanta was implemented.[7] Given that not all elementary schools implemented CMCD℠ in the 1993-1994 school year, this analysis considers 1994-1995 as the first year of Project GRAD. Figures 5.5 and 5.6 present estimated program effects on math and reading. They indicate that, among fourth-graders, average scores on

[7]Chapter 3 discusses the implications of having only one available year of baseline achievement data.

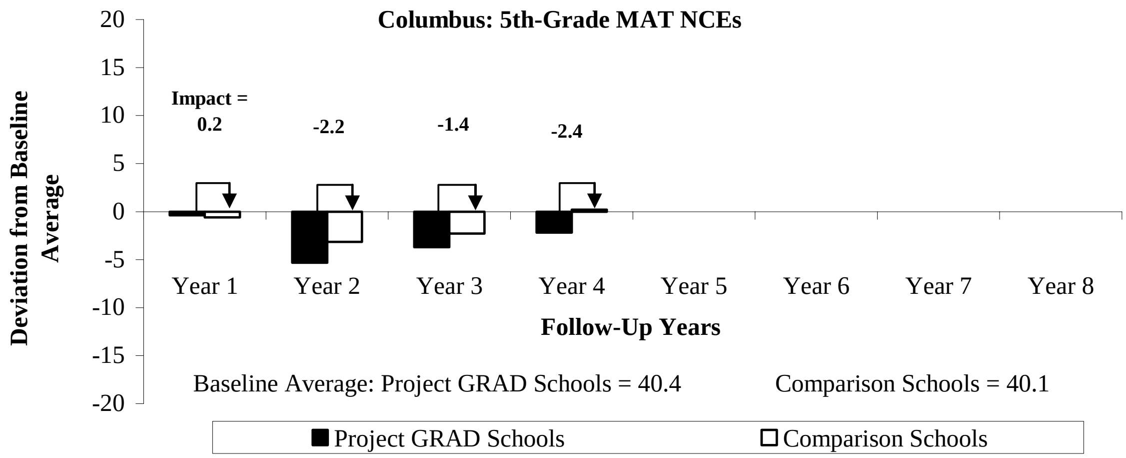
Figure 5.4

Impact Estimates Across All Project GRAD Elementary Schools, Columbus Feeder Pattern, Reading Scores, by Grade



(continued)

Figure 5.4 (continued)



SOURCE: MDRC calculations from individual student school records from Columbus Public Schools.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average. The "deviation from the baseline" for Year 2 was calculated as the difference between the baseline average and the Year 2 average. The "deviation from the baseline" for Year 3 was calculated as the difference between the baseline average and the Year 3 average.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent. Homoscedasticity was assumed when summing degrees of freedom for significance-level calculations.

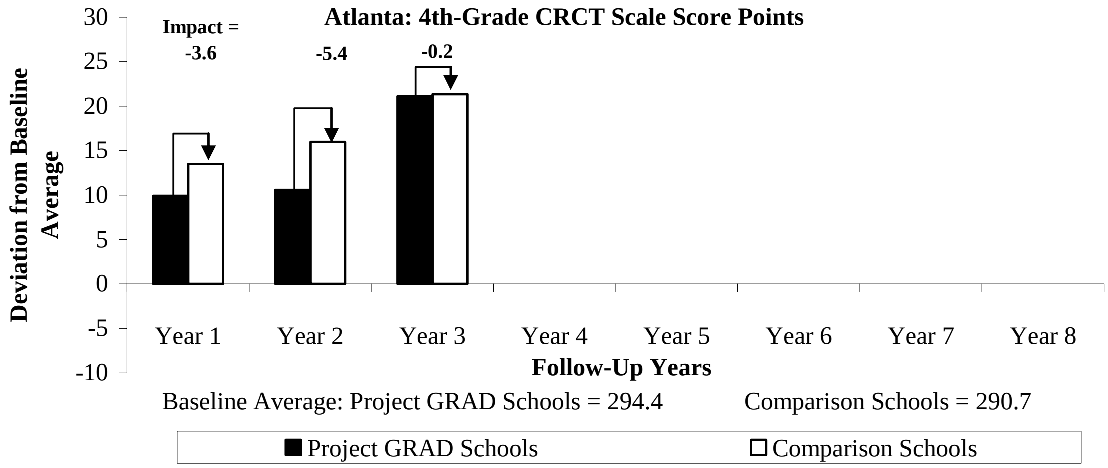
Clusters consist of a Project GRAD school matched with a group of between 6 and 7 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

the CRCT were clearly rising in the three years following implementation of the program. For example, average fourth-grade CRCT reading scores at Project GRAD schools improved by 39.4 scale score points, or 0.90 standard deviation, over the first three years of implementation (Figure 5.6). However, these improvements in state test scores did not exceed the improvements that occurred at the comparison schools. In the first two years of the program, the progress at the

The Project GRAD Evaluation

Figure 5.5

Impact Estimates Across All Project GRAD Elementary Schools, Atlanta Feeder Pattern, Fourth-Grade Math Scores



SOURCE: MDRC calculations from individual student school records from Atlanta Public Schools.

NOTES: Sample consists of students who were present during the testing period.

The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for for students' background characteristics and prior achievement.

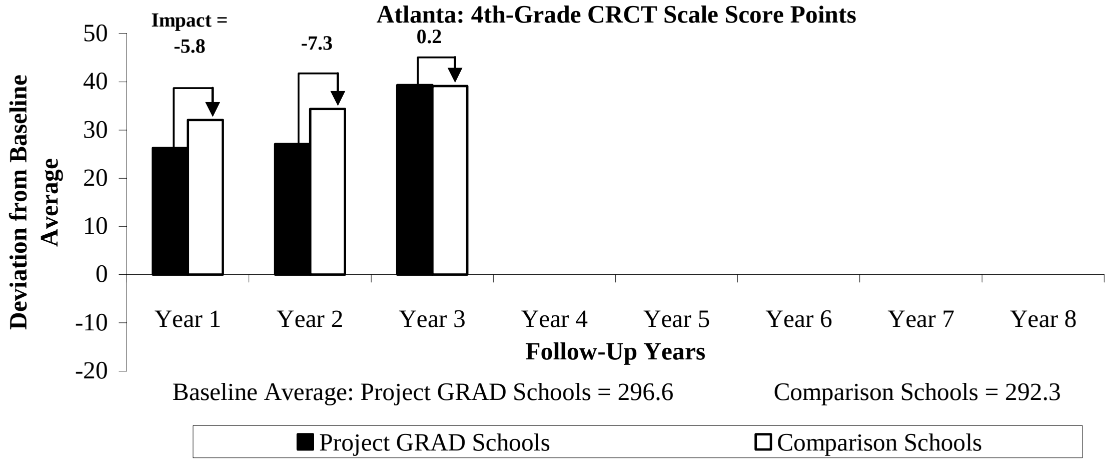
A two-tailed t-test was applied to differences in deviations from the baseline between Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 12 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

The Project GRAD Evaluation

Figure 5.6

Impact Estimates Across All Project GRAD Elementary Schools, Atlanta Feeder Pattern, Fourth-Grade Reading Scores



SOURCE: MDRC calculations from individual student school records from Atlanta Public Schools.

NOTES: Sample consists of students who were present during the testing period.

The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline between Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 12 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

comparison schools was somewhat greater than at the Project GRAD schools, though the differences are not statistically significant. By the third year of the program, the gains relative to the baseline period at both sets of schools were similar.[8]

[8]The State of Georgia also administers nationally norm-referenced tests to third-grade students in Atlanta Public Schools on an annual basis. Changes in test administration make the estimated effects on these outcomes difficult to interpret. In particular, the State of Georgia administered the Iowa Test of Basic Skills (continued)

Conclusions

Looking across Project GRAD's first three expansion sites, the overall pattern of effects is somewhat mixed. In Newark, the first expansion site, Project GRAD's estimated impacts on second- and third-graders' achievement on nationally norm-referenced assessments were substantial. In particular, the program appears to have reversed a substantial decline in achievement that would have otherwise continued. Though results are available only through the first two years of implementation in Newark, they are similar to the findings regarding norm-referenced achievement tests in Houston. Together, these findings indicate that, under some circumstances, Project GRAD's elementary-level reforms can have a substantial positive effect on student achievement.

In the other two expansion sites — Columbus and Atlanta — the analysis did not find evidence that Project GRAD improved elementary students' achievement *over and above the changes that occurred in comparison schools from the same districts.* In Columbus, where implementation faltered, student achievement scores at Project GRAD schools seem to have fallen behind the scores at comparison schools. While much of this decline appears to have faded after the first few years, some of these negative effects persisted into the fourth year of implementation. As Chapter 2 discusses, program implementation in Columbus was characterized by the presence of conflict, uneven efforts, and several attempts to restart the process. The pattern of impacts in Columbus is consistent with the hypothesis that these experiences may have resulted in a temporary disruption to the school routine that, for a time, undermined progress. The question going forward is whether or not these patterns reverse themselves if and when some measure of stable, thorough implementation is achieved.

In Atlanta — a reform-rich environment like Houston — student achievement on state-mandated criterion-referenced tests at Project GRAD schools increased substantially in the several years following implementation. However, as in Houston, the progress at the program schools was mirrored by substantial progress at the comparison schools. This is consistent with the implementation finding that, in Atlanta (as well as in Houston), there were meaningful reforms driven by the central offices as well as myriad comprehensive school reforms in place at comparison schools throughout the district.

There are several potential explanations for this pattern. First, it is possible that both Project GRAD and the reforms that were in place at the comparison schools were effective at raising student achievement. To the extent that other reform efforts in the district did not affect the Project

(ITBS) through the fall of 2001. In 2001, the first year of Project GRAD Atlanta, the third-grade test was changed to the SAT-9. In the third year of the program, 2003, the test was changed back to the ITBS. Moreover, due to a scoring error by the test publisher, third-grade test score data for 2002 — the second year of the program — are unavailable. Given the difficulty of distinguishing program effects in the context of the additional "noise" created by these changes, these estimates are not included in this report's analysis.

GRAD schools and that test scores improved at both groups of schools, one can argue that both sets of reforms were "effective." However, based on the evidence presented here, one cannot say whether Project GRAD actually drove the improvement at these schools or whether the improvements would have occurred without the presence of the program. In other words, it is possible that the reforms, including districtwide effects that drove improvement at the comparison schools, also affected achievement at the Project GRAD schools and that the program did not generate any additional progress over and above these reforms. It is also possible that — in the absence of Project GRAD — the district would have implemented the set of reforms that were generally present at comparison schools throughout the district. The estimated effects suggest that, even in the absence of Project GRAD, program schools in Atlanta would likely have experienced gains on the state tests that would have been similar to the gains that were actually observed.

