This product is part of the RAND Corporation monograph series. RAND monographs present major research findings that address the challenges facing the public and private sectors. All RAND monographs undergo rigorous peer review to ensure high standards for research quality and objectivity.

# Ending Social Promotion Without Leaving Children Behind

## The Case of New York City

Jennifer Sloan McCombs, Sheila Nataraj Kirby,
and Louis T. Mariano, *Editors*

Prepared for the New York City Department of Education

RAND EDUCATION

The research described in this report was prepared for the New York City Department of Education and conducted within RAND Education, a unit of the RAND Corporation.

**Library of Congress Cataloging-in-Publication Data**

McCombs, Jennifer Sloan, 1970-
Ending social promotion without leaving children behind : the case of New York City / Jennifer Sloan McCombs, Sheila Nataraj Kirby, Louis T. Mariano.
p. cm.
Includes bibliographical references.
ISBN 978-0-8330-4778-6 (pbk. : alk. paper)
1. Promotion (School)—New York (State)—New York—Case Studies. 2. Grade repetition—New York (State)—New York—Case Studies. I. Kirby, Sheila Nataraj, 1946– II. Mariano, Louis T. III. Title.

LB3063.M285 2010
371.2'8097471—dc22

2009032280



*Cover design by Eileen Delson La Russo*



Published 2009 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

# Preface

Many states and districts are moving toward test-based requirements for promotion at key transitional points in students' schooling careers, thus ending the practice of "social promotion"—promoting students without regard for how much they have learned. The rationale for retention is that repetition of the grade will give students an additional year to master the academic content that they failed to master the previous year, and, thus, students should be less at risk for failure when they go on to the next grade. Opponents of grade retention argue that prior research has shown that grade retention disproportionately affects low-income and minority children and is associated with low self-esteem, problem behaviors, and an increased risk of dropping out of school.

In 2003–2004, the New York City Department of Education (NYCDOE) implemented a new promotion and retention policy for 3rd-grade students in New York City (NYC) public schools. The policy was extended to grade 5 in 2004–2005, grade 7 in 2005–2006, and grade 8 in 2008–2009. NYCDOE asked the RAND Corporation to conduct an independent longitudinal evaluation of the 5th-grade social promotion policy and to examine the outcomes for two cohorts of 3rd-grade students. This study—conducted between March 2006 and August 2009—examined (1) policy implementation, factors affecting implementation, and implementation progress over time; (2) the impact of the policy on student academic and socioemotional outcomes; and (3) the links between implementation and desired outcomes.

This monograph presents the results of the study. Two other reports (Marsh et al., 2009, and Xia and Kirby, 2009) document the results of two additional tasks that were part of the overall study. The first reviews lessons learned regarding the design and implementation of promotion policies in a selection of states and districts with promotion policies similar to that of NYC, and the second presents a detailed and comprehensive review of the literature on grade retention. All three publications should interest policymakers, practitioners, and researchers involved in designing, implementing, or studying interventions to improve outcomes for low-performing students.

This research was conducted by RAND Education, a unit of the RAND Corporation.

# Contents

# Figures

# Tables

# Summary

The decision to retain students in grade when they fail to meet promotion criteria on standardized tests remains a controversial and hotly debated policy, but the practice of "social promotion"—promoting students who fail to meet academic standards and requirements at the same pace as their peers—has come under increasing attack and criticism. The rationale behind retention policies is that repetition of the grade will give students an additional year to master the knowledge and skills needed for that grade and, thus, "students should be less at risk for failure when they go on to the next grade" (Shepard and Smith, 1990, p. 84). However, opponents of grade retention point out that research has shown that retention disproportionately affects low-income and minority children and is associated with low self-esteem, problem behaviors, and an increased risk of dropping out of school.

As part of an ambitious reform initiative, the New York City Department of Education (NYCDOE), the largest school district in the country, implemented a new promotion and retention policy for students in grade 3 in 2003–2004. The policy was extended to grade 5 in 2004–2005, grade 7 in 2005–2006, and grade 8 in 2008–2009. The policy bases promotion decisions on students' scores on the New York State assessments, which established four levels of performance:

- Level 4—exceeds the standards
- Level 3—meets all the standards
- Level 2—meets some of the standards or partially meets the standards
- Level 1—shows serious academic difficulties.

Under the NYCDOE promotion policy, general education students in the gateway grades are required to score at or above performance Level 2 on both the English language arts (ELA) and mathematics assessments in order to be promoted (i.e., "passing" the promotion benchmark). Performance at or above Level 3 is considered "proficient" under the No Child Left Behind Act (NCLB), a higher standard than the promotion benchmark.

The policy places considerable emphasis on identifying struggling students early, providing them with additional instructional time, and continuously monitoring their

progress. Students who have been identified as in need of services at the beginning of the school year (based on their performance on the previous year's assessments, teacher recommendations, or being previously retained in grade) are mandated to receive academic intervention services (AIS) in school. In addition, schools can offer a variety of out-of-school support services, including Saturday school (previously called Saturday Preparatory Academies). Students who fail to score Level 2 or higher on the mathematics or ELA assessments administered in the spring are offered several opportunities to meet the promotion standards and can be promoted based on (1) a review of a portfolio of their work in the spring, (2) performance on the summer standardized assessment, (3) a review of a portfolio of their work in August, or (4) an appeal process. Students who do not meet the standards when their portfolios are reviewed in the spring are required to enroll in the Summer Success Academy, which offers additional hours of intensive instruction in mathematics and ELA for several weeks in the summer.

## This Study Examined Several Aspects of NYC's Promotion Policy

NYCDOE asked the RAND Corporation to conduct an independent longitudinal evaluation of the 5th-grade promotion policy. This study—conducted between March 2006 and August 2009—examined several aspects of the policy: (1) factors affecting implementation and implementation progress over time; (2) the impact of the policy on student academic and socioemotional outcomes; and (3) the links between implementation, components of the policy, and desired outcomes. This monograph reports the results of the overall study and is one of three reports documenting the overall evaluation of NYC's 5th-grade promotion policy. The two other reports—Marsh et al. (2009) and Xia and Kirby (2009)—help situate the policy in the broader national context and in the context of prior literature on similar policies and their outcomes.

This monograph is one of only a few longitudinal studies that examine the implementation and effects of a promotion policy on both academic and nonacademic outcomes for a large group of students. Using both qualitative and quantitative data, it provides a rich portrait of the implementation and effects of a promotion policy in the largest public school system in the country. It also highlights the factors that enabled or hindered the implementation of New York City's (NYC's) promotion policy, and it uses data from NYC and other locales to draw lessons for the design and implementation of promotion policies.

In this monograph, we focus on two groups of low-performing students:

- The first group includes students who scored Level 1 or low Level 2 on the 4th-grade assessments or had been retained in grade. These students were identified as needing additional help and likely targeted for AIS in schools and offered services through Saturday programs or Saturday Preparatory Academies. We refer to this group as "students needing services."

- The second group consists of students who failed the spring assessments (i.e., scored Level 1 on at least one of the spring assessments) and so were "at risk of retention" in 5th grade. This second group was mandated to attend sessions in the summer through the Summer Success Academies and offered other chances to meet the promotion standards (through a portfolio review or summer assessments). We refer to this group as "students at risk of retention."

We tracked four cohorts of 5th-grade students: three cohorts (2004–2005, 2005–2006, and 2006–2007) that were subject to the policy and a comparison cohort (2003–2004) that was not held to the policy. To improve readability and make clear the distinction between the policy and comparison cohorts, we refer to the policy cohorts as the P1, P2, and P3 cohorts and the comparison cohort as the C0 cohort.

## The Study Used a Variety of Data and Methods

The study used a combination of quantitative and qualitative methods to examine the implementation and impact of NYC's 5th-grade promotion policy on the immediate and future academic and socioemotional outcomes of students.

*To understand the implementation of the policy,* we collected and analyzed data from 80 case studies of schools, Saturday Preparatory Academies, and Summer Success Academies; interviews with 16 leaders of organizations or agencies responsible for providing support to schools (or regions, prior to 2007); and census surveys of principals, administrators of Saturday Preparatory Academies and Summer Success Academies, and school AIS team leaders (weighted to adjust for nonresponse, where feasible). We also reviewed a variety of documents from NYCDOE and from city and state Web sites.

*To track trends in the performance of at-risk students,* we obtained and analyzed demographic and achievement data for four cohorts, each with approximately 60,000 5th-grade students,[1] until 2007–2008.

*To analyze the effectiveness of the various supports offered under the policy to eligible students in terms of 5th-grade and future outcomes*, we employed multiple analytic methods. First, we analyzed the links between various school instructional strategies and other implementation measures and conducted further analyses of the relationship between 5th-grade achievement outcomes and participation in out-of-school instructional services (Saturday Preparatory Academies and Summer Success Academies). To

---

1 We also tracked outcomes for two 3rd-grade cohorts held to the policy (2003–2004 and 2004–2005) and one comparison cohort (2002–2003).

estimate the effects of supportive interventions and of being retained in grade, we used two quasi-experimental methods to define plausible comparison groups that could be used as controls in our models. These models were coupled with a "difference-in-difference" approach that allowed us to estimate the effect of the promotion policy separately from other concurrent reform efforts.

We tracked the performance of students in different cohorts into the 6th and 7th grades. We estimated the effect of being mandated to attend the summer academy and being retained in grade on future outcomes using same-grade comparisons. (For example, we compared these students' outcomes in 6th grade with the 6th-grade performance of promoted students from their cohort, though retained students took the assessment one year later.) Such comparisons explicitly estimate the effect of the additional year of schooling that retained students receive.

*To understand differences in the socioemotional status of three groups of students held to the promotion policy*—at risk and retained, at risk but promoted, and not at risk—we surveyed students in our study cohorts in the fall of the 2006–2007, 2007–2008, and 2008–2009 school years regarding their attitudes and beliefs about school belonging, mathematics, and reading.

*To understand what was known about the design and implementation of promotion policies more generally,* we conducted interviews in 2006 and 2008 with officials from a sample of six states and 12 districts with policies similar to that of NYC. We analyzed the data to identify promising strategies and challenges and to draw lessons learned about the design and implementation of such policies.

## The Study Faced Two Challenges

The first major challenge stemmed from data constraints and a change in test administration that limited our ability to compare outcomes across cohorts and over time. In response to the requirements of NCLB, New York State began administering new state assessments in grades 3 through 8 in ELA and mathematics in spring 2006. This meant that we lacked comparable data over time for some cohorts in certain years, which restricted our ability to compare longitudinal outcomes over time for some cohorts in certain years.

The second challenge arose from the fact that the promotion policy was only one piece of a larger set of reforms adopted by the city and state—reforms that were set against the larger backdrop of the federal NCLB mandates regarding student proficiency and annual goals for school progress. These reforms were associated with substantial improvements in test scores in both NYC and the rest of the state during the course of our study. Across the board, between 2006 and 2008, in almost every grade, the percentage of students scoring at the proficiency level (Level 3 or higher) in NYC schools increased dramatically, while the percentage of students scoring Level 1 declined equally dramatically. This same trend was evident across the rest of the state as well. It was beyond the scope of this study to determine whether this upward trend is a result of true learning or other factors.[2]

To isolate the effect of the promotion policy, we adjusted for the estimated upward trajectory in overall test scores to determine the "net" effect of the promotion policy on student outcomes. If we have not been entirely successful in this adjustment, then our estimated effects still reflect some of these confounding influences.

## Findings Showed Improved Student Performance Under the Policy

### Few Students Were Retained Under the Promotion Policy

Overall, approximately 60,000 5th-grade students were held to the promotion policy in each of the three cohorts subject to the policy, accounting for 75–78 percent of students in each of the cohorts. Special education students, those classified as English-language learners for three or fewer years, and those in charter schools were exempt from the policy. Between 19 and 24 percent of 5th graders were categorized as needing services at the beginning of the school year. Among entering 5th graders, more students needed services in ELA than in mathematics but students were more likely to be retained for failing to meet promotion criteria in mathematics.

Few students were retained under the policy, and the percentage of retained students dropped over time. About 2–3 percent of students in the first two policy cohorts were retained in grade. By the P3 cohort, this number had fallen to 1 percent (about 600 students out of approximately 58,000 students in the cohort compared to more than 1,700 in the previous two cohorts).

### Capacity Limited Some Schools' Ability to Provide Services

Students needing services were not evenly distributed across schools; their numbers ranged from 0 to 80 percent of a school's 5th-grade students in 2005. Schools seemed to provide academic intervention services to as many students as they had the capacity to serve. Schools with high levels of need (those with larger percentages of in-need students) were less likely to serve all students needing services, compared to other schools. However, high-need schools were significantly more likely than low-need schools to

---

2 We relied on the assumption that all scale and level scores provided were reasonable and consistent measures of achievement over time. Technical documentation about the assessments indicates that level scores are equated within grade and across years. However, the number of questions a student had to answer to score Level 2 on the state assessment declined over time, which increased the probability that a student could "guess" his or her way into Level 2. For the grade levels in our study years (2006–2008), this probability remained quite low—less than 1 percent, with one exception (7th-grade ELA in 2008, which was 5 percent). In 2009, the probability of scoring Level 2 by "guessing" increased, but this occurred after the period of our study.

have additional resources for students, such as mathematics specialists, reading coaches, and mathematics coaches.

Schools provided additional services to students both during the school day and outside the school day. Across the three survey years (2006–2008), almost all schools reported serving students in need of services through small-group tutoring during the school day. Fewer schools were able to offer one-on-one tutoring during the school day (64–80 percent). Exploratory models found that being in a school that offered one-on-one tutoring in mathematics (to some or all in-need students) increased in-need students' probability of meeting promotion standards in mathematics. Results for ELA one-on-one tutoring were promising as well. While in-school services were valuable to students, our case-study visits suggest that services may not have been offered consistently when AIS providers were pulled away for other duties. Teaching test-taking strategies using professionally compiled test-preparation materials (such as those produced by Kaplan) or teacher-generated test-preparation materials was a focus to a moderate or great extent both during and outside the school day.