Chapter 6

Conclusions and Implications

Though the ultimate, perhaps even fundamental, goal of Project Graduation Really Achieves Dreams (GRAD) is helping poor and minority children graduate from high school and succeed in college, the effectiveness of its feeder-pattern reform strategy depends in part on its ability to improve elementary student achievement. Project GRAD's early experiences in Houston — as well as other available research on the subject — suggest that improvement in students' basic academic skills is a prerequisite to improving their high school outcomes and preparing them for college. Because students' deficits in basic skills begin early in their education, so must the interventions that are designed to address them. For this reason, though Project GRAD is focused on high school graduation and college success, the vast majority of its *instructional* improvement efforts are targeted toward elementary schools. Consequently, understanding whether or not Project GRAD improves elementary student achievement is a key step in assessing the initiative's effectiveness.[1]

To address this question, this report focuses on implementation and program effects of Project GRAD's elementary school reforms in its flagship site in Houston, Texas, and in three early expansion sites across the country: Newark, New Jersey; Columbus, Ohio; and Atlanta, Georgia. Several important findings emerge from this study.

- **The ambitiousness of the Project GRAD model and the desire of its developers to scale up the initiative influenced the evolution of the model and the Project GRAD organization in several important ways.**

Unlike almost all other comprehensive school reform models, Project GRAD attempts to simultaneously address students' needs at the elementary, middle, and high school levels. While Project GRAD does not operate as a district-level reform, it pushes beyond reforms that are focused on a single school by working in "feeder patterns" of elementary and middle schools that send students to particular high schools. Though Project GRAD began as a college scholarship program for high school students, the program developers' desire to affect the skills that students bring with them to high school resulted in the expansion to the entire feeder system and the development of additional program components, including those designed to improve students' achievement in the elementary grades.

[1]A companion report from MDRC directly addresses the effects of Project GRAD on high school outcomes. See Snipes, Holton, Doolittle, and Sztejnberg (2006).

The scope of this effort — combined with the program developers' desire to expand what they believed was an effective model to new sites — required the creation of a national organization (Project GRAD USA) that developed its own technical assistance capacity, worked with other stakeholders to create local Project GRAD organizations, and learned how to work together with these local organizations to support the interventions in the expansion sites. However, even prior to the creation of Project GRAD USA, the efforts to expand into additional feeder patterns in the original site and simultaneously to reach new sites stretched the capacity of each program component's developers to support their reforms.

- **In general, Project GRAD was able to operate in varying local contexts that had differing trends in achievement, local capacity, existing reforms, and staffing rules. Across the sites, however, the success with which the components were implemented varied in important ways, with Houston and Atlanta achieving the strongest implementation.**

Some of the Project GRAD sites had strong existing district-level supports for improving instruction, while others were just beginning these efforts. In Atlanta and Houston, Project GRAD was implemented in a context of districtwide instructional reform and rising student achievement. In other sites, though districts attempted to mount reform efforts, they had not yet made substantial changes in the core of instructional activities and supports, and test scores were flat or declining. In Newark, for example, though there were several reform efforts affecting the district (including a great deal of pressure resulting from a recent state takeover), there was not yet a comprehensive approach to improving reading and math instruction, particularly at the elementary level.

The sites also had varying success in fully implementing the program's components. Though there were challenges and exceptions, Project GRAD was able to achieve complete implementation of the elementary school components in most schools in Houston and Atlanta. On the other hand, in Columbus, implementation appeared to proceed in "fits and starts," experiencing both resistance and limitations in the capacity of the local stakeholders. In Newark, strong support by funders helped the site overcome strained local administrative capacity, and, in the early phases of implementation, Project GRAD Newark met important milestones and put in place some of the model's key components.

This experience — coupled with Project GRAD's multiple-component design and implementation strategy (whereby components were phased in gradually over several years, and in a different order at each site) — means that the evaluation is unable to disentangle the contributions of individual components. As a result, the report's observations about the efficacy of Project GRAD address the model as a composite set of reform strategies that are evolving over time.

Regarding the initiative's effects on student achievement, an important story emerges from the study's findings: Project GRAD has the potential to help participating schools avoid some of the "downsides" of focusing on specific academic competencies as measured by state tests used in accountability systems. Though data limitations prevent a full examination of this theme, the Project GRAD schools with reasonably good implementation appear to have achieved similar improvements on state tests as other, similar local schools while avoiding declines in scores on national tests that measure broader academic skills not linked to specific curricula.

- **Across the sites, achievement on state standards tests improved at many Project GRAD schools, especially when the program was reasonably well implemented. Where this occurred, student achievement tended to improve by similar amounts at comparison schools from the same districts.**

- **In two of the Project GRAD districts, performance on national norm-referenced tests at the comparison schools appeared to fall in the years following program implementation. Project GRAD often prevented or lessened this deterioration in performance relative to national norms, resulting in substantial improvements in student achievement over and above the changes that would have occurred in the absence of the program.**

Implementing Project GRAD takes resources, administrative effort, and organizational development, and there are typically other substantial reforms in the Project GRAD districts. This suggests that the fundamental evaluation question involves *differential* effects: whether or not Project GRAD's considerable efforts at reform result in improvement in student achievement *over and above what would have been observed without the program.*

This report's impact estimates suggest that the answer to this question is complicated and depends in part on the specific achievement outcomes examined. In Atlanta and Houston, the districts were mounting reforms that focused on and actually produced progress toward state-mandated performance criteria. In such a setting, Project GRAD did not appear to generate systematic improvements on state assessments that were greater than those that occurred at the comparison schools, which are the best proxy for what would have happened without the presence of the program. On the other hand, in both Houston and Newark, performance on nationally norm-referenced tests was declining in schools like those that implemented Project GRAD. In these sites, the Project GRAD schools either reversed the declines or avoided them.

The net result was performance levels that were substantially higher relative to national norms than would have been the case without the program. Nationally norm-referenced tests are not linked as closely to state accountability mechanisms, and scores on them are typically seen as more difficult to improve than scores on state assessments. Moreover, several researchers have questioned whether improvements in state measures — absent progress on norm-referenced tests — represent genuine improvement in students' academic skills.[2] Thus, in Houston and Newark,

[2]Koretz (2002); Klein, Hamilton, McCaffrey, and Strecher (2000); Haney (2000).

there is evidence that Project GRAD substantially improved overall measures of elementary student achievement relative to the levels that would have occurred without the program.[3]

Finally, in Columbus, where implementation of Project GRAD faltered, the presence of the program seemed to disrupt progress in student achievement on both local and national assessments. While some of these effects were temporary, not all of them dissipated over time.

- **The study's finding are based on a small number of feeder patterns in districts that were largely at an early stage of program implementation, so it is important to be cautious about the implications of these findings for the future work of Project GRAD.**

Though the analysis in this report includes many schools, the evaluation represents the experience of only four district sites and six feeder patterns. The expansion sites were the first of the new districts added to the Project GRAD network. Important revisions in the current implementation process — many growing out of this early experience — are not captured in this evaluation. In addition, with the exception of Project GRAD Houston, this evaluation focuses on a relatively early period in the life cycle of these programs. Many argue that it takes at least five years for effective educational reforms to take hold and show results, which highlights the potential that results in Columbus and Atlanta might improve in the future.[4] Finally, though there are positive results in Newark, the district's change in tests does not provide enough follow-up data to know whether or not these effects will persist over time. Thus, there is simply not sufficient evidence to assume that implementation of the evolving Project GRAD model in different sites and different circumstances will necessarily follow the same patterns as those shown here.

- **These findings suggest that site selection may be an important element in Project GRAD's future success in elementary settings. The most promising settings are districts that have a clear need for *added* instructional improvement efforts and where there is likely to be strong local support for implementation.**

The manner in which the reform context, program implementation, and program effects vary across districts may have implications for how one interprets the results in this report. In some districts — even in low-performing districts serving large proportions of economically disadvantaged students — ongoing reforms may be producing rising achievement scores, even though achievement levels may still be lower than desired. This is particularly likely to be the case for performance on state-mandated standards-based assessments. In these settings, Project GRAD may not fill a gap in the efforts already in place to improve elementary instruction in ways that help meet these standards, and the initiative may possibly compete for attention and

[3]In Atlanta, the frequent changes in testing using national tests prevents a similar analysis.
[4]Bloom (2001); Borman (2002).

support. Even in these contexts, however, Project GRAD may improve (or at least prevent the erosion of) student performance on the more general skills that are not necessarily measured by state standards tests.

In other districts, existing reform efforts may not yet be providing adequate support to improve elementary-level instruction, and Project GRAD's programmatic and structural elements may meet an important need. In other words, Project GRAD may have the greatest potential to improve elementary student achievement in districts where there is limited support for basic elementary reading and math curricula and where elementary student achievement has been flat or declining. This would not mean focusing only on districts with low achievement and high levels of disadvantaged and minority students (for example, Atlanta and Houston); rather, it would imply focusing on districts where Project GRAD's emphasis on elementary-level instruction in reading and math and on classroom management and social service supports would represent a difference or added value over and above the district- and school-level reforms that are already in place. Project GRAD's core instructional components are not tailored to meet the specific standards in each district, and its track record suggests stronger effects on norm-referenced achievement tests than on state standards tests. Therefore, evaluating the appropriateness of Project GRAD for particular sites probably involves making some judgments about the relative importance of progress on the specific state standards tests in place in each district.

Project GRAD is in place in a variety of school districts across the country, including some where the reform field is not especially crowded. In these sites, if there is still substantial room for improvement in the implementation of Project GRAD, the findings here suggest that there may be significant payoffs to additional investments in achieving full implementation. In other words, if Project GRAD can improve implementation in sites where other effective reforms do not exist, the program may have the potential to improve outcomes on student achievement.

To the extent that Project GRAD USA seeks to work in districts that already have substantial and apparently effective reforms in place, the organization could examine the similarities and differences between the content of its reforms and other existing reform approaches to discover whether the model could be supplemented or changed to be more effective. In particular, it might be possible to build on and extend existing reforms, rather than substituting one reform for another. Clearly, though state test scores are improving in such districts as Atlanta and Houston, there is further still to go. Districts like these still need assistance in improving the dimensions of student achievement beyond what is measured by state standards tests, as well as help in raising student achievement to the levels needed to succeed in high school and beyond. Finally, to the extent that there are limitations on what can be accomplished in the short run in terms of elementary-level achievement, Project GRAD may want to consider increasing the instructional support components at the middle and high school levels of the model, to address the skill deficits that hamper progress in secondary schools in most urban school districts.

Appendix A

Impact Effect Sizes

Appendix Table A.1 reports estimates of baseline achievement levels, deviations from the baseline average, and program effects on math performance, as measured by both the Texas Learning Index (TLI) score[1] and the pass rate, among third-graders in the Jefferson Davis High School feeder pattern in Houston. Appendix Table A.2 presents a similar analysis for reading achievement.

- Section I of the tables shows the average outcomes and the differences between baseline performance and average performance in each year, that is, the deviation from baseline, for the Project Graduation Really Achieves Dreams (GRAD) feeder pattern and for its comparison schools during the first four years of program implementation.

- Section II of the tables compares the differences between the deviations from baseline attendance at Project GRAD schools and at the comparison schools. This difference represents the estimated "impact" of the program on student attendance.

- Section III of the tables shows the impact effect size, which is a standardized measure that is calculated by dividing the impact by the standard deviation for the baseline years.

In general, over the course of the follow-up period, test scores went up relative to the baseline levels. While there are some positive and negative differences for particular grades and years, the differences do not generally suggest a pattern of sustained positive effects. As an example, in the first year of the program, math TLI scores among third-graders at Project GRAD schools appear to have improved, moving from an average 60.1 points at baseline to 65.3 points in Year 1 of the program, an improvement of 5.2 points. At the same time, the average TLI score at the comparison schools fell from 63.1 at baseline to 62.7 points in Year 1, a difference of –0.4 point. This suggests a statistically significant impact of 5.6 points. In other words, the estimates indicate that, during the first year of implementation, Project GRAD had a positive effect of nearly 6 points on the TLI scores in the Davis feeder pattern.

[1]The TLI is a continuous score that describes a student's performance on both the mathematics and the reading tests of the Texas Assessment of Academic Skills (TAAS). The raw score on the TAAS is simply the number answered correctly on the test, and since this raw score can be interpreted only in relation to the number of items on the test, the score is limited in use. The TLI makes it possible to relate student performance to a passing standard and to compare student performance from year to year. In each year, the raw scores are standardized into TLI scores relative to the state's passing standard of 70. For TLI frequency distributions for each grade and subject, see Texas Education Center (2005).

For each feeder pattern in the evaluation of Project GRAD, Appendix Table A.3 presents achievement impact estimates, translated into an “effect size” metric. In particular, the table compares the average difference in the deviation from baseline at the Project GRAD schools with the deviation from baseline at the comparison schools, divided by the standard deviation of the outcome in question. The estimated program effect is divided by the standard deviation of the outcome in question across the entire analysis sample for each district during the baseline period, that is, during the years immediately preceding the implementation of Project GRAD.[2]

[2]For analyses of Project GRAD in Houston, where there were three separate feeder patterns, the standard deviation is calculated across the 1993-1994 through 1995-1996 school years.

The Project GRAD Evaluation

Appendix Table A.1

Interrupted Time Series and Impact Estimates for Selected Third-Grade Math Test Score Outcomes, Nine-Year Follow-Up Results, Davis Feeder Pattern

	I. Outcome Levels Compared with Baseline Year and Follow-Up Years									
	Project GRAD Schools					Comparison Schools				
Outcome	Baseline	Year 1	Year 2	Year 3	Year 4	Baseline	Year 1	Year 2	Year 3	Year 4
Math total										
TLI score	60.1	65.3	68.5	70.5	70.9	63.1	62.7	67.4	72.5	71.3
Deviation from baseline average		5.2	8.4	10.4	10.8		-0.4	4.3	9.4	8.1
Pass rate	65.1	55.4	59.6	54.2	68.4	63.6	43.5	54.8	65.0	60.8
Deviation from baseline average		-9.7	-5.5	-10.9	3.3		-20.1	-8.8	1.4	-2.8

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools				*III. Impact Effect Size* [a]			
	Year 1	Year 2	Year 3	Year 4	Year 1	Year 2	Year 3	Year 4
Math total								
TLI score								
Deviation from baseline average	5.6 *	4.1	1.1	2.7	0.3 *	0.2	0.1	0.2
Pass rate								
Deviation from baseline average	10.4	3.3	-12.3	6.1	0.2	0.1	-0.3	0.1

(continued)

Appendix Table A.1 (continued)

	I. Outcome Levels Compared with Baseline Year and Follow-Up Years											
	Project GRAD Schools						Comparison Schools					
Outcome	Baseline	Year 5	Year 6	Year 7	Year 8	Year 9	Baseline	Year 5	Year 6	Year 7	Year 8	Year 9
Math total												
TLI score	60.1	67.1	72.9	74.7	82.4		63.1	70.5	73.0	75.0	76.7	
Deviation from baseline average		7.0	12.8	14.6	22.3			7.4	9.9	11.9	13.6	
Pass rate	65.1	54.6	66.0	64.9	91.6	91.1	63.6	61.2	63.7	68.8	73.0	72.1
Deviation from baseline average		-10.5	0.9	-0.2	26.5	25.4		-2.4	0.1	5.2	9.4	9.5

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools					*III. Impact Effect Size* [a]				
	Year 5	Year 6	Year 7	Year 8	Year 9	Year 5	Year 6	Year 7	Year 8	Year 9
Math total										
TLI score										
Deviation from baseline average	-0.3	2.9	2.7	8.7 ***		0.0	0.2	0.2	0.5 ***	
Pass rate										
Deviation from baseline average	-8.0	0.8	-5.3	17.1 *	15.9	-0.2	0.0	-0.1	0.4 *	0.3

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average.

The TLI score is not available in Year 9 because the state test changed from the TAAS, which was issued in the prior years, to the TAKS in 2003.

The "impact" was calculated as the difference between the "deviation from the baseline" for the Project GRAD schools and the "deviation from baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 8 comparison schools. Results in the Comparison Schools columns reflect averages across these groups of non-Project GRAD schools.

[a]The "impact effect size" was calculated by dividing the "impact" by the standard deviation from school years 1998-2000 of outcomes for all 3rd-grade students in the district's nonselective comprehensive elementary schools.

The Project GRAD Evaluation

Appendix Table A.2

Interrupted Time Series and Impact Estimates for Selected Third-Grade Reading Test Score Outcomes, Nine-Year Follow-Up Results, Davis Feeder Pattern

	I. Outcome Levels Compared with Baseline Year and Follow-Up Years									
	Project GRAD Schools					Comparison Schools				
Outcome	Baseline	Year 1	Year 2	Year 3	Year 4	Baseline	Year 1	Year 2	Year 3	Year 4
Reading total										
TLI score	70.2	72.7	67.6	69.4	74.0	70.7	69.3	68.8	73.5	75.5
Deviation from baseline average		2.5	-2.6	-0.9	3.8		-1.4	-1.9	2.8	4.8
Pass rate	68.5	67.4	54.7	56.2	71.3	64.0	58.2	59.9	66.0	72.5
Deviation from baseline average		-1.1	-13.8	-12.3	2.8		-5.8	-4.0	2.0	8.5

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools				*III. Impact Effect Sizes* [a]			
	Year 1	Year 2	Year 3	Year 4	Year 1	Year 2	Year 3	Year 4
Reading total								
TLI score								
Deviation from baseline average	3.9	-0.7	-3.6	-1.0	0.2	0.0	-0.2	-0.1
Pass rate								
Deviation from baseline average	4.7	-9.7	-14.3 **	-5.7	0.1	-0.2	-0.3 **	-0.1

(continued)

Appendix Table A.2 (continued)

	I. Outcome Levels Compared with Baseline Year and Follow-Up Years											
	Project GRAD Schools						Comparison Schools					
Outcomes	Baseline	Year 5	Year 6	Year 7	Year 8	Year 9	Baseline	Year 5	Year 6	Year 7	Year 8	Year 9
Reading total												
TLI score	70.2	75.2	77.6	77.4	84.2		70.7	77.0	78.5	78.2	79.1	
Deviation from baseline average		5.0	7.4	7.2	14.0			6.4	7.8	7.5	8.4	
Pass rate	68.5	70.9	80.4	76.1	92.7	48.5	64.0	74.3	79.2	75.3	78.8	47.3
Deviation from baseline average		2.4	11.9	7.6	24.2	-19.8		10.3	15.3	11.3	14.8	-15.7

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools					*III. Impact Effect Sizes* [a]				
	Year 5	Year 6	Year 7	Year 8	Year 9	Year 5	Year 6	Year 7	Year 8	Year 9
Reading total										
TLI score										
Deviation from baseline average	-1.3	-0.4	-0.3	5.5 **		-0.1	0.0	0.0	0.3 **	
Pass rate										
Deviation from baseline average	-7.9	-3.4	-3.7	9.4	-4.1	-0.2	-0.1	-0.1	0.2	-0.1

(continued)

Appendix Table A.2 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average.

The TLI score is not available in Year 9 because the state test changed from the TAAS, which was issued in the prior years, to the TAKS in 2003.

The "impact" was calculated as the difference between the "deviation from the baseline" for the Project GRAD schools and the "deviation from baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 8 comparison schools. Results in the Comparison Schools columns reflect averages across these groups of non-Project GRAD schools.

[a]The "impact effect size" was calculated by dividing the "impact" by the standard deviation from school years 1998-2000 of outcomes for all 3rd-grade students in the district's nonselective comprehensive elementary schools.

The Project GRAD Evaluation

Appendix Table A.3

ITS[a] Impact Effect Sizes, Follow-Up Results, by Feeder Pattern

Feeder Pattern	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Houston, Davis								
Math								
3rd-grade TAAS	0.34 *	0.25	0.06	0.16	-0.02	0.17	0.16	0.53 ***
4th-grade TAAS	-0.24	-0.39 ***	-0.10	0.09	-0.28 *	0.00	-0.07	-0.10
5th-grade TAAS	-0.28 ***	-0.12	0.13	0.19	-0.29 ***	-0.04	0.06	-0.04
Reading								
3rd-grade TAAS	0.24	-0.04	-0.22	-0.06	-0.08	-0.02	-0.02	0.34 **
4th-grade TAAS	-0.31 *	-0.36 **	-0.03	0.06	-0.22	0.00	-0.14	-0.08
5th-grade TAAS	-0.08	0.08	0.22 *	0.41 ***	-0.23 *	0.07	0.20 *	0.14
Houston, Yates								
Math								
3rd-grade TAAS	0.01	0.19	0.06	0.06	0.01	-0.13		
4th-grade TAAS	-0.11	0.07	-0.15	-0.21 *	0.00	-0.09		
5th-grade TAAS	0.02	0.07	-0.03	0.00	0.02	-0.07		
Reading								
3rd-grade TAAS	-0.06	0.18 *	0.08	-0.01	0.09	-0.03		
4th-grade TAAS	-0.29 ***	0.09	-0.30 ***	-0.24 **	-0.09	-0.17		
5th-grade TAAS	0.04	0.10	-0.10	-0.03	0.01	-0.09		
Houston, Wheatley								
Math								
3rd-grade TAAS	0.10	0.18 *	0.06					
4th-grade TAAS	0.20 **	0.06	0.08					
5th-grade TAAS	0.17 **	0.11	0.10					
Reading								
3rd-grade TAAS	-0.04	0.04	0.00					
4th-grade TAAS	0.08	0.03	-0.01					
5th-grade TAAS	0.08	0.11	0.15 *					

(continued)

Appendix Table A.3 (continued)

Feeder Pattern	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Newark								
Math								
2nd-grade SAT[b]	0.32 *	0.29						
3rd-grade SAT	0.44 **	0.35 *						
Reading								
2nd-grade SAT	0.22	0.05						
3rd-grade SAT	0.37 *	0.24						
Columbus								
Math								
2nd-grade MAT	-0.17	-0.16	-0.20	-0.01				
3rd-grade MAT	-0.21 *	-0.04	-0.07	0.00				
4th-grade OPT	-0.26	-0.35 *	-0.32	-0.07				
5th-grade MAT	-0.02	-0.14 *	-0.22 **	-0.19 **				
Reading								
2nd-grade MAT	-0.09	-0.18	-0.13	-0.09				
3rd-grade MAT	-0.23 **	-0.21 *	-0.09	-0.08				
4th-grade OPT	-0.12	-0.13	0.06	0.12				
5th-grade MAT	0.01	-0.10	-0.07	-0.11				
Atlanta								
Math								
4th-grade CRCT	-0.13	-0.19	-0.01					
Reading								
4th-grade CRCT	-0.13	-0.19	-0.01					

(continued)

Appendix Table A.3 (continued)

SOURCES: MDRC calculations from individual student school records from the Houston Independent School District, Newark Public Schools, Columbus Public Schools, and Atlanta Public Schools.