#### Supports Offered Under the Policy Helped Students Meet Promotion Standards

Our analyses found small, positive effects of supportive services on the achievement of in-need students during the 5th-grade year. In-need students in the P1 cohort increased their 5th-grade spring assessment scores by between 0.10 and 0.20 standard deviations in ELA and by less than 0.10 standard deviations in mathematics over their expected performance in the absence of the policy. However, for the small group of students entering 5th grade at low Level 1, we found that additional promotion-policy services during the school year had little effect on performance.

More frequent attendance at the Saturday Preparatory Academies did not appear to have an impact on ELA outcomes in the P2 and P3 cohorts. However, in mathematics, there was a small benefit to attending 14–15 sessions compared to 6–7 sessions (about 0.10 standard deviations).

Summer Success Academies under the promotion policy had smaller class sizes and were more structured than summer schools offered prior to the policy. In addition, they were accompanied by an additional potential motivator for students—the very real and near-term threat of being retained in grade. We found that, relative to summer school prior to the policy, the Summer Success Academy option offered in the first policy year was somewhat more effective in improving student performance in mathematics (about 0.20 standard deviations) but not in ELA. In addition, we found that more frequent attendance at Summer Success Academy was associated with a small increase in mathematics performance but had no effect on ELA achievement. Prior studies also reported that summer school appears to have more beneficial effects in mathematics than in reading.

### Positive Effects of Promotion-Policy Services Continued Into Later Grades

We examined how specific groups of low-performing students subject to the promotion policy performed in later grades relative to comparable groups of students. Overall, our estimates show small to moderate positive effects of components of the promotion policy in the 6th and 7th grades:

- *Small, positive effects of early identification and intervention.* Students in the P1 cohort who needed services at the beginning of 5th grade scored higher on the 7th-grade assessments than comparable students in the comparison cohort (by 0.10–0.20 standard deviations in ELA and mathematics).
- *Small, positive effects of summer school.* Compared to students who were just above the Level 2 cutoff on the spring assessment, students in the P1 cohort who scored just below the cutoff and were required to attend summer school scored somewhat higher in both subjects (0.10–0.15 standard deviations higher) in both the 6th and 7th grades.
- *Moderate, positive effects of an additional year of instruction due to retention.* Compared to the students in the C0 cohort, retained students in the P1 cohort did considerably better on the 7th-grade assessments (by 0.60 standard deviations in ELA and 0.40 standard deviations in mathematics). Our analyses that focused on the highest Level 1 retained students in the P1 and P2 cohorts (compared to their peers scoring immediately above the Level 2 cutpoint) also point to moderate, positive effects of retention over time (approximately 0.40 standard deviations in ELA and mathematics). The results imply that these students, after an additional year of 5th-grade instruction and intervention, would be expected to score well within the Level 2 range on the 7th-grade assessments.

### Retained Students Did Not Report Negative Socioemotional Effects

Our student surveys showed that retention did not have negative effects on students' sense of school belonging or confidence in mathematics and reading; retained students reported comparable or higher levels than those of their at-risk promoted peers. In addition, retained students reported a greater sense of school connectedness than at-risk promoted students and not-at-risk students, even three years after the retention decision. The mean differences were small but statistically significant. These results mirror what other studies have found.

### School Staff Tended to Be Positive About the Policy and Its Supports

Principals and teachers tended to be positive about many aspects of NYC's promotion policy, and approximately three-quarters of AIS leaders and principals agreed that the policy focused their schools' instruction efforts in a positive way and that it made parents and guardians more concerned about their child's progress. However, the majority of respondents thought that the promotion policy relied too heavily on state assessment scores and, interestingly, that the policy made it more difficult to retain students who should be retained but passed the test.

Across the three years, the majority of principals reported receiving useful support from their region or school support organization as part of the policy.

## Near-Term Benefits Hold Promise for the Possibility of Longer-Term Benefits

We found positive near-term benefits of NYC's promotion policy. Students affected by the 5th-grade promotion policy performed better than they would have in absence of the policy in the 5th grade and into 7th grade. In addition, the study found no negative effects of retention on students' sense of school belonging or confidence in mathematics and reading over time.

However, the effectiveness of the promotion policy will ultimately be judged by whether the benefits of the policy outweigh its costs in the long term. This question has two components. First, from an individual student's point of view, the question is whether these short-term gains will persist into high school and result in improved probability of graduation and higher proficiency at graduation. Of particular concern is that prior studies have shown that retained students have an increased probability of dropping out. Second, from a societal point of view, the question is whether the overall benefits associated with implementing a promotion policy and holding students back a year outweigh the costs. Eide and Goldhaber (2005), for example, focused strictly on the economic consequences of retention and concluded that the overall longer-term benefits from retention fall far short of covering the costs to society in terms of financing an extra year of education and loss of future earnings. While the NYC policy has not been in place long enough to address these larger, long-term questions, the near-term benefits found by the study hold the possibility of longer-term benefits as well.

## The Study Identified Lessons for the Design and Implementation of Promotion Policies

We analyzed NYC's experience along with results from our interviews with officials from states and districts with similar promotion policies to identify lessons and practical insights into implementation for those who have adopted or are considering changes to promotion and retention policies.

With respect to design, the National Research Council (Heubert and Hauser, 1999, p. 135) pointed out that the validity and fairness of test-based promotion decisions can be enhanced by identifying at-risk or struggling students early so they can be targeted for extra help, providing students with multiple opportunities to demonstrate their knowledge through repeated testing with alternative forms or other appropriate means, and considering other relevant information about individual students.

NYC's promotion policy follows these tenets: It emphasizes early identification of at-risk students and provides them with additional instructional services both in school and out of school, and it offers several opportunities and means for students to meet the promotion standards, including the summer assessment, the spring portfolio review, the August portfolio review, and an appeals process that allows other information to be taken into account.

Additional lessons from NYC and other states and districts regarding design and implementation include the following:

- Invest in building support and maintaining ongoing communication with parents and educators.
- Anticipate the need to handle requests for exceptions to the policy.
- Identify students early and expand support beyond the promotion gateway grades.
- Provide adequate professional development for teachers.
- Invest in monitoring implementation and effects.
- Link the policy to a broader set of supports for at-risk students.
- Provide adequate funding.[3]

NYC, for example, rolled out its policy in stages—starting with students in grade 3 and then expanding it to higher grades as it gained more experience with the policy. NYC emphasized the need for open communication with parents, including sending out letters to parents in nine different languages to overcome language barriers. In addition, the policy was linked to a broad set of supports for schools and students, and considerable funding for both teacher professional development and student supportive services was provided.

## Several Policy Recommendations Emerge from the Findings

Based on our findings, we offer some recommendations for policymakers and administrators in NYC at the city, district, and school levels. While targeted at NYC, these recommendations may be of use to other districts and states considering or implementing test-based promotion policies.

*Continue early identification of students and provision of academic intervention services.* Our findings suggest that the process of early identification and support helped students meet promotion standards and had positive effects on student achievement in

[3] See Marsh et al. (2009) for a full description of these and related findings.

future grades. Results from exploratory models suggest that the provision of one-on-one tutoring may be a particularly helpful form of AIS and should be continued and expanded when possible.

*Enable AIS providers to work consistently with students who need services.* Our qualitative findings suggest that the AIS provision was inconsistent because providers were pulled away for substitute teaching, lunch duty, or proctoring assessments. Teachers we interviewed considered AIS important and helpful for their students' academic growth, but they noted that AIS could be more effective if provided on a consistent basis.

*Consider the expected duration and participation when planning Saturday programs.* Principals now have the authority to choose whether to operate a Saturday program for their students. Our results suggest that students attending at least six to seven sessions typically have higher achievement outcomes. Thus, principals who decide to offer the program need to pay careful attention to ensuring more frequent student attendance to maximize the program's benefits.

*Continue to encourage struggling students to attend summer school.* Summer school attendance appears to have a positive relationship with achievement on the summer assessment, particularly in mathematics, and summer school may also have a positive impact on future achievement in grades 6 and 7. Our results suggest that these benefits accrue, in particular, for students close to the Level 2 cutoff. Thus, principals may want to encourage high Level 1 and low Level 2 students to attend summer school programs.

*Collect and use data on the interventions being provided to at-risk students.* Our study highlighted a few instructional strategies that appeared to hold promise for helping low-performing students. Under the current structure, principals have considerable autonomy over the supports provided to these students, and there is no centralized data-collection effort to track what individual students are receiving. However, it is important to collect and analyze these data to determine what works. The school data system implemented by NYC—the Achievement Reporting and Innovation System—may enable the collection of such data.

*Continue to monitor the effects of retention on students.* One of the most important questions regarding the effects of retention on students is whether short-term positive effects persist over the longer term and whether the policy is cost-effective, compared to alternatives. We could not answer these questions with our data, but they remain important topics for NYC to address in the future.

# Acknowledgments

We thank the sponsor of this work, the New York City Department of Education, and its staff for their support, provision of data, and constructive comments throughout the project. We particularly thank those directly involved in supporting the project—Jennifer Bell-Ellwanger, Thomas Gold, and Amy Marsman.

We owe an enormous debt of gratitude to our reviewers—Adam Gamoran (University of Wisconsin–Madison), Melissa Roderick (University of Chicago), and J.R. Lockwood (RAND)—for their thoughtful and constructive reviews of the draft. This monograph is considerably improved as a result of their suggestions and recommendations.

We are indebted to a wide range of study participants in New York City—the principals, administrators, academic intervention team members, teachers, students, NYCDOE officials, regional directors of intervention services, and school support organization officials—who participated in the study and generously shared their valuable time and insights with us. We are also grateful to officials from other states and districts who took the time to describe their experiences implementing test-based promotion policies.

We thank Kadidja Diallo and Adriana Villavicencio for their help with the case studies. In addition, this project benefited from the input of a number of our RAND colleagues—Susan Bodilly, Daniel McCaffrey, Amelia Haviland, Paco Martorell, Greg Ridgeway, Vicki Park, Jennifer Steele, and Jeffery Marshall. We also thank Laura Hamilton for her thoughtful and thorough review of the report. We are grateful to Amber Price for her help with the research and to Emily Bever and Luetta Pope for the help they provided throughout the study, particularly with the student survey.

We are deeply indebted to Lauren Skrabala, our editor, for her impeccable editing and extraordinary patience. She is the editor for the series of three RAND reports documenting the study and worked heroically to bring clarity to the writing—a task made difficult by the large number of authors involved—and to ensure consistency in terminology and usage across the three reports. The reports are greatly improved as a result. Any remaining errors are our own or a result of not listening to her advice.

## Abbreviations

| | |
|---|---|
| AIS | academic intervention services |
| AYP | adequate yearly progress |
| ELA | English language arts |
| ELL | English-language learner |
| ESEA | Elementary and Secondary Education Act |
| ESL | English as a second language |
| ESO | Empowerment Support Organization |
| GAMM | generalized additive mixed models |
| GPA | grade point average |
| IEP | Individualized Education Program |
| IRT | item response theory |
| K-S | Kolmogorov-Smirnov |
| LSO | Learning Support Organization |
| MSE | mean square error |
| NAEP | National Assessment of Educational Progress |
| NCLB | No Child Left Behind Act |
| NYC | New York City |
| NYCDOE | New York City Department of Education |
| NYSED | New York State Education Department |
| OLS | ordinary least squares |

| | |
|---|---|
| PIP | personal intervention plan |
| PK | prekindergarten |
| PSO | Partner Support Organization |
| RDD | regression discontinuity design |
| RDIS | regional director of instructional services |
| SES | socioeconomic status |
| SPA | Saturday Preparatory Academy |
| SSA | Summer Success Academy |
| SSO | school support organization |
| TIMSS | Third International Mathematics and Science Study |
| twang | Toolkit for Weighting and Analysis of Nonequivalent Groups |

CHAPTER ONE

# Introduction

*Sheila Nataraj Kirby, Jennifer Sloan McCombs, and Louis T. Mariano*

In an era emphasizing educational standards and accountability, many states and districts are moving toward test-based requirements for promotion at key transitional points in students' schooling careers, thus ending the practice of "social promotion"—promoting students who have failed to meet academic standards and requirements for their grade.[1] Retention, in contrast, is a practice that holds back students who have failing grades or who fail to meet promotion criteria, which are often linked to standardized assessments. These "test-based" promotion policies typically use standardized tests as the main criterion to make high-stakes decisions about whether a student should be promoted to the next grade.

The rationale behind promotion policies is that repetition of the grade will give students an additional year to master the academic content that they failed to master the previous year and that, "by catching up on prerequisite skills, students should be less at risk for failure when they go on to the next grade" (Shepard and Smith, 1990, p. 84). Indeed, opponents of social promotion argue that the practice creates many problems: It can frustrate unprepared but promoted students by placing them in grades in which they are not ready for the work; it sends the message to all students that they can get by without working hard, adversely affecting student motivation and effort; it requires teachers to deal with underprepared students while trying to teach those who are ready to learn; it gives parents a false sense of their children's progress; it leads employers to conclude that diplomas are meaningless; and it "dumps" poorly educated students into a society in which they are not prepared to perform (Hartke, 1999; Thompson and Cunningham, 2000). However, opponents of grade retention argue that prior research has shown that retention disproportionately affects low-income and minority children and is associated with low self-esteem, problem behaviors, and an increased risk of dropping out of school.

---

[1] As of 2005, 18 states had policies that specified that an assessment was to be used in determining student eligibility for promotion or retention, and several others permitted local authorities to establish promotion policies or consider specific criteria for promotion (Zinth, 2005). However, it is surprisingly difficult to identify the number of states and districts implementing test-based promotion policies because many states and districts change the content of policies over time and sometimes decide to end the policy altogether (Marsh et al., 2009).