NOTES: Texas Assessment of Academic Skills (TAAS); Stanford Achievement Test, versions 8 and 9 (SAT); Metropolitan Achievement Test (MAT); Ohio Proficiency Test (OPT); Criterion-Referenced Competency Test (CRCT).

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Impact estimates generated by comparative interrupted time series (ITS) analysis is described in detail in Chapter 3.

[b]In 1999, the test administered changed from the SAT-8 to the SAT-9.

Appendix B

Selecting Comparison Schools

Comparison schools are included in the evaluation of Project Graduation Really Achieves Dreams (GRAD) in order to build on the information already provided in the baseline patterns at Project GRAD schools. Combined with the information regarding student achievement at baseline, information from the comparison schools provides an estimate of the outcomes that would have been observed at the program schools in the absence of an intervention. Specifically, the comparison schools refine the estimates of how student achievement outcomes in the program schools would have *changed* in the absence of Project GRAD. For example, if student achievement grows dramatically at schools that are similar to the Project GRAD schools in terms of students' demographic characteristics and prior achievement, then — if appropriate comparison schools have been selected — it may be reasonable to conclude that Project GRAD schools would have experienced a similar amount of progress even in the absence of the program. On the other hand, if outcomes at a set of comparison schools decline after the baseline period, the most reasonable conclusion may be that test scores at the Project GRAD schools would have declined absent the intervention of the program.

If the comparison schools are to serve this purpose, they must be as similar as possible to the program schools in the analysis. Therefore, the goal in choosing comparison schools is to find a set of schools that, absent any intervention, would be expected to perform in a manner similar to that of the Project GRAD schools. Both logic and evidence suggest that the most accurate predictor of future performance on any given outcome is past performance on that same outcome. Inasmuch as the elementary-level analysis focuses primarily on the effects of Project GRAD on student achievement, prior academic achievement at the schools in question was the primary basis on which comparison schools were chosen.

In order to ensure "face validity" as well as to guard against the possibility that schools serving different populations of students could evolve differently in response to the same local events, the analysis also limits comparison schools to those that served demographically similar student bodies. In particular, for each Project GRAD school, the selected set of comparison schools from the same school district met the following criteria:

- Average achievement in reading and math during the baseline period (typically the three years prior to Project GRAD's implementation) were each within 0.25 standard deviation of average achievement at the Project GRAD school.

- The percentages of students in key racial/ethnic groups were within 20 percentage points of the percentage of levels for modal racial/ethnic groups at the Project GRAD school.

In the case of Project GRAD Newark, because there was a negative trend in achievement in the several years prior to program implementation, comparison schools were limited to those that had baseline trends that were in the same direction as the Project GRAD school in question.

Appendix Tables B.1 through B.6 present the characteristics of Project GRAD and comparison schools during the baseline period for each feeder pattern in the analysis. These tables show that, during the years immediately preceding the implementation of Project GRAD, the comparison schools chosen for the analysis served similar populations of students and exhibited similar levels of achievement. The tables also show that, in general, the Project GRAD schools and their comparison counterparts served higher proportions of poor and minority students than the other elementary schools in their districts.

The Project GRAD Evaluation

Appendix Table B.1

Characteristics of Davis Feeder Pattern Elementary Schools, 1991-1992 Through 1993-1994

Characteristic	Project GRAD Schools	Comparison Schools	All Project GRAD and Comparison Schools in the District
Average school size			
1st grade	108	134	127
2nd grade	94	119	112
3rd grade	90	116	109
4th grade	85	116	107
5th grade	87	111	105
Total	463	596	560
Race/ethnicity (%)			
Black	9.5	12.5	11.7
White	1.9	5.3	4.4
Hispanic	88.4	80.4	82.6
Asian	0.2	1.7	1.3
Other	0.0	0.0	0.0
Gender (%)			
Male	52.2	51.8	51.9
Female	47.8	48.2	48.1
Classified as English speaker of other language (ESOL) (%)	78.2	79.7	79.3
Classified for special education (%)	6.7	6.2	6.3
Eligible for free or reduced-price lunch (%)	98.7	98.1	98.3
Characteristics of 3rd-grade students			
Overage for grade[a] (%)	40.4	35.2	36.6
Attendance rate[b] (%)	89.8	88.0	88.5
TAAS test scores			
Reading total			
Texas Learning Index (TLI) score	70.6	70.1	70.3
Pass rate	60.6	58.9	59.4
Math total			
Texas Learning Index (TLI) score	60.2	62.2	61.7
Pass rate	32.0	39.4	37.4

(continued)

Appendix Table B.1 (continued)

Characteristic	Project GRAD Schools	Comparison Schools	All Project GRAD and Comparison Schools in the District
Characteristics of 5th-grade students			
Overage for grade[a] (%)	50.0	43.4	45.2
Attendance rate[b] (%)	90.2	89.2	89.5
TAAS pretest scores			
Reading total			
Texas Learning Index (TLI) score	74.8	77.0	76.2
Pass rate	44.2	36.9	38.9
Math total			
Texas Learning Index (TLI) score	73.1	75.5	74.7
Pass rate	53.7	47.8	49.4
TAAS test scores			
Reading total			
Texas Learning Index (TLI) score	72.0	71.6	71.7
Pass rate	59.9	59.4	59.5
Math total			
Texas Learning Index (TLI) score	66.0	65.0	65.3
Pass rate	49.3	44.8	46.0
Total number of schools	6	16	22

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: All values are for the pre-Project GRAD school years 1997-1998 to 1999-2000.

Clusters consist of a Project GRAD school matched with a group of between 3 and 8 comparison schools. Results in the Comparison Schools columns reflect averages across these groups of non-Project GRAD schools.

[a]A student is defined as overage for grade if he or she turns 9 before the start of the 3rd grade or 11 before the start of the 5th grade.

[b]Attendance rate is calculated by dividing the number of days present by 180.

The Project GRAD Evaluation

Appendix Table B.2

Characteristics of Yates Feeder Pattern Elementary Schools, 1993-1994 Through 1995-1996

Characteristic	Project GRAD Schools	Comparison Schools	All Project GRAD and Comparison Schools in the District
Average school size			
1st grade	90	117	110
2nd grade	86	105	100
3rd grade	83	101	96
4th grade	76	94	90
5th grade	73	92	87
Total	408	509	483
Race/ethnicity (%)			
Black	85.7	80.1	81.6
White	1.8	2.2	2.1
Hispanic	11.7	16.1	14.9
Asian	0.8	1.6	1.4
Other	0.0	0.1	0.1
Gender (%)			
Male	51.3	51.6	51.5
Female	48.7	48.4	48.5
Classified as English speaker of other language (ESOL) (%)	55.0	57.0	56.5
Classified for special education (%)	12.9	11.5	11.9
Eligible for free or reduced-price lunch (%)	94.3	93.5	93.7
Characteristics of 3rd-grade students			
Overage for grade[a] (%)	28.2	27.6	27.8
Attendance rate[b] (%)	90.4	88.2	88.7
TAAS test scores			
Reading total			
Texas Learning Index (TLI) score	72.0	73.0	72.8
Pass rate	64.5	67.0	66.4
Math total			
Texas Learning Index (TLI) score	66.1	67.4	67.1
Pass rate	51.3	55.5	54.4

(continued)

Appendix Table B.2 (continued)

Characteristic	Project GRAD Schools	Comparison Schools	All Project GRAD and Comparison Schools in the District
Characteristics of 5th-grade students			
Overage for grade[a] (%)	38.8	37.0	37.5
Attendance rate[b] (%)	91.3	89.3	89.8
TAAS pretest scores			
Reading total			
Texas Learning Index (TLI) score	73.6	72.3	72.6
Pass rate	55.7	51.8	52.8
Math total			
Texas Learning Index (TLI) score	64.8	64.5	64.6
Pass rate	45.2	42.2	43.0
TAAS test scores			
Reading total			
Texas Learning Index (TLI) score	73.9	73.6	73.7
Pass rate	65.2	64.7	64.9
Math total			
Texas Learning Index (TLI) score	67.1	67.3	67.2
Pass rate	51.2	51.1	51.1
Total number of schools	11	31	42

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: All values are for the pre-Project GRAD school years 1997-1998 to 1999-2000.

Clusters consist of a Project GRAD school matched with a group of between 4 and 15 comparison schools. Results in the Comparisons Schools columns reflect averages across these groups of non-Project GRAD schools.

[a]A student is defined as overage for grade if he or she turns 9 before the start of the 3rd grade or 11 before the start of the 5th grade.

[b]Attendance rate is calculated by dividing the number of days present by 180.

The Project GRAD Evaluation

Appendix Table B.3

Characteristics of Wheatley Feeder Pattern Elementary Schools, 1996-1997 Through 1998-1999

Characteristic	Project GRAD Schools	Comparison Schools	All Project GRAD and Comparison Schools in the District
Average school size			
1st grade	85	131	125
2nd grade	78	120	115
3rd grade	72	112	107
4th grade	69	109	104
5th grade	64	104	99
Total	369	576	550
Race/ethnicity (%)			
Black	45.0	40.5	41.1
White	1.3	6.3	5.7
Hispanic	53.1	51.0	51.3
Asian	0.5	2.1	1.9
Other	0.1	0.1	0.1
Gender (%)			
Male	51.7	51.0	51.1
Female	48.3	49.0	48.9
Classified as English speaker of other language (ESOL) (%)	58.0	60.3	60.1
Classified for special education (%)	11.3	10.4	10.5
Eligible for free or reduced-price lunch (%)	93.4	85.0	86.1
Characteristics of 3rd-grade students			
Overage for grade[a] (%)	27.8	21.3	22.1
Attendance rate[b] (%)	87.5	85.8	86.0
TAAS test scores			
Reading total			
Texas Learning Index (TLI) score	77.2	78.3	78.2
Pass rate	74.7	77.2	76.9
Math total			
Texas Learning Index (TLI) score	72.3	73.4	73.2
Pass rate	65.9	70.5	70.0

(continued)

Appendix Table B.3 (continued)

Characteristic	Project GRAD Schools	Comparison Schools	All Project GRAD and Comparison Schools in the District
Characteristics of 5th-grade students			
Overage for grade[a] (%)	34.4	28.3	29.0
Attendance rate[b] (%)	88.1	86.5	86.7
TAAS pretest scores			
Reading total			
Texas Learning Index (TLI) score	81.1	79.3	79.5
Pass rate	80.4	76.5	77.0
Math total			
Texas Learning Index (TLI) score	77.4	75.9	76.1
Pass rate	78.3	73.5	74.1
TAAS test scores			
Reading total			
Texas Learning Index (TLI) score	79.8	81.0	80.9
Pass rate	75.5	77.0	76.8
Math total			
Texas Learning Index (TLI) score	77.5	78.5	78.3
Pass rate	76.4	78.9	78.6
Total number of schools	12	85	97

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: All values are for the pre-Project GRAD school years 1997-1998 to 1999-2000.

Clusters consist of a Project GRAD school matched with a group of between 2 and 19 comparison schools. Results in the Comparison Schools columns reflect averages across these groups of non-Project GRAD schools.

[a]A student is defined as overage for grade if he or she turns 9 before the start of the 3rd grade or 11 before the start of the 5th grade.

[b]Attendance rate is calculated by dividing the number of days present by 180.

The Project GRAD Evaluation

Appendix Table B.4

Characteristics of Nonselective Comprehensive Elementary Schools in the Newark, New Jersey, School District, 1995-1996 to 1997-1998

Characteristic	Project GRAD Schools	Comparison Schools	All Newark Elementary Schools in the District
Average school size			
1st grade	88	75	70
2nd grade	80	64	67
3rd grade	72	57	62
Total	241	196	199
Race/ethnicity (%)			
Black	87.7	94.0	69.6
White	0.4	0.0	6.7
Hispanic	12.0	5.8	22.3
Asian	0.0	0.0	0.5
Other	0.0	0.2	0.8
Gender (%)			
Male	49.8	49.6	49.2
Female	50.2	50.4	50.8
Characteristics of 2nd-grade students			
Overage for grade[a] (%)	11.0	14.5	12.5
SAT-9 test scores			
Reading total			
Normal curve equivalent (NCE) score	45.0	43.5	46.5
At or above grade level (%)	39.6	37.6	42.9
Math total			
Normal curve equivalent (NCE) score	45.8	46.8	51.2
At or above grade level (%)	43.4	43.7	51.8
Characteristics of 3rd-grade students			
Overage for grade[a] (%)	14.8	18.9	15.8
SAT-9 test scores			
Reading total			
Normal curve equivalent (NCE) score	37.9	37.4	43.3
At or above grade level (%)	26.7	25.2	37.4
Math total			
Normal curve equivalent (NCE) score	43.6	42.2	48.7
At or above grade level (%)	39.0	34.6	47.4
Total number of schools	7	9	50

(continued)

Appendix Table B.4 (continued)

SOURCE: MDRC calculations from individual student school records from Newark Public Schools.

NOTES: All values are for the pre-Project GRAD school years 1995-1996 to 1997-1998.

Clusters consist of a Project GRAD school matched with a group of between 1 and 7 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

[a]A student is defined as overage for grade if he or she turns 8 before the start of the 2nd grade, 9 before the start of the 3rd grade, or 11 before the start of the 5th grade.

The Project GRAD Evaluation

Appendix Table B.5

Characteristics of Nonselective Comprehensive Elementary Schools in the Columbus, Ohio, School District, 1996-1997 to 1998-1999

Characteristic	Project GRAD Schools	Comparison Schools	All Columbus Elementary Schools in the District
Average school size			
1st grade	75	75	74
2nd grade	75	66	69
3rd grade	69	63	65
4th grade	66	58	61
5th grade	61	55	58
Total	346	317	327
Race/ethnicity (%)			
Black	85.6	83.1	53.4
White	10.7	13.3	41.2
Hispanic	0.7	0.8	1.1
Asian	1.1	0.8	2.2
Other	0.1	0.2	0.2
Gender (%)			
Male	52.0	51.7	51.9
Female	48.0	48.3	48.1
Classified as English speaker of other language (ESOL) (%)	2.5	1.1	1.9
Classified as disabled student (%)	10.3	9.8	10.8
Classified as economically or academically disadvantaged (%)	81.8	83.4	66.8
Characteristics of 2nd-grade students			
Overage for grade[a] (%)	19.5	18.2	17.4
Attendance rate[b] (%)	92.0	91.2	92.3
MAT test scores			
Reading comprehension			
Normal curve equivalent (NCE) score	35.1	36.2	41.9
At or above grade level (%)	17.5	20.7	31.9
Math total			
Normal curve equivalent (NCE) score	32.2	34.7	41.0
At or above grade level (%)	18.0	23.1	33.9

(continued)

Appendix Table B.5 (continued)

Characteristic	Project GRAD Schools	Comparison Schools	All Columbus Elementary Schools in the District
Characteristics of 3rd-grade students			
Overage for grade[a] (%)	22.8	19.5	18.3
Attendance rate[b] (%)	92.5	91.6	92.7
MAT pretest scores			
Reading comprehension			
Normal curve equivalent (NCE) score	34.5	36.3	42.5
Math total			
Normal curve equivalent (NCE) score	33.2	35.3	42.2
MAT test scores			
Reading comprehension			
Normal curve equivalent (NCE) score	37.5	37.4	44.5
At or above grade level (%)	20.5	21.2	35.4
Math total			
Normal curve equivalent (NCE) score	33.6	34.8	42.8
At or above grade level (%)	20.7	23.4	37.1
Characteristics of 5th-grade students			
Overage for grade[a] (%)	26.4	22.4	21.7
Attendance rate[b] (%)	92.3	92.1	92.4
MAT pretest scores			
Reading comprehension			
Normal curve equivalent (NCE) score	39.4	39.3	45.7
Math total			
Normal curve equivalent (NCE) score	38.2	37.7	45.4
MAT test scores			
Reading comprehension			
Normal curve equivalent (NCE) score	40.4	41.2	47.1
At or above grade level (%)	24.6	28.8	41.8
Math total			
Normal curve equivalent (NCE) score	35.3	34.9	42.8
At or above grade level (%)	20.5	19.1	34.1
Total number of schools	7	19	88

(continued)

Appendix Table B.5 (continued)

SOURCE: MDRC calculations from individual student school records from Columbus Public Schools.

NOTES: All values are for the pre-Project GRAD school years 1996-1997 to 1998-1999.

Clusters consist of a Project GRAD school matched with a group of between 6 and 7 comparison schools. Results in the Comparison Schools column reflects averages across these groups of non-Project GRAD schools.

[a]A student is defined as overage for grade if he or she turns 8 before the start of the 2nd grade, 9 before the start of the 3rd grade, or 11 before the start of the 5th grade.

[b]Attendance rate is calculated by dividing the number of days present by the number of days enrolled.

Appendix Table B.6

Characteristics of Nonselective Comprehensive Elementary Schools in the Atlanta, Georgia, School District, 1997-1998 to 1999-2000

Characteristic	Project GRAD Schools	Comparison Schools	All Atlanta Elementary Schools in the District
Average school size			
1st grade	84	94	93
2nd grade	85	90	89
3rd grade	85	88	87
4th grade	77	79	78
5th grade	73	73	75
Total	403	425	421
Race/ethnicity (%)			
Black	97.7	98.1	87.9
White	1.5	0.5	7.5
Hispanic	0.6	1.2	3.1
Asian	0.1	0.1	1.0
Other	0.0	0.0	0.1
Gender (%)			
Male	48.7	50.2	50.2
Female	51.3	49.8	49.8
Classified as English speaker of other language (ESOL) (%)	0.3	0.4	1.9
Classified for special education (%)	1.1	0.6	0.6
Eligible for free or reduced-price lunch (%)	84.0	85.1	77.6
Characteristics of 3rd-grade students			
Overage for grade[a] (%)	0.0	0.0	0.0
Attendance rate[b] (%)	82.3	81.1	82.5
ITBS pretest scores			
Reading total			
Normal curve equivalent (NCE) score	55.1	54.3	55.6
Math total			
Normal curve equivalent (NCE) score	60.7	59.7	61.5

(continued)

Appendix Table B.6 (continued)

Characteristic	Project GRAD Schools	Comparison Schools	All Atlanta Elementary Schools in the District
ITBS test scores			
Reading total			
Normal curve equivalent (NCE) score	46.3	46.3	47.5
At or above grade level (%)	42.0	44.7	45.8
Math total			
Normal curve equivalent (NCE) score	48.0	48.3	50.7
At or above grade level (%)	46.7	45.9	49.8
Characteristics of 5th-grade students			
Overage for grade[a] (%)	17.1	18.9	20.2
Attendance rate[b] (%)	83.9	83.2	84.1
ITBS pretest scores			
Reading total			
Normal curve equivalent (NCE) score	45.4	49.2	49.8
Math total			
Normal curve equivalent (NCE) score	50.8	52.0	54.0
ITBS test scores			
Reading total			
Normal curve equivalent (NCE) score	44.4	47.8	48.4
At or above grade level (%)	39.6	46.1	47.6
Math total			
Normal curve equivalent (NCE) score	45.3	45.6	48.0
At or above grade level (%)	43.2	42.8	47.0
Total number of schools	9	26	68

SOURCE: MDRC calculations from individual student school records from Atlanta Public Schools.

NOTES: All values are for the pre-Project GRAD school years 1997-1998 to 1999-2000.

Clusters consist of a Project GRAD school matched with a group of between 3 and 12 comparison schools. Results in the Comparison Schools column reflect averages across these groups of non-Project GRAD schools.

[a]A student is defined as overage for grade if he or she turns 9 before the start of the 3rd grade or 11 before the start of the 5th grade.

[b]Attendance rate is calculated by dividing the number of days present by 180.

Appendix C

The Cumulative Effects of Project GRAD on Elementary Student Achievement

The analysis presented in the body of this report focuses on the impact of Project Graduation Really Achieves Dreams (GRAD) on student achievement, that is, Project GRAD's effect on elementary student achievement, *over and above what would have happened in the absence of the program.* In addition to focusing on Project GRAD's effects over and above whatever interventions were occurring at the comparison schools, much of MDRC's analysis focuses on what can be called the *incremental* effect of Project GRAD in each grade. For most grades, this means that the estimates reflect only the effect of Project GRAD over and above the effects in any previous grades. For example, in Project GRAD Houston, the impact estimates for third-grade achievement controlled for such student characteristics as race/ethnicity and eligibility for free or reduced-price lunch. The estimates did not control, however, for each student's prior achievement, inasmuch as third grade was the earliest grade for which achievement data were available.

On the other hand, estimated effects on student achievement in the fourth and fifth grades did control for prior achievement. For fourth-graders, impacts on achievement were estimated controlling for each student's third-grade achievement. Among fifth-grade students, program effects were estimated using regression controls for each student's fourth-grade achievement level. Therefore, while the estimated effects on third-grade achievement represent the cumulative impact of Project GRAD through the third grade, the estimated impacts in the fourth and fifth grades each represent the incremental effect of the program on achievement between that grade and the previous grade, that is, the incremental effect of the program for that particular grade.

Though unlikely, some may argue that these effects underestimate the cumulative effect. In other words, it is possible, for example, for incremental effects in the fourth and fifth grades to be statistically distinguishable from zero but for the cumulative effect through the fifth grade to be positive.

In order to ascertain whether or not this is the case, MDRC undertook several types of sensitivity analyses. First, the program effects among fifth-graders in all the sites were reestimated, taking out all statistical controls. In addition to testing the sensitivity of the results to specification, this provides estimates of *cumulative* program effects.