Test-based promotion policies are often implemented with additional short-term remedial supports, such as after-school programs, summer schools, and, sometimes, early identification of and interventions targeted toward at-risk students (Marsh et al., 2009). The premise of such policies is that the threat of retention along with additional intervention programs will both motivate and help students meet grade standards. Supporters argue that the threat of retention will provide incentives for students to work harder, for parents to monitor their child's progress, and for teachers to focus on the development of basic skills among low-achieving students—all of which should lead to increases in student achievement (Allensworth, 2005; Jacob, Stone, and Roderick, 2004; Roderick and Engel, 2001; Roderick, Nagaoka, and Allensworth, 2005; Roderick, Bryk, et al., 1999). However, critics of test-based promotion policies believe that standardized tests tend to narrowly direct teaching efforts and school resources toward raising student test scores at the expense of teaching other important skills. They contend that such tests have measurement errors and often fail to accurately define a student's achievement in a subject area. Moreover, a single measure may not adequately characterize a child's progress (Hartke, 1999; Heubert and Hauser, 1999; Roderick, Nagaoka, and Allensworth, 2005; Thompson and Cunningham, 2000).

Although philosophical differences regarding the effects of retention remain and the debate continues, the push for standards and test-based accountability has led to a growth in test-based promotion policies. This monograph presents new and rigorous evidence from a large, longitudinal study of New York City's (NYC's) 5th-grade promotion policy on the effects of grade retention on students' academic and nonacademic outcomes and the effects of supportive interventions, such as Saturday school and summer school. The findings detailed here should help inform the current heated debate about whether retention is "helpful, harmful, or harmless" (Roderick and Nagoaka, 2005).

## The Current Study

As part of an ambitious reform initiative, the New York City Department of Education (NYCDOE) implemented a new test-based promotion policy for students in grade 3 in 2003–2004. The policy was extended to grade 5 in 2004–2005, grade 7 in 2005–2006, and grade 8 in 2008–2009. General education students in these grades are required to meet promotion criteria on the English language arts (ELA) and mathematics assessments in order to be promoted, i.e., they must score at Performance Level 2 or higher on both these assessments. The grade promotion policy is not based on a sole criterion: Students may demonstrate basic proficiency in ELA and mathematics either through their performance on standardized tests administered during the school year or in August or through a review of a portfolio of their work.

As described in greater detail in Chapter Three, the policy places considerable emphasis on identifying struggling students early, providing them with additional instructional time, and continuously monitoring their progress. Students who have been identified as needing services at the beginning of the school year (those who scored Level 1 or low Level 2 on the previous year's assessments or those identified by teachers as needing additional help) are to receive additional instructional support in school, including differentiated instruction in the classroom and other intervention services targeting specific student areas of difficulty with small-group instruction. In addition to providing instructional support in school, schools may offer a variety of out-of-school support services, including Saturday school, which, in early years, was organized as a series of Saturday Preparatory Academies (SPAs). Students who fail to meet promotion standards on the mathematics or ELA assessments or portfolio evaluation in the spring are mandated to enroll in the Summer Success Academy (SSA), which offers additional hours of intensive instruction in mathematics and ELA over several weeks in the summer. Student progress is to be monitored by academic intervention services (AIS) or other school teams using pre- and post-assessments in reading, writing, and problem solving; periodic progress reports; and a student portfolio assessment. Particular attention is to be given to monitoring students who have already been retained.

NYCDOE asked the RAND Corporation to conduct an independent longitudinal evaluation of the 5th-grade promotion policy, with a follow-up of outcomes for 3rd-grade students. This study—conducted between March 2006 and August 2009—examined several aspects of the 5th-grade promotion policy: (1) policy implementation, factors affecting implementation, and implementation progress over time; (2) the impact of the policy on student academic and socioemotional outcomes; and (3) the links between implementation and desired outcomes. It also included an analysis of achievement outcomes for two cohorts of 3rd graders held to the promotion policy. This monograph presents the results of the overall study and is one of three reports documenting the overall evaluation of NYC's 5th-grade promotion policy. The two other reports—Marsh et al. (2009) and Xia and Kirby (2009)—help situate the NYC policy in the broader national context and in the context of prior literature. The first presents the results of a review of states and districts (other than NYC) that have implemented grade retention policies similar to the NYC promotion policy; the review included interviews with state and district officials regarding their successes, challenges, and lessons learned regarding policy design and implementation. The second provides a comprehensive and detailed review of the literature on grade retention—in particular, what we know about the characteristics of retained students and the short- and longer-term effects on student outcomes (both academic and nonacademic). This monograph includes key material from each of these reports.

## Contribution of the Study

This monograph builds on and extends a body of research examining the effects of retention on students. Along with its companion reports, it makes three major contributions to research, policy, and practice. First, it is one of a handful of longitudinal studies to examine the implementation and effects of a promotion policy on both academic and nonacademic outcomes over time and for a large group of students. Our study followed close to 60,000 students in each of three 5th-grade cohorts subject to the policy and one comparison 5th-grade cohort that was not subject to the policy for two to four years beyond the promotion gate grade. Our analyses are based on both qualitative and quantitative data collected from several different sources using a variety of methods and offer a rich portrait of the implementation and effects of a grade promotion policy in the largest public school system in the country.

Second, it adds to a small body of literature that attempts to model the effects of grade retention and supportive interventions on at-risk or retained students using quasi-experimental designs that compare outcomes in a treatment group to those in a plausibly similar control group. The issue of an appropriate comparison group is a problem that has plagued many earlier studies and produced inconsistent results. Here, we use two powerful quasi-experimental methods to provide compelling evidence regarding treatment effects. An important extension of the literature is to examine the relationship between intensity of treatment (for example, the number of Saturday or summer school sessions attended) and student outcomes to understand whether there are threshold levels of participation before beneficial effects are seen.

Third, it highlights the factors that enabled or hindered the implementation of the NYC promotion policy and uses data from both NYC's experience and those of other states and districts with similar test-based promotion policies to draw lessons learned regarding the design and implementation of promotion policies that should prove valuable to policymakers considering adopting such policies.

Overall, the study focused on seven major questions regarding NYC's promotion policy:

1. What types of supports did schools and Saturday and summer sessions provide to students?
2. What were the trends in student achievement for students held to the policy?
3. Relative to comparable groups of students, how did 5th graders needing services at the beginning of the year perform on the spring assessments?
4. Relative to comparable groups of students, how did students at risk of retention perform on the summer assessments?
5. Relative to comparable groups of students, what were the future academic outcomes of at-risk students?

6. What were the socioemotional outcomes of students subject to the promotion policy?
7. What are the lessons learned about the design and implementation of policies from NYC's experience and those of other states and districts with similar promotion policies?

The last-mentioned research question is fully addressed in the companion report, Marsh et al. (2009). We do not discuss it here, with the exception of noting some of the important lessons learned about policy design and implementation in the conclusions and summary. The remainder of this monograph focuses on the first six major research questions.

## Terminology

We focus on two groups of low-performing students in this monograph:

- The first group includes students who scored Level 1 or low Level 2 on the 4th-grade assessments or who had been retained in grade. These students were identified as needing additional help, likely targeted for AIS in schools, and offered services through Saturday programs/SPAs. We refer to this group as "students needing services."
- The second group consists of students who failed the spring assessments and so were "at risk of retention" in 5th grade. This second group was mandated to attend summer sessions through the SSAs and offered other chances to meet the promotion standards (through a portfolio review or summer assessments). We refer to this group as "students at risk of retention."

We tracked four cohorts of 5th-grade students: three cohorts (2004–2005, 2005–2006, and 2006–2007) that were subject to the policy and a comparison cohort (2003–2004) that was not held to the policy. To improve readability and make clearer the distinction between the policy and comparison cohorts, we refer to the policy cohorts as the P1, P2, and P3 cohorts and the comparison cohort as the C0 cohort.

## Organization of This Monograph

Part of the larger study involved a rigorous review of the literature on grade retention. We also examined what was known about the design and implementation of promotion policies and the impact of supportive interventions. A summary of this literature review, fully documented in Xia and Kirby (2009), is provided in Chapter Two. Chapter Three describes the Children First Initiative, a broad set of reforms undertaken by

NYC to improve student performance and within which NYC's promotion policy was enacted. It also presents the conceptual framework underpinning our analyses and situates the NYC policy in the broader state and federal context. Chapter Four discusses our data and methods, along with study limitations. Chapters Five and Six address the implementation of the policy: Chapter Five presents findings from administrator surveys and case studies about the in-school supports provided to in-need students and the challenges that schools face in doing so; Chapter Six uses administrator surveys and case studies to describe SPAs and SSAs and the factors that foster or hinder the implementation of these two supportive interventions. Chapter Seven examines overall trends in student performance, focusing on the outcomes of retained and at-risk but promoted students. Chapters Eight through Ten present results from our formal statistical analyses: Chapter Eight focuses on proximal year 5th-grade outcomes and presents estimates of the effects of the supports provided during the school year and summer; Chapter Nine focuses on the 6th- and 7th-grade outcomes of students at risk of retention who were either promoted or retained; and Chapter Ten uses data from student surveys to examine differences in the socioemotional outcomes of three groups of students over a three-year period—those retained in grade; those at risk of retention but eventually promoted via a portfolio review, by passing the summer assessment, or through appeal; and those not at risk of retention and promoted. Conclusions and policy implications are discussed in Chapter Eleven.

A series of appendixes provide supporting information. Appendix A is the technical appendix for the statistical methods presented in Chapter Four and used in the models presented in Chapters Eight and Nine. The next three appendixes provide supporting tables and figures for various chapters: Appendix B for the analyses of implementation discussed in Chapter Five, Appendix C for the performance trends discussed in Chapter Seven, and Appendix D for the analyses of the socioemotional outcome models presented in Chapter Ten. Appendix E presents data and analyses for the 3rd-grade cohorts: (1) performance trends for the comparison cohort and the two cohorts that were subject to the 3rd-grade promotion policy, similar to the analyses presented for the 5th-grade cohorts in Chapter Seven; (2) modeling results of the effects of the interventions, similar to those discussed in Chapters Eight and Nine; and (3) analyses of socioemotional outcomes, similar to those presented in Chapter Ten.

CHAPTER TWO

# What We Know About the Effects of Grade Retention and Implementation of Promotion Policies

*Nailing Xia and Sheila Nataraj Kirby*

As part of the overall study, we conducted a systematic and rigorous search of the literature on grade retention—in particular, that on the characteristics of retained students and the short- and longer-term effects on student outcomes (both academic and nonacademic). The full set of findings is documented in Xia and Kirby (2009), which also provides a brief summary of each of the 91 articles that met the criteria for inclusion in the review.

Given that the genesis of the literature review was the larger evaluation of NYC's promotion policy, we were also interested in literature that examined the design and implementation of promotion policies and the relationship between supportive components of policies and student outcomes. Findings from this literature—admittedly quite sparse—are presented at the end of this chapter.

## Methods

We conducted a systematic search of the literature to identify relevant studies published since 1980. Three selection criteria were used for inclusion: relevance, methodological rigor, and publication date. To be included, a study had to examine retention in grades K–12, and results had to address at least one of the following issues: characteristics of retained students or the effects of grade retention on several student outcomes. The study had to be either empirical in nature, using well-established statistical methods, or a systematic and rigorous review of past research. Empirical studies needed to use a credible comparison group or statistical method to control for selection bias. Studies based mainly on simple descriptive statistics were excluded. For a review essay to be included, it had to have a clear analytical method for synthesizing past research, such as a systematic search of the literature, selection criteria for inclusion, and statistical procedures for combining the results of past studies (meta-analysis, for example). Only studies published between 1980 and 2008 were included in the review. Ninety-one studies met all three criteria and were included in our review. Among the 91 selected studies, 87 were empirical, three were meta-analyses, and one was a systematic review of past research. Table 2.1 shows the number of studies by topic area.

**Table 2.1**
**Number of Studies, by Topic**

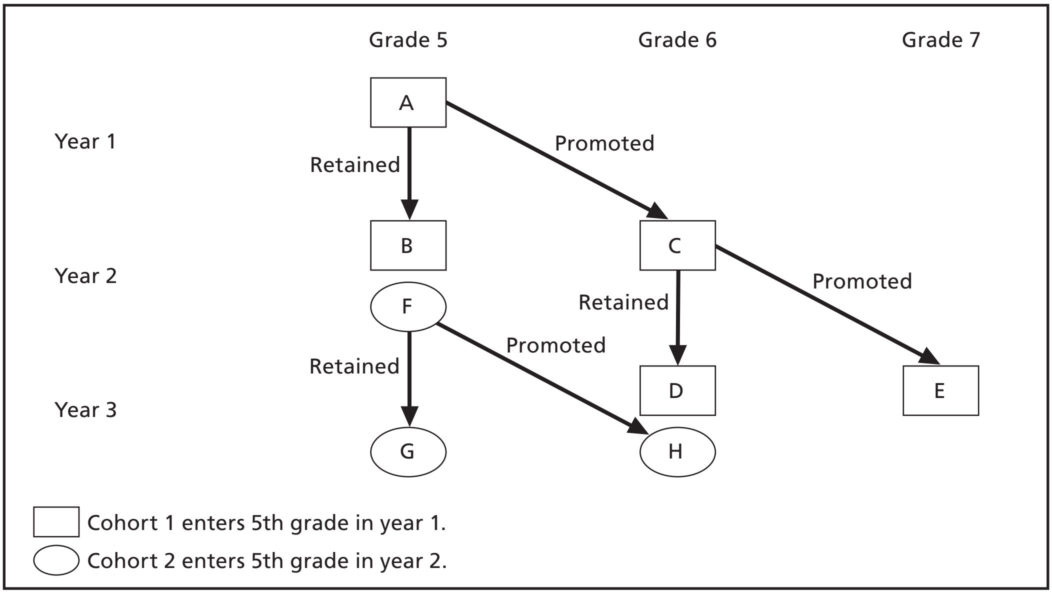
| Topic | Total Number of Studies | Number of Empirical Studies | Number of Reviews |
|---|---|---|---|
| Characteristics of retained students | 32 | 32 | 0 |
| Effects of grade retention | 77 | 73 | 4 |
| Academic outcomes | 55 | 52 | 3 |
| Socioemotional outcomes | 29 | 26 | 3 |
| Propensity to drop out of school | 17 | 16 | 1 |

NOTE: Some studies examined both the characteristics of retained students and the effects of retention and thus are counted in both categories. A total of 91 studies were included in the review.