Appendix Table C.1 compares "unadjusted" and "adjusted" impact effect sizes for the highest grade available in each feeder pattern. In particular, the table presents the program impacts among fifth-grade students in the three Project GRAD Houston feeder patterns, among third-grade students in Project GRAD Newark, and among fourth-grade students in Atlanta and Columbus. The program impacts are divided by the standard deviation of the outcome in question, in order to translate them into an "effect size" metric. The rows of the table present the

The Project GRAD Evaluation

Appendix Table C.1

Adjusted Versus Unadjusted ITS[a] Impact Effect Sizes on Test Scores, Follow-Up Results, by Feeder Pattern

	Adjusted Versus Unadjusted Achievement Outcomes							
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Houston, Davis: 5th grade								
Math TAAS								
Adjusted	-0.28 ***	-0.12	0.13	0.19	-0.29 ***	-0.04	0.06	-0.04
Unadjusted	-0.21 *	-0.17	0.06	0.27 **	-0.23 *	-0.12	0.11	0.06
Reading TAAS								
Adjusted	-0.08	0.08	0.22 *	0.41 ***	-0.23 *	0.07	0.20 *	0.14
Unadjusted	-0.07	-0.04	0.06	0.43 ***	-0.21	-0.10	0.12	0.09
Houston, Yates: 5th grade								
Math TAAS								
Adjusted	0.02	0.07	-0.03	0.00	0.02	-0.07		
Unadjusted	0.01	-0.01	0.07	-0.03	-0.07	-0.08		
Reading TAAS								
Adjusted	0.04	0.10	-0.10	-0.03	0.01	-0.09		
Unadjusted	0.01	-0.01	0.02	-0.06	-0.07	-0.17		
Houston, Wheatley: 5th grade								
Math TAAS								
Adjusted	0.17 **	0.11	0.10					
Unadjusted	0.09	0.09	0.12 *					
Reading TAAS								
Adjusted	0.08	0.11	0.15 *					
Unadjusted	-0.01	0.09	0.14					

(continued)

Appendix Table C.1 (continued)

	Adjusted Versus Unadjusted Achievement Outcomes							
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Newark: 3rd grade								
Math SAT-9								
Adjusted	0.44 **	0.35 *						
Unadjusted	0.33 *	0.34 *						
Reading SAT-9								
Adjusted	0.37 *	0.24						
Unadjusted	0.29 *	0.21						
Columbus: 5th grade								
Math MAT								
Adjusted	-0.02	-0.14 *	-0.22 **	-0.19 **				
Unadjusted	-0.06	-0.13	-0.35 **	-0.25 **				
Reading MAT								
Adjusted	0.01	-0.10	-0.07	-0.11				
Unadjusted	0.00	-0.07	-0.19 *	-0.20 *				
Atlanta: 4th grade								
Math CRCT								
Adjusted	-0.13	-0.19	-0.01					
Unadjusted	-0.16	-0.20	-0.02					
Reading CRCT								
Adjusted	-0.13	-0.17	0.00					
Unadjusted	-0.15	-0.22	-0.05					

(continued)

Appendix Table C.1 (continued)

SOURCES: MDRC calculations from individual student school records from the Houston Independent School District, Newark Public Schools, Columbus Public Schools, and Atlanta Public Schools.

NOTES: Texas Assessment of Academic Skills (TAAS); Stanford Achievement Test, versions 8 and 9 (SAT); Metropolitan Achievement Test (MAT); Criterion-Referenced Competency Test (CRCT).

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

[a]Impact estimates generated by comparative interrupted time series (ITS) analysis are described in detail in Chapter 3.

original impact effect sizes, controlling for prior achievement and other background characteristics. The rows below the "adjusted" impacts present impact effect sizes for "unadjusted" program effects, which do not control for any student characteristics, including prior achievement.

The results in this table indicate that the adjusted and unadjusted program effects are consistent with one another. The unadjusted impact effect sizes are *not* generally larger than the adjusted impacts, suggesting that the original, adjusted impacts that are presented in the body of the report are not somehow masking larger cumulative program effects.

The estimates provided in Appendix Table C.1 do not, however, account for the possibility that program effects are concentrated among students who receive more exposure to Project GRAD during their elementary school years. Given the high rates of student mobility, fifth-grade samples are likely to be made up of a combination of a smaller number students who have been in the same elementary school since the first grade and a larger number of students who have been in several different elementary schools during the first grade. If there are cumulative effects to exposure to Project GRAD, then effects would be larger among students who were exposed to Project GRAD for a longer period of time.

To explore this idea, MDRC identified subsamples of Project GRAD students who had attended the same Project GRAD elementary school long enough to get what might be considered a "higher dose" of the program's interventions. In Houston, within the Project GRAD Yates feeder pattern, data were available for fifth-grade students; a subsample was selected of fifth-graders who had been in the same elementary school for five consecutive school years. In Columbus, since fourth grade was the highest elementary grade for which data were available, a subsample was selected of fourth-graders who had been in the same elementary school for four consecutive school years.[1] Impact estimates among these "nonmobile" subgroups can shed light on whether the patterns of effects appear to be different among Project GRAD students who received more exposure to the program. The nonmobile students from the comparison schools provided a suitable counterpart, that is, a subset of non-Project GRAD students who were similarly stable. By comparing these students with similarly nonmobile students from the comparison schools, these analyses control for some of the bias that is introduced by concentrating on a select group of Project GRAD students.[2]

---

[1]Because this is an effort to identify the effects of Project GRAD on students who had been stable over time and who had been exposed to the program for several years, this analysis is limited to the feeder patterns that had at least four years of both follow-up and baseline data. This process established a baseline among stable students so that achievement patterns among those who had been exposed to Project GRAD for several years could be examined.

[2]To the extent that Project GRAD had an impact on student mobility or on the composition of students who stayed at or left Project GRAD elementary schools, the interpretation of these comparisons as program effects is undermined.

As in the study's primary analysis, the deviations from baseline achievement patterns among the nonmobile subsamples at Project GRAD schools were compared with the deviations from baseline patterns among similarly nonmobile students at the comparison schools. However, in order to allow for the cumulative effect of Project GRAD, these analyses do *not* control for individual student's prior achievement.

Appendix Table C.2 reports program impacts among both mobile and nonmobile students in the Yates and the Columbus feeder patterns. The first two rows of each panel present results for nonmobile students, that is, the students with longer exposure to Project GRAD or to the particular comparison school that they attended. These impacts have been divided by the standard deviation of the outcome in question, in order to translate them into an "effect size" metric. In the Yates feeder pattern, the impact estimates suggest no positive effects over and above the changes that occurred at the comparison schools. This is the case even in Year 5 and beyond, when the Project GRAD students had been exposed to the program for their entire elementary experience. The results in Project GRAD Columbus show a similar pattern: no positive effects among the nonmobile fourth-graders, even in Year 4 of the intervention, when they had been exposed to Project GRAD since entering elementary school.

Appendix Table C.2 also presents results for "mobile" students, that is, students who had *not* been in the same elementary school for as long as the nonmobile group. The results for these subgroups are not substantially different from one another. In sum, these results do not suggest the existence of any substantial cumulative effects that were not detected by the primary analysis presented in this report.

The Project GRAD Evaluation

Appendix Table C.2

Impact Effect Sizes for Mobile and Nonmobile Students, Not Controlling for Pretest, Houston (Yates) and Columbus Feeder Patterns

	Achievement Outcomes						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
Houston, Yates: 5th grade							
Nonmobile students							
Math - TAAS pass rates	0.07	-0.03	0.03	-0.02	-0.06	-0.02	-0.26 **
Reading - TAAS pass rates	0.08	-0.07	0.02	-0.02	-0.04	-0.06	-0.39 ***
Mobile students							
Math - TAAS pass rates	-0.04	0.03	0.01	-0.02	0.05	-0.07	-0.22 **
Reading - TAAS pass rates	0.04	0.08	0.09	0.07	-0.07	-0.17	-0.29 **
Columbus: 4th grade							
Nonmobile students							
Math - OPT pass rates	-0.08	-0.10	-0.04	-0.13			
Reading - OPT pass rates	-0.15	-0.08	0.08	0.08			
Mobile students							
Math - OPT pass rates	-0.17	-0.09	-0.10	-0.07			
Reading - OPT pass rates	0.03	0.04	0.03	0.09			

SOURCES: MDRC calculations from individual student school records from the Houston Independent School District and Columbus Public Schools.

NOTE: Texas Assessment of Academic Skills (TAAS), Ohio Proficiency Test (OPT).

Appendix D

The Impacts of Project GRAD on Elementary Student Promotion Rates

In addition to determining whether Project Graduation Really Achieves Dreams (GRAD) affected elementary students' achievement, MDRC's analysis investigated whether Project GRAD increased student promotion rates over and above what would have been observed in the absence of the intervention.

Promotion rates are based on student-level enrollment data provided to MDRC by the districts participating in the study. In particular, MDRC received annual enrollment files from each district in the study. Using district identification numbers and other information, MDRC identified the individual students enrolled in each school and the grade in which they were enrolled, for every year for which data were available. Students were classified as "promoted" in a given school year if they were enrolled in the subsequent grade during the following school year (regardless of whether or not they were in the same school). Students who were enrolled in the same grade in the following year were classified as "not promoted." Promotion status for students who could not be found in the district's data for the following school year was classified as "missing."

As in analyzing student achievement, MDRC's analysis of promotion rates compares the average rates during the years immediately preceding Project GRAD's implementation and the average rates in each subsequent year. These changes over time at the Project GRAD schools are then compared with the changes over time at the set of comparison schools that were originally chosen for the achievement analysis.

Impact Estimates

Appendix Tables D.1 through D.5 present the estimated impacts on promotion rates in the Houston, Atlanta, and Columbus Project GRAD sites.[1] Appendix Tables D.1 through D.3 present the impact estimates for the three feeder patterns in Project GRAD Houston: Davis, Yates, and Wheatley, respectively. Appendix Table D.4 presents the impact estimates for Project GRAD Atlanta, and Appendix Table D.5 reports the promotion impacts for Project GRAD Columbus. Overall, the results do not suggest that Project GRAD had systematic effects on student promotion rates. In particular, in most sites, changes in promotion rates at Project GRAD and at the comparison schools are similar. The exception is Project GRAD Atlanta, which appears to have had a positive effect on promotion to the third, fourth, and fifth grades in the first year of the program; this effect appears to have dissipated in the second year of implementation (Appendix Table D.4).

The other sites show no consistent patterns of positive or negative effects on promotion rates. Although there were occasional positive or negative effects in particular school years, there were no sustained differences in one direction or the other.

[1]Unfortunately, this analysis could not be performed for Project GRAD Newark because the necessary student-level enrollment data were not available.

The Project GRAD Evaluation

Appendix Table D.1

Interrupted Time Series and Impact Estimates for Promotion, Eight-Year Follow-Up Results, Davis Feeder Pattern

	I. Outcome Level Compared with Baseline Year and Follow-Up Years									
	Project GRAD Schools					Comparison Schools				
Outcome	Baseline	Year 1	Year 2	Year 3	Year 4	Baseline	Year 1	Year 2	Year 3	Year 4
Promoted to 4th grade (%)	88.4	90.9	85.8	87.3	83.2	84.1	85.7	83.4	83.9	81.7
Deviation from baseline average		2.6	-2.5	-1.0	-5.2		1.6	-0.7	-0.1	-2.4
Promoted to 5th grade (%)	87.3	88.6	92.3	90.5	90.4	83.9	89.0	86.7	89.6	85.3
Deviation from baseline average		1.2	5.0	3.2	3.1		5.2	2.8	5.7	1.5
Promoted to 6th grade (%)	88.2	88.5	84.0	85.9	83.7	86.1	84.5	86.5	83.4	78.6
Deviation from baseline average		0.3	-4.3	-2.4	-4.6		-1.6	0.4	-2.7	-7.5

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools				*III. Impact Effect Sizes* [a]			
	Year 1	Year 2	Year 3	Year 4	Year 1	Year 2	Year 3	Year 4
Promoted to 4th grade (%)								
Deviation from baseline average	1.0	-1.9	-0.9	-2.8	0.03	-0.05	-0.02	-0.08
Promoted to 5th grade (%)								
Deviation from baseline average	-3.9	2.2	-2.5	1.6	-0.11	0.06	-0.07	0.05
Promoted to 6th grade (%)								
Deviation from baseline average	1.9	-4.7	0.3	2.9	0.05	-0.13	0.01	0.08

(continued)

Appendix Table D.1 (continued)

	I. Outcome Level Compared with Baseline Year and Follow-Up Years									
	Project GRAD Schools					Comparison Schools				
Outcome	Baseline	Year 5	Year 6	Year 7	Year 8	Baseline	Year 5	Year 6	Year 7	Year 8
Promoted to 4th grade (%)	88.4	71.7	85.4	83.2	84.4	84.1	74.4	81.6	80.7	82.2
Deviation from baseline average		-16.7	-3.0	-5.1	-3.9		-9.7	-2.4	-3.4	-1.9
Promoted to 5th grade (%)	87.3	85.3	84.2	81.8	83.8	83.9	79.7	76.5	78.7	85.2
Deviation from baseline average		-2.0	-3.1	-5.6	-3.5		-4.2	-7.4	-5.1	1.3
Promoted to 6th grade (%)	88.2	84.4	79.3	84.4	83.1	86.1	79.3	79.2	78.8	81.2
Deviation from baseline average		-3.9	-8.9	-3.9	-5.1		-6.7	-6.8	-7.3	-4.8

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools				*III. Impact Effect Sizes* [a]			
	Year 5	Year 6	Year 7	Year 8	Year 5	Year 6	Year 7	Year 8
Promoted to 4th grade (%)								
Deviation from baseline average	-7.0 **	-0.5	-1.8	-2.0	-0.20 **	-0.02	-0.05	-0.06
Promoted to 5th grade (%)								
Deviation from baseline average	2.2	4.2	-0.4	-4.8 *	0.06	0.12	-0.01	-0.14 *
Promoted to 6th grade (%)								
Deviation from baseline average	2.8	-2.1	3.4	-0.3	0.08	-0.06	0.09	-0.01

(continued)

Appendix Table D.1 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average.

The TLI score is not available in Year 9 because the state test changed from the TAAS, which was issued in the prior years, to the TAKS in 2003.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 8 comparison schools. Results in the Comparison Schools columns reflect averages across these groups of comparison schools.

[a]The "impact effect size" was calculated by dividing the "impact" by the standard deviation from school years 1998-2000 of outcomes for all 3rd-grade students in the district's nonselective comprehensive elementary schools.

The Project GRAD Evaluation

Appendix Table D.2

Interrupted Time Series and Impact Estimates for Promotion, Six-Year Follow-Up Results, Yates Feeder Pattern

	I. Outcome Level Compared with Baseline Year and Follow-Up Years							
	Project GRAD Schools				Comparison Schools			
Outcome	Baseline	Year 1	Year 2	Year 3	Baseline	Year 1	Year 2	Year 3
Promoted to 4th grade (%)	87.5	86.8	84.9	77.5	85.3	81.6	82.5	74.6
Deviation from baseline average		-0.7	-2.6	-10.0		-3.7	-2.8	-10.7
Promoted to 5th grade (%)	90.6	89.5	89.4	85.3	87.6	84.6	84.8	81.7
Deviation from baseline average		-1.1	-1.2	-5.2		-3.0	-2.8	-5.8
Promoted to 6th grade (%)	86.8	86.5	85.8	82.7	86.5	85.8	84.7	83.1
Deviation from baseline average		-0.3	-0.9	-4.1		-0.7	-1.8	-3.4

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools			*III. Impact Effect Sizes* [a]		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
Promoted to 4th grade (%)						
Deviation from baseline average	3.0	0.2	0.7	0.08	0.00	0.02
Promoted to 5th grade (%)						
Deviation from baseline average	1.9	1.6	0.6	0.06	0.05	0.02
Promoted to 6th grade (%)						
Deviation from baseline average	0.4	0.9	-0.7	0.01	0.02	-0.02

(continued)

Appendix Table D.2 (continued)

	I. Outcome Level Compared with Baseline Year and Follow-Up Years							
	Project GRAD Schools				Comparison Schools			
Outcome	Baseline	Year 4	Year 5	Year 6	Baseline	Year 4	Year 5	Year 6
Promoted to 4th grade (%)	87.5	79.4	81.0	82.1	85.3	80.0	79.0	81.5
Deviation from baseline average		-8.1	-6.5	-5.4		-5.3	-6.3	-3.8
Promoted to 5th grade (%)	90.6	82.9	84.4	83.8	87.6	81.7	80.7	83.8
Deviation from baseline average		-7.6	-6.2	-6.8		-5.8	-6.8	-3.7
Promoted to 6th grade (%)	86.8	84.0	78.4	86.0	86.5	80.4	81.4	82.1
Deviation from baseline average		-2.8	-8.3	-0.8		-6.1	-5.1	-4.4

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools			*III. Impact Effect Sizes* [a]		
	Year 4	Year 5	Year 6	Year 4	Year 5	Year 6
Promoted to 4th grade (%)						
Deviation from baseline average	-2.8	-0.3	-1.5	-0.08	-0.01	-0.04
Promoted to 5th grade (%)						
Deviation from baseline average	-1.8	0.6	-3.0	-0.05	0.02	-0.09
Promoted to 6th grade (%)						
Deviation from baseline average	3.3 *	-3.3 *	3.6 **	0.09 *	-0.09 *	0.10 **

(continued)

Appendix Table D.2 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 4 and 15 comparison schools. Results in the Comparison Schools columns reflect averages across these groups of comparison schools.

[a]The "impact effect size" was calculated by dividing the "impact" by the standard deviation from school years 1998-2000 of outcomes for all 3rd-grade students in the district's nonselective comprehensive elementary schools.

The Project GRAD Evaluation

Appendix Table D.3

Interrupted Time Series and Impact Estimates for Promotion, Three-Year Follow-Up Results, Wheatley Feeder Pattern

	I. Outcome Level Compared with Baseline Year and Follow-Up Years							
	Project GRAD Schools				Comparison Schools			
Outcome	Baseline	Year 1	Year 2	Year 3	Baseline	Year 1	Year 2	Year 3
Promoted to 4th grade (%)	81.0	80.4	84.0	81.4	79.6	79.6	79.0	81.0
Deviation from baseline average		-0.5	3.0	0.4		-0.1	-0.6	1.3
Promoted to 5th grade (%)	86.8	79.7	82.7	87.0	84.4	79.1	80.0	83.3
Deviation from baseline average		-7.1	-4.1	0.3		-5.4	-4.4	-1.1
Promoted to 6th grade (%)	88.5	83.2	84.7	88.2	82.8	77.8	79.0	80.7
Deviation from baseline average		-5.3	-3.8	-0.3		-5.0	-3.8	-2.1

	II. Impact of Project GRAD Schools Compared with Impact of Comparison Schools			*III. Impact Effect Sizes* [a]		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
Promoted to 4th grade (%)						
Deviation from baseline average	-0.4	3.7	-0.9	-0.01	0.10	-0.03
Promoted to 5th grade (%)						
Deviation from baseline average	-1.7	0.3	1.4	-0.05	0.01	0.04
Promoted to 6th grade (%)						
Deviation from baseline average	-0.4	0.0	1.8	-0.01	0.00	0.05

(continued)

Appendix Table D.3 (continued)

SOURCE: MDRC calculations from individual student school records from the Houston Independent School District.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" for the comparison schools.

Estimates are regression-adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 2 and 19 comparison schools. Results in the Comparison Schools column reflect averages across these groups of comparison schools.

[a]The "impact effect size" was calculated by dividing the "impact" by the standard deviation from school years 1998-2000 of outcomes for all 3rd-grade students in the district's nonselective comprehensive elementary schools.

The Project GRAD Evaluation

Appendix Table D.4

Interrupted Time Series and Impact Estimates for Promotion, Two-Year Follow-Up Results, Atlanta Feeder Pattern

	Project GRAD Schools			Comparison Schools			Impact		Impact Effect Sizes[a]	
Outcome	Baseline	Year 1	Year 2	Baseline	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Promoted to 2nd grade[b] (%)	77.5	76.0	71.0	78.6	77.8	74.0				
Deviation from baseline average		-1.5	-6.5		-0.7	-4.6	-0.8	-1.9	-0.02	-0.05
Promoted to 3rd grade[c] (%)	77.6	81.0	77.3	77.5	76.3	74.6				
Deviation from baseline average		3.4	-0.3		-1.1	-2.9	4.6 *	2.6	0.11 *	0.06
Promoted to 4th grade[d] (%)	74.9	75.9	68.4	76.6	73.1	68.3				
Deviation from baseline average		1.0	-6.6		-3.5	-8.3	4.5	1.7	0.11	0.04
Promoted to 5th grade[e] (%)	81.5	84.2	75.5	81.7	78.0	72.7				
Deviation from baseline average		2.7	-6.0		-3.7	-9.0	6.5 **	3.0	0.17 **	0.08

(continued)

Appendix Table D.4 (continued)

SOURCE: MDRC calculations from individual student school records from Atlanta Public Schools.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" from the non-Project GRAD schools.

Estimates are regression adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 3 and 12 comparison schools. Results in the Comparison Schools columns reflect averages across these groups of comparison schools.

[a]The "impact effect size" was calculated by dividing the "impact" by the standard deviation from school years 1998-2000 of outcomes for all 4th-grade students in the district's nonselective comprehensive elementary schools.

[b]1st-grade students were considered promoted if they were listed as 2nd-graders in the next year's administrative data file. Students in this sample were on file in both their 1st- and 2nd-grade years.

[c]2nd-grade students were considered promoted if they were listed as 3rd-graders in the next year's administrative data file. Students in this sample were on file in both their 2nd- and 3rd-grade years.

[d]3rd-grade students were considered promoted if they were listed as 4th-graders in the next year's administrative data file. Students in this sample were on file in both their 3rd- and 4th-grade years.