The studies included in our review used a variety of analytical approaches. They differed on such issues as types of comparisons, comparison groups, and statistical methods, which may have affected the findings. Three types of comparison strategies were used in the retention literature: same-age comparison, same-grade comparison, and "across-year" comparison. Same-age comparisons involve comparing retained students with their peers who are the same age but promoted to higher grade levels. Same-grade comparisons use measurements of the performance of retained and promoted students in the same grade as the basis for comparison. There are two variations of same-grade comparisons. In the first variation, the outcomes of retained students measured in a given year are compared to the outcomes of same-age promoted peers measured a year earlier. Essentially, this approach involves comparing data from different years for retained and promoted students. The second variation uses younger, nonretained students as the comparison group. In other words, the comparison groups are in the same grade but are younger than the retained students at the time of the comparison.

An example helps illustrate the differences between same-age and same-grade comparisons. Suppose a student is retained in 5th grade for one year. At the end of the retention year, a same-age comparison would compare his current performance with that of his former classmates, who are now completing their 6th-grade year (i.e., comparing B with C, as illustrated in Figure 2.1). The first variation of same-grade comparison would compare the retained student's performance at the end of the repeated year with the outcomes of the 6th graders who were measured one year earlier, when they were his classmates (i.e., comparing B with A). Both outcomes are measured at the end of the 5th grade, but in two different years. The second variation of same-grade comparison would compare the retained student's performance at the end of the retention year with that of his current 5th-grade classmates who are one year younger than the retained student, assuming that both are on grade when they enter 5th grade (i.e., comparing B with F).

**Figure 2.1**
**Illustration of Comparison Strategies**




SOURCE: Adapted from Alexander, Entwisle, and Dauber (2003).
RAND *MG894-2.1*

Researchers also used different matching strategies to define a comparison group, including matching retained students with their peers who were recommended for retention but eventually promoted, matching students based on a selected list of variables, propensity score methods, or regression discontinuity designs (RDDs).[1] We discuss these issues in further detail in Chapter Four when we describe the data and methods used in our study.

[1] In the past decade, the latter two methods have received widespread attention as powerful quasi-experimental methods that can provide credible evidence regarding treatment effects in the absence of experimental studies (Shadish, Cook, and Campbell, 2002; West, Biesanz, and Pitts, 2000). Here, we briefly note that (1) propensity score methods account for differences between treatment and control groups by modeling the process by which a study participant is assigned to the treatment being studied (Rosenbaum and Rubin, 1983; McCaffrey, Ridgeway, and Morral, 2004), and (2) an RDD compares students who fall just below a particular threshold (a cutoff score, for example, assigns them to receive a particular intervention, like summer school) to those who score just above it with respect to particular outcomes (Imbens and Lemieux, 2008; Shadish, Cook, and Campbell, 2002). The first method uses all the treated participants; the second focuses on a subset of treated and control students who are close to the threshold on the assignment variable. Among the studies reviewed here, Hong and Raudenbush (2005); Hong and Yu (2007, 2008); and Wu, West, and Hughes (2008) used propensity score methods, while Jacob and Lefgren (2002), Roderick and Nagaoka (2005), Greene and Winters (2006), and Matsudaira (2008) used RDD models.

## Characteristics of Retained Students

There is general consensus in the literature on the characteristics of retained students. Retained students were more likely to be male, minority, and of lower socioeconomic status (SES). Most studies reported that African-American students were at an increased risk of retention, compared to their white peers (Blair, 2001; Dauber, Alexander, and Entwisle, 1993; Frymier, 1997; Hauser, Pager, and Simmons, 2000; Hong and Yu, 2007; Lorence et al., 2002; Roderick, Bryk, et al., 1999), while some found that Hispanics were also more likely to repeat a grade (Frymier, 1997; Hauser, Pager, and Simmons, 2000). Students in the lowest income quartile were 41 percent more likely to be retained than those in the highest quartile (Hauser, Frederick, and Andrew, 2007), and among high school graduates, those in the lowest SES quartile were twice as likely as high-SES graduates to have been retained (Fine and Davis, 2003).

Retained students were more likely to come from poor households (Byrd and Weitzman, 1994; Corman, 2003; Dauber, Alexander, and Entwisle, 1993; Frederick and Hauser, 2006; Guevremont, Roos, and Brownell, 2007; Lorence et al., 2002) and single-parent families (Corman, 2003; El-Hassan, 1998; Hong and Raudenbush, 2005). Parents of retained students were more likely to have lower IQ scores (Jimerson, 1999; Jimerson, Carlson, et al., 1997), lower educational levels (Byrd and Weitzman, 1994; Corman, 2003; Dauber, Alexander, and Entwisle, 1993; El-Hassan, 1998; Frederick and Hauser, 2006, 2008; Liddell and Rae, 2001), lower occupational levels (El-Hassan, 1998; Frederick and Hauser, 2006, 2008), poorer attitudes toward their child's education (Hong and Raudenbush, 2005; Willson and Hughes, 2006), lower expectations of their child's educational attainment (Hong and Raudenbush, 2005), and less involvement in school (Jimerson, Carlson, et al., 1997; Marcon, 1993; McCoy and Reynolds, 1999; Reynolds, 1992).

Compared to promoted students, retained students were found to fare poorly on cognitive and academic measures, including early academic standing (Alexander, Entwisle, and Dauber, 2003; Dauber, Alexander, and Entwisle, 1993; McCoy and Reynolds, 1999); IQ scores or cognitive test scores (Blair, 2001; Liddell and Rae, 2001; Mantzicopoulos et al., 1989; Safer, 1986); academic achievement prior to retention (Fine and Davis, 2003; Gottfredson, Fink, and Graham, 1994; Hong and Raudenbush, 2005; Hong and Yu, 2007; Liddell and Rae, 2001; Mantzicopoulos et al., 1989; Marcon, 1993; Reynolds, 1992; Robles-Pina, Defrance, and Cox, 2008; Safer, 1986; Willson and Hughes, 2006); special education referrals and placements (Ferguson, Jimerson, and Dalton, 2001); and socioemotional adjustment, skills, and behavior prior to retention (Alexander, Entwisle, and Dauber, 2003; Blair, 2001; Byrd and Weitzman, 1994; Dauber, Alexander, and Entwisle, 1993; Ferguson, Jimerson, and Dalton, 2001; Hong and Raudenbush, 2005; Jimerson, 1999; Jimerson, Carlson, et al., 1997; Mantzicopoulos et al., 1989; Reynolds, 1992; Robles-Pina, Defrance, and Cox, 2008; Safer, 1986).

## Effects of Grade Retention on Students' Academic and Nonacademic Outcomes

### Effect on Academic Outcomes

Overall, research on the academic effect of grade retention shows that retention alone is ineffective in raising student achievement. Studies that reported positive or mixed findings focused on short-term effects,[2] used same-grade comparisons, or evaluated retention policies with supportive components.

While retained students appeared to make significant gains during the retention year, improvements were often not big enough to bring them to the same performance level as their promoted peers (Alexander, Entwisle, and Dauber, 2003; Karweit, 1999). Moreover, those gains were typically short-lived and tended to fade in subsequent years (Alexander, Entwisle, and Dauber, 2003; Baenen, 1988; Jacob and Lefgren, 2002; Jimerson, Carlson, et al., 1997; Jimerson, 2001; Karweit, 1999; Lorence and Dworkin, 2006; Lorence et al., 2002; Mantzicopoulos and Morrison, 1992; Nagaoka and Roderick, 2004; Peterson, DeGracie, and Ayabe, 1987; Roderick and Nagaoka, 2005). Several studies reported that academic gains found among retained students in the short term disappeared several years later, and retained students eventually fell behind again (Baenen, 1988; Mantzicopoulos and Morrison, 1992; Nagaoka and Roderick, 2004; Peterson, DeGracie, and Ayabe, 1987; Roderick and Nagaoka, 2005).

Findings on the academic outcomes of retention varied, depending on the basis of comparison. Same-age comparisons generally suggested negative consequences of retention on student performance. Same-grade comparisons often yielded mixed findings, depending on other features of the research design, such as how long students were followed after retention. Studies using matched control groups of low-achieving but promoted students tended to find no academic benefit, or even negative impact on retained students (Holmes and Saturday, 2000; Karweit, 1999; Roderick, 1994; Tanner and Galis, 1997). Studies using same-grade comparison without a matched control group often showed no harm or benefit to retained students.

Although many studies did not report whether retention was accompanied by supportive interventions, in a few studies that found positive academic outcomes in certain grades, retained students received targeted interventions designed to help them overcome individual problems (Lorence and Dworkin, 2006; Lorence et al., 2002; Greene and Winters, 2004, 2006, 2007, 2009; Peterson, DeGracie, and Ayabe, 1987). Some researchers commented that it was unclear whether the positive outcomes were the result of retention or the supportive components (Fager and Richen, 1999; Hauser, 1999).

For example, Roderick and Nagaoka (2005) and Greene and Winters (2006) were able to exploit data from recently implemented test-based promotion policies

---

[2] We define *short term* as within three years after retention and *long term* as four or more years after retention.

in Chicago and Florida, respectively, and used methods similar to those used in our study. They relied on both "across-year" designs that compared students from a prior cohort not subject to the policy with students in a cohort subject to the policy and an RDD, in which they compared students with test scores that were very close to the cutoff score (above and below). Both studies used same-age comparisons focusing on one subject: ELA. The Chicago study (Roderick and Nagoaka, 2005) found a small positive impact on the performance of retained 3rd- and 6th-grade students relative to promoted students in the year in which students were retained, but these gains disappeared among 3rd graders and were reversed two years after the baseline year among 6th graders. In contrast, Greene and Winters (2006), using similar analytical models and designs, found small positive but statistically significant gains among retained students in Florida—gains that increased over time. For example, retained students benefited by between 0.01 and 0.05 standard deviations after one year and by 0.15–0.16 standard deviations after two years, depending on how strictly the researchers defined the region around the threshold (within 25 or 50 points of the cutoff score). They hypothesized that differences in findings may be due to differences in the design and implementation of the promotion policies and the fact that the Chicago policy underwent several changes during its implementation.

Only four studies examined whether the timing of grade retention matters in terms of academic effects. Two studies found no statistically significant difference between students retained in early grades and those retained in later grades (Baenen, 1988; Silberglitt, Appleton, et al., 2006). Two reported that retention in later grades was associated with poorer academic outcomes than retention in early grades (Hagborg et al., 1991; Meisels and Liaw, 1993).

### Effect on Socioemotional Outcomes

Conventional wisdom predicts that retention will have negative effects on students' emotional health and social adjustment by lowering their self-esteem, causing emotional distress, and decreasing their peer acceptance. Yet, empirical findings on the socioemotional effects of retention are mixed and inconclusive.

Among the studies included in our review, some used teacher ratings[3] to measure student motivation, some used student responses to questionnaires or interviews designed to assess their socioemotional adjustment in response to retention or their attitudes toward school, and a few used a combination of teacher- and student-reported measures.

Compared to promoted peers, retained students were significantly more attached to school almost a year after retention (Gottfredson, Fink, and Graham, 1994) or had

---

[3] Studies that use teacher ratings to measure student attitudes are subject to threats to validity because teacher ratings may be influenced by teachers' attitudes toward grade retention. For example, teachers who do not support a retention policy may be more likely to report negative attitudes among retained students.

more positive attitudes toward school (Bonvin, Bless, and Schuepbach, 2008). But these results varied by comparison group (Bonvin, Bless, and Schuepbach, 2008). In contrast, Shepard and Smith (1987, 1989) reported that retained kindergartners had more negative attitudes toward school than nonretained students after the 1st grade. One meta-analysis reported that retained students had less favorable attitudes toward school than their promoted peers (Holmes and Matthews, 1984). However, it is important to note that this meta-analysis included studies with methodological problems, including a lack of statistical controls, small sample sizes, and older data.

Results from studies examining retention's effects on self-concept are mixed. Depending on the study, retention was associated with positive effects on academic self-concept (Bonvin, Bless, and Schuepbach, 2008; Reynolds, 1992), significantly lower perceptions of cognitive competence (Pierson and Connell, 1992), or no significant impact on self-concept (McCoy and Reynolds, 1999; Pomplun, 1988; Shepard and Smith, 1987, 1989).

Among the studies that measured it, some reported that retained students had significantly lower self-esteem (Frymier, 1997; Hagborg et al., 1991; Jimerson, Carlson, et al., 1997; Setencich, 1994) while others indicated higher self-esteem (Hong and Yu, 2008; Plummer and Graziano, 1987). One study found generally positive effects among the 2nd and 3rd graders after adjusting for preretention and demographic factors. However, students retained in 1st grade continued to score lower on self-esteem than promoted peers, although the gaps in scores after the retention year were considerably smaller than those prior to retention (Alexander, Entwisle, and Dauber, 2003).

Two meta-analyses (Holmes and Matthews, 1984; Jimerson, 2001) reported that retained students scored significantly lower than promoted peers on measures of social, emotional, and behavioral adjustment. In another meta-analysis, Holmes (1989) reported negative results on measures of personal adjustment but no statistically significant differences with respect to any of the subcategories of personal adjustment (i.e., social, emotional, and behavioral adjustment).

### Effect on the Propensity to Drop Out of School

Among the empirical studies that examined the relationship between retention and dropping out of school, there was general consensus that students retained for one or more grades were more likely to drop out of school than their promoted peers (Alexander, Entwisle, and Dauber, 2003; Allensworth, 2004, 2005; Eide and Showalter, 2001; Grissom and Shepard, 1989; Guevremont, Roos, and Brownell, 2007; Jacob and Lefgren, 2007; Jimerson, 1999; Jimerson and Ferguson, 2007; Jimerson, Ferguson, et al., 2002; Lorence and Dworkin, 2004; Roderick, 1994). Four empirical studies indicated that grade retention was one of the strongest predictors of dropping out (Goldschmidt and Wang, 1999; Ou and Reynolds, 2008; Rumberger and Larson, 1998; Rush and Vitale, 1994).