[e]4th-grade students were considered promoted if they were listed as 5th-graders in the next year's administrative data file. Students in this sample were on file in both their 4th- and 5th-grade years.

The Project GRAD Evaluation

Appendix Table D.5

Interrupted Time Series and Impact Estimates for Promotion, Three-Year Follow-Up Results, Columbus Feeder Pattern

	I. Outcome Levels Compared with Baseline Year and Follow-Up Years							
	Project GRAD Schools				Comparison Schools			
Outcome	Baseline	Year 1	Year 2	Year 3	Baseline	Year 1	Year 2	Year 3
Promoted to 3rd grade[b] (%)	88.2	89.5	93.2	84.6	86.6	88.8	89.5	85.7
Deviation from baseline average		1.4	5.0	-3.6		2.2	2.9	-0.9
Promoted to 4th grade[c] (%)	91.3	94.4	95.1	88.7	89.4	92.5	91.9	89.7
Deviation from baseline average		3.1	3.8	-2.6		3.1	2.6	0.3
Promoted to 5th grade[d] (%)	91.3	93.3	96.0	89.4	91.3	93.5	92.0	87.1
Deviation from baseline average		2.0	4.6	-1.9		2.2	0.7	-4.3

	Impact of Project GRAD Schools Compared with Impact of Comparison Schools			*III. Impact Effect Sizes*[a]		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
Promoted to 3rd grade[b] (%)						
Deviation from baseline average	-0.9	2.1	-2.8	-0.02	0.06	-0.08
Promoted to 4th grade[c] (%)						
Deviation from baseline average	-0.1	1.2	-2.9	0.00	0.03	-0.08
Promoted to 5th grade[d] (%)						
Deviation from baseline average	-0.2	3.9 *	2.4	-0.01	0.12 *	0.07

(continued)

Appendix Table D.5 (continued)

SOURCE: MDRC calculations from individual student school records from Columbus Public Schools.

NOTES: The "deviation from the baseline" for Year 1 was calculated as the difference between the baseline average and the Year 1 average. The "deviation from the baseline" for Year 2 was calculated as the difference between the baseline average and the Year 2 average. The "deviation from the baseline" for Year 3 was calculated as the difference between the baseline average and the Year 3 average.

The "impact" was calculated as the difference between the "deviation from the baseline" for Project GRAD schools and the "deviation from the baseline" from the non-Project GRAD schools.

Estimates are regression-adjusted using ordinary least squares, controlling for students' background characteristics and prior achievement.

A two-tailed t-test was applied to differences in deviations from the baseline for Project GRAD and comparison schools. Standard errors and statistical significance levels of deviations from the baseline are adjusted to account for cohort effects. Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

Clusters consist of a Project GRAD school matched with a group of between 6 and 7 non-Project GRAD schools. Results in the Comparison Schools column reflects averages across these groups of non-Project GRAD schools.

[a]The "impact effect size" was calculated by dividing the "impact" by the standard deviation from school years 1997-1999 of outcomes for all 2nd-grade students in the district's nonselective comprehensive elementary schools.

[b]2nd-grade students were considered promoted if they were listed as 3rd-graders in the next year's administrative data file. Students in this sample were on file in both their 2nd- and 3rd-grade years.

[c]3rd-grade students were considered promoted if they were listed as 4th-graders in the next year's administrative data file. Students in this sample were on file in both their 3rd- and 4th-grade years.

[d]4th-grade students were considered promoted if they were listed as 5th-graders in the next year's administrative data file. Students in this sample were on file in both their 4th- and 5th-grade years.

References and Bibliography

Bifulco, Robert, William Duncombe, and John Yinger. 2005. "Does Whole-School Reform Boost Student Performance? The Case of New York City." *Journal of Policy Analysis and Management* 24, 1.

Bloom, Howard. 2003. "Using Short Interrupted Time-Series Analysis to Measure the Impacts of Whole School Reforms." *Evaluation Review* 27, 1: 3-49.

Bloom, Howard, JoAnn Rock, Sandra Ham, Laura Melton, and Julieanne O'Brien. 2001. *Evaluating the Accelerated Schools Approach: A Look at Early Implementation and Impacts on Student Achievement in Eight Elementary Schools*. New York: MDRC.

Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2002. *Comprehensive School Reform and Student Achievement: A Meta-Analysis*. Baltimore: Center for Research on the Education of Students Placed At Risk (CRESPAR).

Center on Educational Policy. 2004. *State High School Exams: A Maturing Reform.* Washington, DC: Center on Educational Policy.

Community Training and Assistance Center. 2000. *Myths and Realities: The Impact of the State Takeover on Students and Schools in Newark.* Boston: Community Training and Assistance Center.

Haney, Walter. 2000. "The Myth of the Texas Miracle in Education." *Education Policy Analysis Archives* 8, 41. Web site: http://epaaa.asu.edu/epaa/v8n41.

Haycock, Kati. 2001. "Closing the Achievement Gap." *Educational Leadership* 58, 6: 6-11.

Holland, Holly. 2005. *Whatever It Takes: Transforming American Schools: The Project GRAD Story.* New York: Teachers College Press.

Kemple, James J., Corinne M. Herlihy, and Thomas J. Smith. 2005. *Making Progress Toward Graduation: Evidence from the Talent Development High School Model.* New York: MDRC.

Klein, Stephen P., Laura S. Hamilton, Daniel F. McCaffrey, and Brian M. Strecher. 2000. "What Do Test Scores in Texas Tell Us?" Issue Paper IP-202. RAND Corporation.

Koretz, Daniel. 2002. "Limitations in the Use of Achievement Tests as Measures of Educators' Productivity." *Journal of Human Resources* 37, 4: 752-777.

Lipsey, Mark. 1990. *Design Sensitivity: Statistical Power for Experimental Research.* Newbury Park, CA: Sage Publications.

McAdams, Donald R. 2000. *Fighting to Save Our Urban Schools . . . and Winning! Lessons from Houston.* New York: Teachers College Press.

National Center for Education Statistics (NCES). 1999. "International Education Indicators, Upper Secondary Graduation Rates, 1999." Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Opuni, Kwame A. 1997. *Project GRAD: 1996-97 Program Evaluation Report*. Houston: Project GRAD USA.

Opuni, Kwame A. 1999. *Project GRAD: 1998-99 Program Evaluation Report.* Houston: Project GRAD USA.

Opuni, Kwame A. 2001. *Project GRAD: A Comprehensive School Reform Model: 1999-2000 Program Evaluation Report.* Houston: Project GRAD USA.

Opuni, Kwame A. 2005. *Project GRAD Newark: 2003-04 Program Evaluation Report: An Evaluation of the First Six Years*. Houston: Project GRAD USA.

Opuni, Kwame A., and Mary L. Ochoa. 2002a. *Project GRAD: A Comprehensive School Reform Model: Project GRAD Houston: 2001-2002 Program Evaluation Report.* Houston: Project GRAD USA.

Opuni, Kwame A., and Mary L. Ochoa. 2002b. *Project GRAD Houston: The Fine Arts Program: 2001-02 Program Evaluation Report.* Houston: Project GRAD USA.

Opuni, Kwame A., and Mary L. Ochoa. 2004. *Project GRAD: A Comprehensive School Reform Model: Project GRAD Houston: An Assessment of Conditions/Factors Affecting the Model's Implementation: 2003 Formative Evaluation Report.* Houston: Project GRAD USA.

Project GRAD USA. 2004a. "Project GRAD Works: Graduation Really Achieves Dreams." Houston: Project GRAD USA.

Project GRAD USA. 2004b. "Project GRAD: Working to Close the Academic Achievement Gap." Houston: Project GRAD USA.

Project GRAD USA: 2004c. *New Site Development Handbook.* Houston: Project GRAD USA.

Quint, Janet, Howard S. Bloom, Alison Rebeck Black, and LaFleur Stephens with Theresa M. Akey. 2005. *The Challenges of Scaling Up Educational Reform: Findings and Lessons from First Things First*. New York: MDRC.

Snipes, Jason. 2003. "Using Interrupted Time Series with Comparison Groups to Estimate the Effects of Educational Interventions on Student Achievement." Presented at the annual conference of the Association for Public Policy Analysis and Management.

Snipes, Jason, Fred Doolittle, and Corinne Herlihy. 2002. *Foundations for Success: Case Studies of How Urban School Districts Improve Student Achievement.* New York and Washington, DC: MDRC for the Council of the Great City Schools.

Snipes, Jason C., Glee Ivory Holton, Fred Doolittle, and Laura Sztejnberg. 2006. *Striving for Student Success: The Effect of Project GRAD on High School Student Outcomes in Three Urban School Districts.* New York: MDRC.

Texas Education Center. 2005. "Student Assessment Division: Frequency Distribution." Web site: http://www.tea.state.tx.us/student.assessment/reporting/freq/index.html.

MDRC Publications on Project GRAD

Striving for Student Success
The Effect of Project GRAD on High School Student Outcomes in Three Urban School Districts
2006. Jason C. Snipes, Glee Ivory Holton, Fred Doolittle, Laura Sztejnberg

Charting a Path to Graduation
The Effect of Project GRAD on Elementary School Student Outcomes in Four Urban School Districts
2006. Jason C. Snipes, Glee Ivory Holton, Fred Doolittle

Building the Foundation for Improved Student Performance
The Pre-Curricular Phase of Project GRAD Newark
2000. Fred Doolittle, Sandra Ham, Glee Ivory Holton, with Ana Maria Ventura and Rochanda Jackson

About MDRC

MDRC is a nonprofit, nonpartisan social and education policy research organization dedicated to learning what works to improve the well-being of low-income people. Through its research and the active communication of its findings, MDRC seeks to enhance the effectiveness of social and education policies and programs.

Founded in 1974 and located in New York City and Oakland, California, MDRC is best known for mounting rigorous, large-scale, real-world tests of new and existing policies and programs. Its projects are a mix of demonstrations (field tests of promising new program approaches) and evaluations of ongoing government and community initiatives. MDRC's staff bring an unusual combination of research and organizational experience to their work, providing expertise on the latest in qualitative and quantitative methods and on program design, development, implementation, and management. MDRC seeks to learn not just whether a program is effective but also how and why the program's effects occur. In addition, it tries to place each project's findings in the broader context of related research — in order to build knowledge about what works across the social and education policy fields. MDRC's findings, lessons, and best practices are proactively shared with a broad audience in the policy and practitioner community as well as with the general public and the media.

Over the years, MDRC has brought its unique approach to an ever-growing range of policy areas and target populations. Once known primarily for evaluations of state welfare-to-work programs, today MDRC is also studying public school reforms, employment programs for ex-offenders and people with disabilities, and programs to help low-income students succeed in college. MDRC's projects are organized into five areas:

- Promoting Family Well-Being and Child Development
- Improving Public Education
- Raising Academic Achievement and Persistence in College
- Supporting Low-Wage Workers and Communities
- Overcoming Barriers to Employment

Working in almost every state, all of the nation's largest cities, and Canada and the United Kingdom, MDRC conducts its projects in partnership with national, state, and local governments, public school systems, community organizations, and numerous private philanthropies.