## Effect of Supportive Components of Promotion Policies on Student Achievement

While grade retention essentially requires students to repeat a grade, the experiences of students who are retained or at risk of being retained can differ in important ways, depending on whether and how additional programs are designed to support struggling students. Examples of supportive programs include early identification of at-risk students, individualized education plans, small-group instruction, after-school programs, summer school, and continuous evaluation of student performance. Given the structure of NYC's promotion policy, we were particularly interested in literature that looked at the relationship between supportive components of policies and student outcomes. We found 24 studies that provided details on the decisionmaking process and supportive programs. Among them, 18 studies evaluated test-based promotion policies adopted in Chicago, Florida, and Texas.

In this section, we first present four studies—three empirical studies and one meta-analysis—that examined the relationship between summer school programs and student achievement. We then discuss other studies that looked at additional components of the promotion policies.

### Summer School Programs and Student Achievement

Three empirical studies examined the effects of summer school on academic achievement in districts that have adopted test-based promotion policies, and a meta-analysis reviewed 93 studies of the effects of summer school. All four studies reported increases in student achievement as a result of attending summer school.

Jacob and Lefgren (2002) attempted to disentangle the effects of retention from summer school attendance and reported mixed findings. Retention appeared to increase the short-term performance of retained students in the 3rd grade but had no impact on mathematics and a negative effect on reading for those in the 6th grade.[4] They found that "summer school increased academic achievement in reading and mathematics and that these positive effects remain substantial at least two years following the completion of the program" (Jacob and Lefgren, 2002, p. 3). When the effects of summer school and retention were combined, the authors reported academic benefits to 3rd graders and zero effects among 6th graders. The authors hypothesized that the zero net effects for sixth graders masked a small positive summer school effect and a negative retention effect.

A second study evaluated Summer Bridge, the mandatory summer school program for at-risk and retained students in Chicago, concluding that students in Summer Bridge, especially those in the 6th and 8th grades, experienced significant increases

---

4 The authors argued that the negative effects on reading among the 6th graders may have been due to differential test incentives faced by retained and promoted students.

in their test scores over the summer (Roderick, Engel, et al., 2003). However, the study found no evidence that the program "affected school year learning rates nor did it address the fact that participating students continued to show low performance" during the school year (Roderick, Engel, et al., 2003, p. 3).

Matsudaira (2008) examined the effects of summer school in another large, urban school district. He found an average positive effect of summer school (0.12 standard deviations) for both mathematics and reading achievement the following year. However, the overall estimates masked considerable heterogeneity in the results. For example, he estimated an effect of 0.24 standard deviations for 5th graders who were mandated to summer school for mathematics, compared to 0.13 standard deviations for 3rd graders. The results were reversed for those mandated to summer school for reading. Here, the effect was larger for 3rd graders than for 5th graders (0.20 versus 0.10 standard deviations).

Cooper et al. (2000) conducted meta-analyses of 93 evaluations of summer school programs and found "a positive impact on the knowledge and skills of participants" (p. v). The meta-analyses included summer school programs focusing on a variety of goals, including remedial interventions and accelerated learning. Although all students appeared to benefit from attending summer school, students from middle-class households showed larger positive effects than students from disadvantaged households. In terms of remedial summer programs, results indicate that students in the earliest grades and in secondary school benefited the most. Moreover, remedial programs appeared to "have more positive effects on mathematics than on reading" (Cooper et al., 2000, p. v), and had larger effects when programs were relatively small and instruction was individualized.

### Findings on Other Supportive Components of Retention Policies

Holmes's (1989) meta-analysis of 63 studies found that those with positive findings often included such program characteristics as early identification of and special help for at-risk students. Students were not merely recycled through the same curriculum, but were supported through individualized education plans, continuous evaluation of academic performance, and low student-teacher ratios. In addition, many students were "mainstreamed into the regular program with their age peers for part of the day" (Holmes, 1989, p. 26).[5]

Following a cohort of 1st-grade students for three years, Karweit (1999) reported from teacher surveys that retained students had a very similar experience in the retention year in terms of classroom organization and instructional content and approaches. The study found positive academic effects based on same-grade comparisons, negative

---

[5] The studies discussed in this section focused primarily on the effects of grade retention on student outcomes. None attempted to directly estimate the relationship between specific supportive components and student outcomes. Rather, they mentioned the supportive components in their discussion of the retention policy.

academic outcomes based on same-age comparisons, and positive outcomes in terms of student attitudes.

Niklason (1987) conducted an analysis of covariance by comparing retained students in a district providing a special remediation program with retained students in other districts. The program offered after-school instruction on basic academic skills. The author reported no significant differences across districts.

Peterson, DeGracie, and Ayabe (1987) followed a cohort of students in the Mesa Public Schools in California, where the retention policy featured early identification and individual education plans for at-risk as well as retained students. They found positive academic effects using both same-grade and same-age comparisons, although gains appeared to diminish in subsequent years.

## Summary

Our review of the literature on grade retention produced the following conclusions:

- Retained students are more likely to be male, minority, of lower SES, and from single-parent families. They are also more likely to have poorer academic performance prior to retention; significantly lower social skills and poorer emotional adjustment; more problem behaviors, such as inattention and excessive absenteeism; more school changes; and poorer health.
- In general, retention alone does not appear to have long-term academic benefits for students. Some studies—including some recent ones—have found academic improvement in the years immediately after retention. However, other studies have shown that these gains are often short-lived and tend to fade over time.
- Being retained is associated with significant increases in the risk of dropping out of school.
- Summer school appears to have a small positive effect on student achievement.
- Findings on the social, emotional, behavioral, and attitudinal outcomes of retained students compared to those of their promoted peers appear mixed, with no overwhelming evidence to suggest that retention affects students negatively.

CHAPTER THREE

# Context and Conceptual Framework for Understanding New York City's Promotion Policy

*Jennifer Sloan McCombs, Sheila Nataraj Kirby,*
*Julie A. Marsh, and Catherine DiMartino*

Examining the implementation and effects of the NYC promotion policy requires understanding not only NYC's implicit theory of change, but also the larger city, state, and federal context against which the policy was implemented. In addition, there is a set of factors and actors that must be aligned for the policy to work as intended. This chapter begins by discussing the broader set of city reforms, then provides a detailed description of the 5th-grade promotion policy as a way of understanding NYC's theory of change. We then outline a conceptual framework for understanding the set of actors and factors that affect both implementation and expected outcomes. As part of the framework, we examine the state and federal context in which the city's efforts were embedded. The framework underpins our evaluation and guides the research questions addressed in this monograph.

## New York City's Reform Initiative

### The Children First Initiative

NYC is the largest public school system in the country, with approximately 1.1 million students, 79,000 teachers, and more than 1,400 schools. In 2002, the state legislature granted Mayor Michael Bloomberg control of the NYC school system in an effort to support a major overhaul of its schools. According to NYCDOE, prior to mayoral takeover, there was little standardization or coordination across the city, and student performance and outcomes were dismal, with too many students leaving school without the skills and knowledge needed for success. Mayor Bloomberg, with school chancellor Joel Klein, launched an ambitious new reform initiative, Children First, in 2002 "to create a system of outstanding schools where effective teaching and learning is a reality for every teacher and child" (NYCDOE, undated[c]).

Children First began with the reorganization of NYCDOE's management structure, focusing on centralizing a decentralized system. The restructuring eliminated the citywide board of education and community school boards and organized the 32 independent community school districts into 10 regions that were responsible for supporting and overseeing the schools. These regions became the focus of the new governance structure, with regional superintendents reporting directly to the chancellor (Fruchter, 2006).

In addition to the reorganization, NYCDOE adopted a host of reform policies (aside from the promotion policy discussed later). It implemented a systemwide approach to instruction in reading, writing, and mathematics that was reinforced with annual testing in grades 3 through 8;[1] a new parent support system that provided a parent coordinator for each school; and a leadership academy to train new school leaders. In 2004, Chancellor Klein launched a pilot program called the "autonomy zone." Principals of schools in the pilot program were provided with additional decisionmaking power over their programs, personnel, and finances in exchange for pledging to meet ambitious achievement targets. In 2006–2007, the program expanded into the Empowerment Schools initiative, which included 332 schools.

NYCDOE worked with the United Federation of Teachers to establish a new contract, which, among other things, allowed NYCDOE to create lead teacher positions with a $10,000 salary differential; provided a $15,000 housing incentive for experienced mathematics, science, and special education teachers who agreed to teach for at least three years in high-need schools; and gave principals the power to make final decisions regarding hiring for vacancies. In addition, the new teacher agreement extended each school day by 37.5 minutes so that teachers could work with students in small groups and provide struggling students with additional help.

In the spring of 2007, Chancellor Klein announced a second restructuring, driven by two goals: promoting school-based decisionmaking and introducing market competition into the school support arena. With the 2007 reform, the mayor and chancellor aimed to empower principals by giving them more decisionmaking power in their individual schools, including influence over curriculum selection, human resources, and the budget. At a Manhattan town hall meeting on February 6, 2007, Chancellor Klein commented,

> Decisions are best for kids when they're happening close to kids at the school level. Starting in 2007–2008, rather than being told what to do by distant bureaucrats, principals and school communities will have decision-making power and they'll be responsible for results.

Unlike the previous system, in which the region selected curricular and intervention materials and organized programs intended to help at-risk students, this new system gave principals the power to make these decisions and to select the type of sup-

[1] In response to the requirements of the No Child Left Behind Act (NCLB), New York State began administering state assessments in grades 3 through 8 in ELA and mathematics in spring 2006, replacing the spring assessments previously given in these two subjects by the city in grades 3, 5, 6, and 7 and by the state in grades 4 and 8. The new state assessments are not equivalent to those administered in prior years, and a separate standard-setting procedure was used to establish the proficiency-level scores.

port organization that best served the needs of their school. There are three distinct types of school support organizations (SSOs): the Empowerment Support Organization (ESO), which grew out of the pilot program; Learning Support Organizations (LSOs), which provide support that is most similar to the previous regional governance model and are, in fact, led by former regional leaders; and Partner Support Organizations (PSOs), which are external private-sector organizations that provide their own models of support.

NYCDOE describes the rationale for the transition as follows:

> Through the regional offices, the Department of Education invested resources and made decisions on behalf of schools. Central and regional decision-making led to uniform solutions, even though each of our schools has unique needs and challenges. While effective at capacity building and bringing coherence to a large system, the one-size-fits-all approach does not maximize the investment in children's futures. (NYCDOE, undated[a])

In 2007–2008, in return for greater autonomy, principals were held accountable for meeting specific targets that measured students' progress. Schools that were not meeting those targets faced consequences, while schools that were meeting or exceeding standards received rewards.

### The Promotion Policy

As part of Children First, NYCDOE implemented a new promotion and retention policy for students in grades 3, 5, 7, and, most recently, 8. General education students in these grades are required to score at or above Performance Level 2 on the mathematics and ELA assessments or demonstrate Level 2 performance through a portfolio review in order to be promoted. New York State established four performance levels for its assessments:

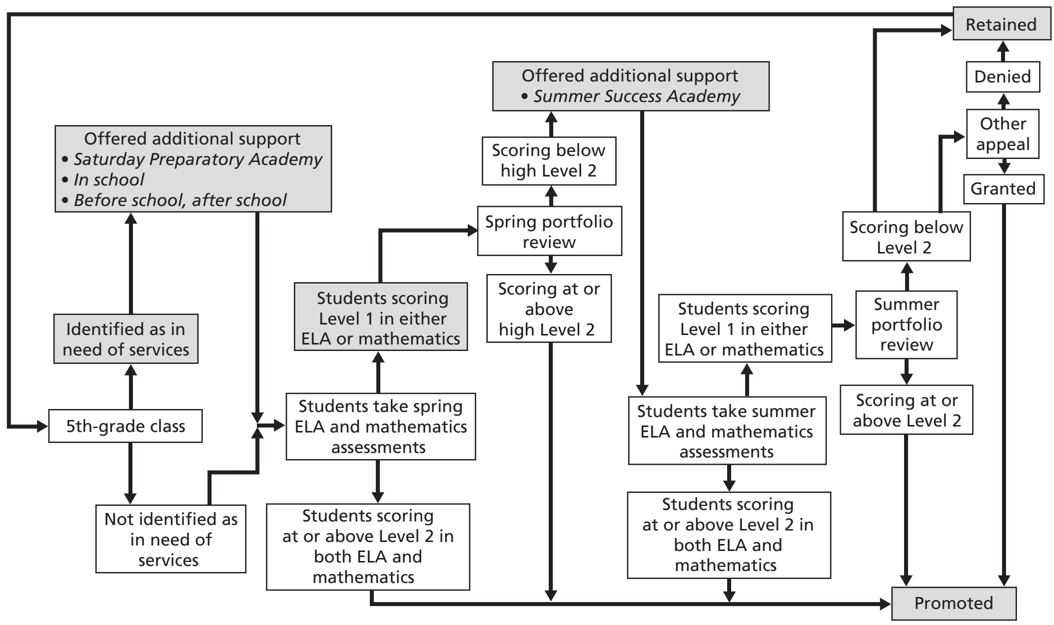
- Level 4—exceeds the standards
- Level 3—meets all the standards
- Level 2—meets some of the standards or partially meets the standards
- Level 1—shows serious academic difficulties.

Performance at or above Level 3 is considered "proficient" under NCLB, a higher standard than the promotion benchmark.

### Criteria for Promotion

Figure 3.1 shows how NYC's 5th-grade promotion policy works. As mentioned, the 5th-grade promotion policy was the focus of our study. Under the policy, promotion is not based on a single criterion; students may demonstrate basic proficiency in ELA

**Figure 3.1**
**The NYC 5th-Grade Promotion Policy**




RAND *MG894-3.1*

and mathematics either through their performance on standardized tests administered during the school year or in August or through a review of a portfolio of their work.

As noted earlier, to be promoted, 5th-grade students must achieve Level 2 or higher on both the ELA and the mathematics standardized tests. There is an automatic appeals process for students who do not meet promotion criteria through the standardized assessments. The work of all students who score at Level 1 in ELA, mathematics, or both is collected in a portfolio by their teachers and assessed using standard criteria. Each portfolio includes a common assessment in ELA, mathematics, or both, developed by NYCDOE, with a specific rubric for grading.[2] This section of the portfolio is mandated, while other original work demonstrating students' abilities is compiled at the discretion of the teacher or school. The portfolio is reviewed in June (and is generally referred to as the *spring portfolio*), and students who do not score a high Level 2 on the portfolio are mandated to attend SSA and take a summer assessment developed and administered by the city.[3] For those who still do not achieve Level 2 performance after the administration of the summer assessments, portfolios are reviewed again in August. Prior to the implementation of the policy, students who failed the

---

2 The purpose of the common assessment was to help standardize the portfolios and ensure comparability of portfolio review across the city.

3 Although New York State took over spring testing in grades 3–8 in 2005–2006, the summer assessment is still developed and administered by NYC.

spring assessments were given the opportunity to attend summer school and take the summer assessments.

All 5th-grade students are held to the new promotion policy with three exceptions: (1) promotion decisions for students with Individualized Education Programs (IEPs) are based on the promotion criteria specified by the IEP, (2) English-language learners (ELLs) enrolled in an English-language school system for less than two years are not held to promotion standards, and (3) charter school students are not held to the policy.[4]

### Key Components

The 5th-grade policy places considerable emphasis on identifying students who may struggle to meet promotion standards at the beginning of the school year (i.e., "students needing services"), providing them with supportive academic services, and continuously monitoring their progress.

**Early Identification.** Schools are to identify students needing services using prior-year test results, in-class assessments, and teacher or principal recommendations and must inform the parents or guardians of their children's status.

**Support Services.** Students who have been identified as needing services are to receive additional instructional support in school, including differentiated instruction in the classroom and other intervention services targeting specific areas of difficulty with small-group instruction. Students were also given an opportunity to receive support through SPAs. In 2004–2005, 2005–2006, and 2006–2007, the city helped support SPAs. In 2007–2008, it stopped requiring the program's operation, instead allowing each principal to decide whether to offer Saturday school sessions to their students. As described earlier, students who scored Level 1 on the mathematics or ELA assessment during the school year are encouraged to enroll in SSAs, which offer small class sizes and four hours of additional intensive instruction in mathematics and ELA four days per week for five weeks in the summer.

**Ongoing Monitoring.** School AIS teams monitor students' progress using pre- and post-assessments in reading, writing, and problem solving; periodic progress reports; and a student portfolio assessment. Particular attention is given to monitoring students who have already been retained and are at risk of being held over again.

---

[4] ELL students enrolled in an English-language school system for more than two years and less than three years must make satisfactory progress based on a comprehensive assessment. ELL students must meet at least one of the following requirements: score at Level 2 on the mathematics assessments, have student work that shows satisfactory progress in English as a second language (ESL) development, meet NYC performance standards in mathematics (taught in the native language or using ESL methodologies), or maintain 90-percent attendance. ELL students enrolled in an English-language school system for three years or more are expected to meet the typical 5th-grade promotional standards unless they have received prior approval.

### Support Structures for the Promotion Policy

Administrators of schools, SPAs, and SSAs receive a variety of supports in their efforts to implement the promotion policy and provide services to students at risk of retention. NYCDOE sets policy and provides guidance and training to support policy implementation. In 2005–2006 and 2006–2007, regions provided direct support for schools' implementation of academic intervention services and provided support for SPAs and SSAs. After the reorganization, in 2007–2008, schools received support from their SSOs, and requirements for SPAs and SSAs shifted.

**The Central Office.** NYCDOE outlined the regulations for the implementation of the citywide promotion policy in Regulation of the Chancellor A-501 (NYCDOE, 2009). The regulation provided an overview of the promotion policy and the academic interventions guaranteed to at-risk students. At the city level, the Division of Assessment and Accountability and the Division of Teaching and Learning shared responsibility for the implementation of the promotion policy and AIS. In the Division of Assessment and Accountability, the Office of Accountability covered all evaluative aspects of the promotion policy, including test administration, scoring, score reporting, portfolio development and evaluation, and the appeals process. To support portfolio development and the appeals process, the division offered a two-day workshop for regional and SSO leaders to outline the content of the portfolios and the logistics of the appeals process. Leaders attending the training were to share what they learned in the workshops with support staff and principals. While regional representatives were required to attend these training sessions, SSO representatives were not.

In the Division of Teaching and Learning, the Office of Academic Intervention Services focused on training schools to identify students needing services, differentiate their level of need, and provide services accordingly. In 2005–2006 and 2006–2007, officials in the Office of Academic Intervention Services met regularly with the regional directors of instructional services (RDISs) from the 10 regions to train them on various aspects of intervention services, such as identifying students and selecting intervention materials. In 2007–2008, the Office of Academic Intervention Services offered training workshops focused specifically on interventions, to which principals could send staff for a fee of approximately $50 to $100 per registrant (NYCDOE, undated[b]).

During the first year of the promotion policy, 2004–2005, NYCDOE centrally administered the Saturday sessions offered through SPAs to students identified as needing services. The promotion policy was announced in September 2004, and NYCDOE was able to open SPAs at the end of October. During this initial year, NYCDOE hired teachers, invited students to SPAs, and centrally ordered curriculum materials for the academies. While regions were provided with a choice to order alternative curricula, only one region selected this option. NYCDOE set the schedule for the SPAs and even provided a socioemotional growth program for the sites to use.

In school years 2005–2006 and 2006–2007, much of the responsibility for running the SPAs was delegated to the regions. However, the central office still provided guidance regarding the length of the programs (including start and end dates), student-teacher ratios, and daily schedules. NYCDOE also held curricular fairs that presented a number of curricula that regions could adopt. In 2007–2008, NYCDOE stopped mandating SPAs, and schools made their own decisions regarding whether to hold Saturday schools.

NYCDOE also set strict guidelines for the SSAs, outlining the length of the programs (including start and end dates), student eligibility, student-teacher ratios, types of teachers (i.e., teachers specializing in particular types of interventions), and daily schedules. NYCDOE central staff also supported facility selection and management, human resources and staffing, purchasing of materials, and food services. Although schools were provided greater flexibility in running SSAs beginning in summer 2007, NYCDOE still mandated that schools provide summer school services to students at risk of being retained, and the lower student-teacher ratios for SSA students remained constant even after the reorganization.

**The Regions.** Within NYCDOE's strict guidelines about various aspects of the promotion policy, the regions had some leeway in terms of implementation. The role of the RDIS, in particular, was to oversee and support the provision of AIS to identified students via the students' regular school—as opposed to providing these services via SPAs or SSAs. The RDISs tended to be experienced teachers and administrators; some were former regional instructional superintendents.

The primary role of the RDIS was to support implementation of AIS as part of the promotion policy. To this end, all held monthly meetings with AIS liaisons or team leaders, visited schools to observe the implementation of AIS and the AIS team meetings, provided additional support to schools that were identified as needing improvement, and provided professional development when requested by the schools on specific topics relating to AIS. During the monthly meetings, the RDISs typically provided information and clarification about guidance from NYCDOE (for example, on personal intervention plans, or PIPs, which the city required for multiple holdovers; promotion portfolios; and the appeals process). In addition, they used the monthly meetings to provide professional development regarding effective teams and intervention services, using data to assess students and track student progress, and individualizing instruction. On average, the majority of the RDISs' time—between 50 and 75 percent—was allocated to school visits. Most RDISs focused on 10–20 schools based on need, prioritizing schools that were identified as needing improvement under NCLB, those with significant numbers of multiple holdovers or low-performing students, and those that were referred for special attention by the superintendent. During school visits, the RDISs generally observed AIS team meetings and provided advice on improving the team structure and interactions as well as the AIS plans. They also observed the implementation of AIS and provided feedback. Multiple-holdover students were given special attention during these visits. Often, the principals of schools that were struggling asked the RDISs to provide professional development on specific topics.

The role of the RDISs in SPAs and SSAs varied across the regions. In some places, instructional superintendents were responsible for these academies and the RDIS had a minimal role, such as identifying the students who should attend the academies. Most RDISs, however, played an important liaison role in the SSAs, helping with the portfolios and the appeals process. In a few of the regions, the RDISs were more directly involved.

**The School Support Organizations.** As discussed earlier in this chapter, beginning with the 2007–2008 school year, the regional structure was dismantled and schools were supported by an SSO of their choice. Almost two-thirds (65 percent) of the schools partnered with the LSOs; this may have been because they were already familiar with the regional leaders and staff. About 30 percent chose to become part of the ESO, and a very small number (5 percent) elected to partner with one of the PSOs. SSOs helped principals to make decisions about such matters as curriculum, scheduling, and budgets. Unlike the regions that provided guidance on the implementation of the promotion policy and directly supported schools in AIS provision through RDISs, SSOs provide customized but not "supervisory" support to schools and respond to needs expressed by the principal, although some may be more proactive. SSO school support is organized around network teams that consist of a team leader and, on average, two to five specialists. These specialists have expertise in instructional support, services for special education and ELL students, data analysis, and business and operations management. Some SSOs offer regular meetings with AIS leaders.

Under the new organizational structure, decisions about Saturday school rested with the individual school. Principals who decided to offer Saturday schools often had prior experience with these programs and did not require assistance in this area. Additionally, the role of the SSOs in supporting SSAs was also limited. Some SSOs provided assistance such as running curriculum fairs, providing logistical information, and creating connections among principals in the same network to enable them to offer a joint SSA.

We now turn to a discussion of the conceptual framework that informs our evaluation.

## Conceptual Framework

Any new program or policy can be thought of as representing a theory, in that a policy or program decisionmaker hypothesizes that a particular treatment, program, or policy will have certain predicted effects or outcomes. Our study design, data collection, and analysis were guided by a conceptual framework grounded in NYC's implicit "theory of change" for the 5th-grade promotion policy, along with research on promotion policies and policy implementation more generally. NYC's theory is that the implementation of a comprehensive promotion policy—including the early identification of students at risk of being retained, parent outreach, the provision of support services to those students (e.g., AIS, SPAs, SSAs), ongoing monitoring of student progress in meeting promotion standards, the provision of an appeals process, and the retention of students who fail to meet standards—will improve student outcomes above and beyond what regular school services can provide. However, the conceptual framework for our study—shown in Figure 3.2—recognizes that the extent of the impact depends on the depth and quality of the implementation of the policy across the schools and the broader context of student background, supports, and school contextual factors that contribute directly and indirectly to student outcomes. Weiss (1972, p. 38) points out the importance of understanding the broader context. If a program or project is unsuccessful, the explanation may be that

> [it] did not activate the "causal process" that would have culminated in the intended goals (this is a failure of program), or it may have set the presumed "causal process" in motion but the process did not "cause" the desired effects (this is a failure of theory).

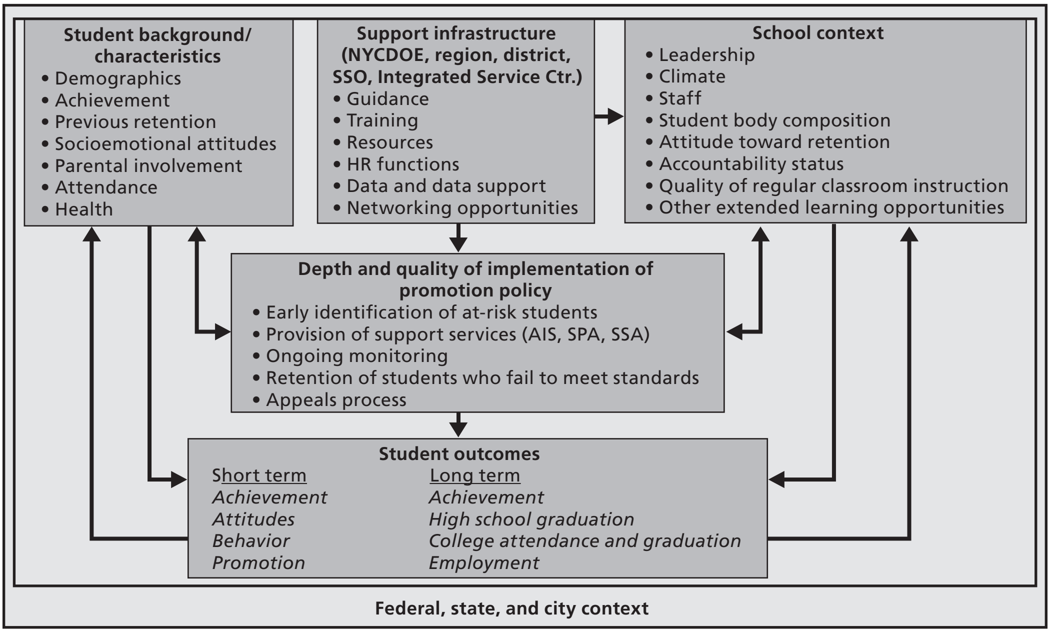
Thus, in the context of our study, it is important to examine both the theory itself and the implementation of that theory when analyzing outcomes.

As shown in the figure, the outcomes of interest are, in the short term, student achievement, attitudes (e.g., engagement in school, self-confidence in reading/mathematics), behavior (e.g., discipline), and promotion; in the long term, they are overall achievement, high school graduation, college attendance and graduation, and future employment. As described earlier, past research has shown that grade retention policies can affect this broad set of outcomes. For the purposes of this evaluation, we were able to measure only short-term student outcomes. We should note that, to keep the figure readable, we do not show the full set of interactions and feedback loops that are likely to exist among the various actors and factors. For example, we acknowledge that promotion policies can also affect parents and educators, who, in turn, could affect student outcomes. For example, participation in the promotion policy could bring teachers closer together or make them better teachers, potentially improving student achievement. Given the focus of our evaluation on student outcomes, we have not emphasized these intermediate outcomes in our framework.

Thus, changes in student outcomes are a function of the following sets of factors.

**Student Background.** Numerous studies have documented how a student's background characteristics can influence achievement, attitudes toward attainment, and changes in these two outcomes (see Coleman et al., 1966; Jencks at al., 1972; Gamoran, 1987, 1992; and Bryk, Lee, and Holland, 1993). There may be other

**Figure 3.2**
**Conceptual Framework for Evaluating the NYC Promotion Policy**




RAND *MG894-3.2*

characteristics that affect success but are not so easily measured or controlled for. Numerous student attributes are linked to greater participation and more successful outcomes, including demographic characteristics (e.g., eligibility for free or reduced-price lunch, minority status, gender), prior achievement, previous retentions, socioemotional attitudes, level of parental involvement in school, attendance, mobility, and health status. These factors may have a direct or indirect impact on student outcomes (in the latter case, for example, students whose parents are not as involved in school may be less likely than students whose parents are more involved to attend intervention programs designed to help those most at risk of retention).

**Depth and Quality of Implementation of Promotion Policy.** There are several hypothesized links between student outcomes and various aspects of the promotion policy. Identifying students needing services early in the school year will provide teachers with sufficient time to provide support to these students as well as a chance to try different and better-targeted interventions and provide students with opportunities to gain in-class and out-of-class instructional support to achieve those standards. Outreach to parents via letters, meetings, and workshops helps improve parent involvement and encourages student participation in intervention programs. Saturday and summer programs provide additional instructional time that will help students gain the knowledge and skills they need to achieve grade-level promotion standards. Specific aspects of these out-of-school programs are also intended to facilitate this learning—including smaller class sizes, more personalized attention from teachers, and a dual focus on content and more general study skills. The frequent monitoring of student progress via periodic assessment results and other data is further intended to help teachers and administrators tailor instruction to the specific needs of at-risk students and ensure that no student slips through the cracks.

The act of retaining students is also intended to result in positive student outcomes. First, as we pointed out earlier, the rationale for retention is that repetition of the grade gives students an additional year to master the academic content they failed to master the previous year, and, thus, students should be less at risk for failure when they go on to the next grade. Second, the threat of retention will motivate all students to take responsibility for learning and work harder in school.

The safeguards built into the system in terms of alternative methods by which students can be promoted—portfolios and appeals—help prevent "unjustified" retentions. These safeguards help determine whether students with special circumstances or those who have a difficult time with standardized tests have the ability to master the work demanded in the higher grade. They also help ensure that such students are promoted to the next grade and avoid possible negative consequences of being retained.

In theory, this is how the promotion policy is expected to unfold. In practice, we know from prior research that the outcomes depend on the depth and quality of implementation and on students' actual experiences. For instance, it is likely that the level of student participation and engagement, the quality of the enacted curriculum and instruction, the amount of instructional time, the timeliness in the receipt of student data, and many other factors will contribute to students' experiences and the outcomes of the support services. It is also possible that the extent to which decisions about retention and appeals are based on data will also affect the quality of those decisions and the extent to which students who truly are meeting promotion standards are permitted to advance to the next grade.

**Support Infrastructure.** Research underscores the importance of the broader organizational environment in enacting policies, especially district provision of support and resources to promote reform (Berends, Bodilly, and Kirby, 2002; Marsh, 2002; Elmore and Burney, 1999; McLaughlin and Talbert, 2002; Rosenholtz, 1991; Snipes, Doolittle, and Herlihy, 2002; Spillane, 1996; Togneri and Anderson, 2003). The district can facilitate and foster change by providing guidance (e.g., on the selection of curricular materials or interventions, on rules and regulations regarding the policy), professional staff development, materials and financial resources, human resources support (e.g., hiring staff), data and data support (e.g., assessments, data reports, technology), and networking opportunities (e.g., opportunities for staff to share and learn about best practices). In the case of NYC, as we discussed earlier, this broader infrastructure of support comes not only from a single district entity, but from a host of potential sources, including NYCDOE, regional and district offices, and, in later years of the reform, SSOs. Again, the depth and quality of that support is likely to influence the depth and quality of the implementation of the promotion policy and programs. For instance, the content, frequency, and quality of training provided to teachers might affect their delivery of instruction to in-need students in and outside of class.

**School Context.** The school itself represents an important context that can influence student outcomes directly or indirectly. In terms of indirect influences, the quality of school leadership, the nature of the school climate (including teacher-teacher relations and student-teacher relations), the nature and quality of staff (including experience level, skills, attitudes, and beliefs about students and retention and interest and incentives for teaching on Saturdays or during the summer), composition of the student body (e.g., percentage of at-risk students), and the school's accountability status under NCLB may affect the implementation of support services and the ways in which retention and appeals decisions are made. For instance, the extent to which support services in and outside of school are coordinated may depend on having a strong instructional leader who knows how to bring that coherence to multiple programs, or collegial staff who regularly communicate about their students. Further, a school that has been identified for improvement under NCLB may operate under an additional set of pressures to improve student performance in all tested grades or among Level 2 students (who are most likely to influence the school's adequate yearly progress, or AYP), as opposed to just the promotion gateway grades or Level 1 students (who are most at risk of retention). These competing pressures may affect the attention and resources given to promotion-grade teachers, students, and programs.

The school is also likely to affect student achievement directly because it is the primary source of instruction for all students and because schools often provide additional services to at-risk students, such as other extended learning programs that are not associated with the promotion policy (e.g., after-school tutoring) and socio-emotional supports (e.g., referrals to social service agencies). The impact of these services will depend, in part, on whether students take advantage of and attend these extra services, particularly when they are optional.

**Federal, State, and City Context.** Finally, the framework for the study recognizes that the promotion policy and the various programs and supports for students are embedded in a broader federal, state, and city context that can influence their implementation and effects.

As mentioned earlier, the reforms adopted by NYC were enacted against a backdrop of substantial reform activity undertaken by the state. New York State's Board of Regents is responsible for the general supervision of all educational activities in the state and oversees the New York State Education Department (NYSED). At an education summit held in 2005, the regents outlined the challenges facing the education system and formulated goals and a plan for action. The regents set two primary goals: to raise student achievement across the state and to close the achievement gap among various racial/ethnic groups, between ELL and non-ELL students, and between students with and without disabilities. To help schools achieve these goals, NYSED invested an additional $3.4 billion in schools over a two-year period and adopted several reforms, including establishing a grade-by-grade curriculum that set clear expectations and was better linked to state standards, encouraging schools to focus on professional development, and helping teachers become "highly qualified" under NCLB. These state efforts provided additional reinforcement for the city's efforts to reform the schools and improve student achievement. This context makes an evaluation of NYC's promotion policy and the attribution of results particularly challenging.

In addition to the reforms undertaken by the city and the state, the federal push for greater accountability has added another set of pressures to which schools need to respond. In response to what was perceived as the continued failure of the public education system to educate students (particularly disadvantaged students) to high standards, Congress passed the No Child Left Behind Act of 2001 (Pub. L. 107-110). NCLB requires states to adopt standards-based accountability systems that set challenging content and performance standards for all students.[5] To ensure that students are meeting these standards, states must test all students annually in grades 3 through 8 in reading and mathematics and establish AYP goals for all students and groups of students (e.g., disadvantaged students, students with disabilities, ELL students). By the end of 12 years, all schools in the state should have reached 100-percent proficiency, meaning that all the students in the school passed the state test. While each state is required to meet the 100-percent proficiency goal and demonstrate AYP toward that goal, the federal government has allowed states to develop their own standards, assessments, proficiency levels, and annual AYP goals.

Schools that fail to make AYP for two consecutive years are identified as being "in need of improvement." NCLB requires states to establish a set of escalating interventions and sanctions for schools in need of improvement that receive federal funds under Title I of the Elementary and Secondary Education Act (ESEA, Pub. L. 89-10, 1965).[6]

Thus, schools are faced with a different set of incentives that require them to focus on improving the proficiency levels of students, particularly those close to the proficient level, rather than simply on students who are failing to meet the promotion standards, which are set below the NCLB standard.

---

5 Content standards should specify what students should know and be able to do in core subjects (including, at least, mathematics and reading), contain coherent and rigorous content, and encourage the teaching of advanced skills. Performance standards operationalize and further define content standards "by connecting them to information that describes how well students are doing in learning the knowledge and skills contained in the content standards," according to the earlier Improving America's Schools Act (Pub. L. 103-382, 1994, §1111).

6 Prior iterations of the reauthorization of ESEA placed requirements only on schools that received ESEA Title I funding. NCLB expanded the authority of the federal law by requiring that all schools and students in a state be held to the same standards and be tested on those standards. However, as allowed under federal law, in New York, sanctions under NCLB apply to Title I schools only.

## Major Research Questions

This conceptual framework underpins our analyses and data-collection efforts. It led to seven research questions, around which the next several chapters are organized.

1. What types of supports did schools, SPAs/Saturday schools, and SSAs provide to students? (Chapters Five and Six)
   - What were the factors that enabled or hindered the implementation of the services provided to students?
   - To what extent, if any, were student outcomes linked to the types of schools and to the strategies used to improve achievement?
2. What were the trends in student achievement for students held to the policy? (Chapter Seven)
   - Which students were affected by the promotion policy, and what were their 5th-grade achievement and promotion outcomes?
   - How did retained and promoted students perform in future grades?
3. Relative to comparable groups of students, how did 5th graders needing services at the beginning of the year perform on the spring assessments? (Chapter Eight)
   - What was the relationship between SPA attendance and performance on the spring assessments?
4. Relative to comparable groups of students, how did students at risk of retention perform on the summer assessments? (Chapter Eight)
   - What was the relationship between SSA attendance and performance on summer assessments?
5. Relative to comparable groups of students, what were the future academic outcomes of at-risk students? (Chapter Nine)
   - What is the impact of the promotion policy on at-risk students over time?
6. What were the socioemotional outcomes of students subject to the promotion policy? (Chapter Ten)
   - For students at risk of retention, what was the relationship between socioemotional status and meeting promotion standards via the summer assessment or portfolio review?
7. What are the lessons learned about the design and implementation of policies from NYC's experience and those of other states and districts with similar promotion policies? (Chapter Eleven)

## Summary

As part of the Children First citywide initiative, NYCDOE implemented a test-based promotion policy for students in grades 3, 5, 7, and, most recently, 8. General education students in these grades are required to meet the promotion criteria on the ELA and mathematics assessments (i.e., score at Performance Level 2 or higher on both assessments) in order to advance to the next grade. Key components of the policy include identifying struggling students early in the school year, providing them with additional instructional time, and continuously monitoring their progress. Students who have been identified as needing services at the beginning of the school year are to receive additional instructional support in school. Schools may also offer a variety of out-of-school support services, including Saturday school. Students who fail to meet the promotion standards on the mathematics or ELA assessments in the spring are encouraged to enroll in summer school to receive additional instruction. The promotion policy is not based on a sole criterion: Students may demonstrate basic proficiency in ELA and mathematics through their performance on standardized tests administered during the school year or in August, through a review of a portfolio of their work, or by appeal. Administrators charged with implementing the promotion policy received a host of supports from NYCDOE, the region, and SSOs.

Our conceptual framework embeds this promotion policy in the larger context of federal and state reforms and outlines a number of factors that influenced the policy's implementation as well as expected outcomes. On top of the city's efforts, the state invested substantial resources in supporting the schools and adopted several reforms, including a well-defined curriculum aligned to state standards and an emphasis on teacher qualifications and professional development. Schools are also facing pressures from the NCLB requirements to bring all their students to the proficiency level and show improvements every year. The framework informs a set of research questions focusing on implementation and student outcomes, which are addressed in the subsequent chapters.

CHAPTER FOUR

# Data and Methods

*Sheila Nataraj Kirby, Louis T. Mariano, and Jennifer Sloan McCombs*

## Data

The conceptual framework outlined in the previous chapter guided our data-collection efforts. We collected data from a variety of sources over a three-year period. These sources included the following:

- interviews with RDISs and SSO leaders
- case studies of a small sample of schools, SPAs, and SSAs
- administrator surveys
- surveys of selected samples of students
- data on school characteristics
- longitudinally linked data on several cohorts of 5th- and 3rd-grade students.

Next, we discuss each of these sources in greater detail.[1]

### Interviews with Regional Directors of Instructional Services and School Support Organization Leaders

In 2007, we conducted telephone interviews with the 10 RDISs using a semistructured protocol. Interviews lasted between 30 and 60 minutes. Directors were asked about their backgrounds; roles and responsibilities; guidance provided by the region to schools regarding AIS; other types of assistance the region provided to the schools in terms of curricula, assessments, AIS materials for identified students; and professional development for AIS leaders and other instructional staff, as well as some of the successes and challenges they faced in providing AIS.

[1] Although not discussed here, to situate the NYC policy in a larger context, we also collected data on promotion policies implemented in other states and districts that were similar to NYC's policy. We conducted two rounds of interviews with representatives from a combined sample of six states and 12 districts in 2006 and 2008. In both rounds of interviews, respondents included officials who oversaw and monitored some or all aspects of the promotion and retention policies and programs in their state or district. In both years, we also reviewed available promotion policy documentation on state and district Web sites (e.g., policy documents, parent guides), as well as other relevant studies, newspaper articles, and reports that we gathered or that respondents provided to us. The results of this effort are detailed in Marsh et al. (2009).

In 2008, we interviewed representatives from six SSOs using a semistructured protocol that was similar in scope to the one used with the RDISs. Interviews lasted between 30 and 60 minutes and focused on the role of the SSO and the types of assistance provided to the school.

### Case-Study Site Visits

In the spring and summer of 2006, 2007, and 2008, we conducted case-study visits to schools, SPAs, and SSAs. During the school visits, we interviewed principals, assistant principals, or administrators and AIS leaders and conducted focus groups with 5th-grade teachers. Although we attempted to interview all 5th-grade teachers, this was not always possible in some schools, SPAs, and SSAs. We asked principals, AIS leaders, and teachers about how students were selected for AIS, services provided, professional development, and enablers and challenges to providing AIS and to their schools in general. SPA and SSA administrators were asked about enrollment, attendance, curriculum, communication with the home school, and enablers and challenges in administering their programs.

### Selection of Case-Study Sites

**Schools.** In 2006, we focused the case-study sample on schools with large numbers of at-risk students that appeared to be performing better than comparable schools, using data provided by NYCDOE. We selected schools that had at least 200 students who took the spring assessments, in which 80 percent or more of students were eligible for free or reduced-price lunch, and that were ranked in performance category 4 or 5 in ELA according to the 2004 results published by NYCDOE, which were based on the city assessment in grades 3, 5, and 6. Two regions had no schools that met the third criterion; in these regions, we included schools ranked in performance category 3.

In 2007, we selected 10 case-study schools. We limited the sample pool to schools with 5th-grade enrollments between 40 and 100 students to eliminate outlier schools. We used the percentage of students scoring Level 1 in mathematics in 2005 to categorize schools as "low," "medium," or "high" on this distribution. We purposely sampled five schools with a medium percentage of Level 1 students and five schools with a high percentage of Level 1 students. We also sampled schools representing the 10 regions.

In 2008, because schools were given the responsibility for deciding whether and how to provide Saturday and summer school services, we increased our school case-study sample to 23 schools. At these schools, we asked to interview supervisors of these services, along with principals, teachers, and intervention leaders. In selecting schools for our site visits, we excluded the following types of schools: schools that participated in case studies for the study in prior years (to minimize respondent burden), schools with fewer than 40 or more than 120 5th-grade students, and schools with fewer than four students retained in 5th grade in 2005–2006. From this group, we randomly selected case-study schools to get a proportional representation by type of SSO partner:

14 schools that had partnered with LSOs, seven schools that had partnered with the ESO, and two schools that had partnered with PSOs.

**Saturday Preparatory Academies.** In 2006, the delayed start of the study precluded our conducting any case studies of SPAs. In winter 2007, we visited 10 randomly selected SPAs—one per region. As noted earlier, in 2008, schools made their own decisions about Saturday school, so we asked a set of questions about Saturday school provision during our 2008 school visits.

**Summer Success Academies.** In summer 2006, we randomly selected two SSAs to visit in each of the 10 regions. In summer 2007, our original sample plan called for one SSA per region—a plan that was made moot by the dismantling of the regional structure as of July 1, 2007. We purposely sampled four SSAs that operated their own stand-alone programs. These sites were members of the empowerment zone during the 2006–2007 school year. We also purposely sampled three schools that had initially indicated to NYCDOE that they would run a shared program and three schools that had initially indicated that they would use a shared-space model. However, our visits revealed that all six of these sites ended up running a shared program. Again, in 2008, we asked schools about their plans to provide SSA.

Table 4.1 shows the number of case-study sites visited, the number of sites from which we collected each type of data, and the total number of individuals interviewed at each site.

### Administrator Surveys

We administered Web-based surveys to principals, AIS team leaders, and SPA and SSA administrators in all three years. In each year, we attempted to reach administrators of all the schools in our sample with a 5th grade. Table 4.2 shows the timing and response rates for the surveys.

Principals and AIS team leaders were asked about the full range of instructional support services provided to at-risk students and how these services were implemented, monitoring of student progress (including through formative assessments), the training and resources provided to teachers, outreach activities offered to families of at-risk students, and the challenges faced by principals and school personnel.

SPA and SSA administrators were asked about the curriculum used, the availability and adequacy of curriculum materials and alignment with regular school curriculum, the use and allocation of time, class size, perceptions of teacher quality, methods of ongoing monitoring, training provided to teachers, and resources provided by the region or district. In addition, all respondents were asked about the perceived effects of the promotion policy.

In 2008, principals were responsible for in-school services provided to at-risk students, as well as Saturday school and SSA services. Given space constraints, we chose to focus the majority of questions on in-school service provision while including a smaller subset of questions about Saturday school and SSA administration.

**Table 4.1**
**Data Collected from Case-Study Sites**

| Site Type | 2006 | | 2007 | | 2008 | |
|---|---|---|---|---|---|---|
| | Number of Sites | Number of Interviewees/ Observations | Number of Sites | Number of Interviewees/ Observations | Number of Sites | Number of Interviewees/ Observations |
| Schools | 7 | | 10 | | 23[a] | |
| Principal interviews | 7 | 7 | 9 | 9 | | 21[b] |
| Assistant principal interviews | 7 | 6[b] | 7 | 7 | | |
| 5th-grade teacher focus groups | 7 | 22 | 10 | 33 | | 77 |
| AIS leader interviews | 7 | 6[b] | 10 | 23 | | 20 |
| Intervention teacher interviews | 7 | 3[b] | 2 | 2 | | 40 |
| SPAs | NA[c] | | 10 | | | |
| Site supervisor interviews | | | 10 | 11 | | 12 |
| 5th-grade teacher focus groups | | | 10 | 28 | | |
| Intervention teacher interviews | | | 2 | 2 | | |
| SSAs | 20 | | 10 | | | |
| Site supervisor interviews | 20 | 20 | 10 | 14 | | 20 |
| 5th-grade teacher focus groups | 19 | 49 | 10 | 17 | | |
| Intervention teacher interviews | 17 | 20 | 8 | 9 | | |

[a] In 2007–2008, principals had the authority to decide whether and how to offer Saturday and summer school services. As result, we increased the sample size of case-study schools and interviewed those who were identified as having responsibility for these services.

[b] At some sites, we were not able to interview certain school staff because they were unavailable.

[c] The study started too late in the year to overlap with the SPA session.

**Table 4.2**
**Web-Based Surveys of Administrators and AIS Team Leaders: Timing and Response Rates**

| Survey Subjects | Period | Number Surveyed[a] | Number of Respondents | Response Rate (%) |
|---|---|---|---|---|
| Principals of all schools with a 5th grade | July–August 2006 | 696 | 440 | 63.2 |
| AIS team leaders in all schools with a 5th grade[b] | May–June 2007 | 631 | 295 | 46.8 |
| Principals of all schools with a 5th grade[c] | July–September 2008 | 673 | 315 | 46.8 |
| Administrators of SPAs serving 5th-grade students[d] | April–May 2007 | 144 | 106 | 73.6 |
| Administrators of SSAs serving 5th-grade students[e] | July–August 2006 | 253 | 89 | 35.2 |

[a] Number of eligible respondents who were surveyed.

[b] We had email addresses for 663 AIS team leaders, of whom 32 were ineligible because they had left the school or were no longer working in a school with a 5th grade.

[c] We had email addresses for 708 principals of schools with a 5th grade, of whom 35 were ineligible because they had left that position or were no longer working in a school with a 5th grade.

[d] We also fielded an SPA administrator survey in 2006; however, response rates were quite low (23 percent) because of the unavoidable concurrent fielding of the principal, SPA administrator, and SSA administrator surveys due to the delayed start date of the study. Due to the very low response rate, we do not present results from the 2006 SPA survey.

[e] In 2007, because of the dissolution of the regions, schools were responsible for the SSA program. NYCDOE was unable to provide us with the email addresses of the SSA administrators that we needed to conduct the planned 2007 online survey.

We were particularly interested in how schools with relatively low and high percentages of students needing services at the beginning of the school year implemented the promotion policy and the challenges they faced in doing so. We defined students needing services as those scoring Level 1 or low Level 2 on the prior year's assessment and used the population distribution of this variable to categorize respondent schools into quartiles.[2] We then used post-stratification weights based on these quartiles to weight the respondent schools to the full population.

We did not have data on the characteristics of SPAs and SSAs, and, thus, we were unable to investigate or correct for differential nonresponse. As such, any findings reported from these two surveys need to be viewed as suggestive, rather than conclusive, and as representative of the responding group, rather than the population.

---

2 To determine the threshold for low Level 2, we examined the standard errors of the scale scores in each year and used the following criterion: any score that was within two standard errors of the upper boundary of Level 2 was considered "high," and the rest were considered "low." In general, this served to close the midpoint of the range of Level 2 scores. However, this method is more defensible because it uses a confidence-level interpretation. "High" Level 2 scores include those that are statistically indistinguishable from the cutoff point of Level 3.

### Student Surveys

A key component of the evaluation was monitoring the socioemotional status of students held to the promotion policy through student surveys conducted over three consecutive years. Data from the surveys provided a basis for assessing students' attitudes toward school, ELA, and mathematics (the subjects included in the promotion policy) and the relationship between student attitudes and meeting the promotion criteria.

Student surveys were administered in the fall of the 2006–2007, 2007–2008, and 2008–2009 school years. Delays in the receipt of data pushed the fielding of the first two surveys to December and January of those school years, while the 2008–2009 survey was fielded in November 2008.

**Survey Instrument.** In developing the survey, we used three socioemotional measures with a strong theoretical basis that had been tested in large-scale research and did not overburden respondents:

- a scale for attitudes and beliefs about school belonging, Goodenow's Psychological Sense of School Membership (Goodenow, 1993)
- a scale for attitudes and beliefs about mathematics from the Third International Mathematics and Science Study (TIMSS), 2003 (Martin, Mullis, and Chrostowski, 2004)
- a scale of attitudes and beliefs about reading, also from the 2003 TIMSS (Martin, Mullis, and Chrostowski, 2004).

These survey measures are discussed in greater detail in Chapter Ten.

**Survey Design.** The objective of the survey was to understand differences in the socioemotional status of three groups of students held to the promotion policy—at risk and retained, at risk but promoted, and not at risk—in four cohorts: 2005–2006 and 2006–2007 5th graders and 2003–2004 and 2004–2005 3rd graders. These cohorts were mapped to the grade in which they would be at the time of the survey, and students in those grades then constituted the sampling population. To ensure that we reached both the retained and promoted students in the 2006–2007 survey, grades 4, 5, and 6 were surveyed. The following year, 2007–2008, we surveyed grades 5, 6, and 7 and, in the fall of 2008, grades 6, 7, and 8.

The sampling frame consisted of all schools in districts 1 through 32 that had at least one of these grades, according to NYCDOE; a school was excluded from the sampling frame if it did not have any students in at least one of the four survey cohorts who were at risk of retention in the cohort policy year. A two-stage sampling design was used, with the objectives of minimizing the number of participating schools while selecting a representative sample from each cohort. In the first stage, schools were selected with probability proportional to numbers of students at risk of retention in the year in which each cohort was subject to the promotion policy, and sampling was stratified by school grade range.[3] In the second stage, for grades that represented the current grade of promoted students (e.g., grades 5 and 6 for the 2006–2007 survey), all eligible classrooms in a sampled school were selected to complete the survey. For the grade that represented the current grade of retained students only (e.g., grade 4 for the 2006–2007 survey), all retained students in the sampled schools were included.

The final samples for the surveys were as follows:[4]

- 2006–2007: 130 schools, ~17,500 students
- 2007–2008: 159 schools, ~22,500 students
- 2008–2009: 183 schools, ~27,500 students.

We increased the target sample size over time to account for students in mainstream classrooms who entered the system after 5th grade, the unexpectedly large proportions of students in mainstream classrooms who were not held to the policy, and the somewhat lower-than-expected participation rate. Approximately 15 percent of students in the 5th- and 6th-grade classrooms sampled in 2006–2007 were members of at least one of the four survey cohorts but were not held to the promotion policy in either 3rd or 5th grade. Students in this category included special education and ELL students in that status for three or fewer years.

**Survey Administration.** We provided detailed administration instructions to school personnel who administered the survey, along with information for parents (including an option to decline participation on behalf of the student). In the first two years, due to data delays, the survey-fielding period extended through January. This was problematic because the fielding period overlapped with the ELA assessment in both these surveys. However, analyses of responses received before and after the ELA assessment did not show differences between the two groups. In the third year, the fielding period covered the first three weeks in November and thus avoided overlap with the ELA assessment.

**Response and Participation Rates.** School participation rates fell over time across the three surveys (see Table 4.3). The participation rate was close to 70 percent in the first year, 59 percent in the second year, and 54 percent in the third year. The decline in years 2 and 3 occurred even though schools were offered a $150 honorarium for

---

3 Grade-range stratification was used to acknowledge the possibility that similar students in schools that served different grade ranges might have differing experiences with respect to the promotion policy.

4 In the first two years, because of a delay in the receipt of data and the planned fall survey administrations, complete school population data were not available in time for sampling. Therefore, each school's population with respect to each subcohort to be sampled was estimated using the data from the prior year. Because of the unanticipated sharp decline in the number of retained students in the 2006–2007 5th-grade cohort, the distribution of students retained from the 2005–2006 5th-grade cohort was not a good estimate for the distribution of students retained in the 2006–2007 5th-grade cohort. This caused a serious shortfall in the number of retained students from that cohort represented in the sample in the second survey year.

**Table 4.3**
**Survey Sample and Usable Responses, 2006–2007, 2007–2008, and 2008–2009 Student Surveys**

| Student Sample Characteristic | 2006–2007 | 2007–2008 | 2008–2009 |
|---|---|---|---|
| Number in sample | 17,497 | 22,459 | 27,500 |
| Number of eligible respondents in sample[a] | 14,508 | 17,813 | 21,268 |
| Number of eligible responses | 8,109 | 7,415 | 8,973 |
| Number of usable responses[b] | 7,982 | 7,260 | 8,769 |
| Adjusted student response rate (%) | 55.0 | 40.8 | 41.2 |

[a] Eligible students are defined as students who were held to the promotion policy in at least one of the four survey cohorts.

[b] Usable responses are defined as complete and partially complete surveys.

participating in and returning the student surveys, which was not offered in year 1. We calculated an adjusted response rate as the number of eligible respondents divided by the number of eligible students in the sample; the number of responses did not include incomplete surveys (in which less than 75 percent of items in each of the three scales were answered). The percentage of usable surveys was 55 percent in 2006–2007, 41 percent in 2007–2008, and 41 percent in 2008–2009.

One reason for the decline in participation may have been that the sample contained middle schools as well as elementary schools in the second survey year and consisted almost entirely of middle schools in the third survey year. These schools and students may have been less motivated to participate in a survey whose purpose was to track the impact of the 5th-grade promotion policy—indeed, from the responses we received, some schools felt that they were targeted for the survey in error because they did not have any 5th graders. A second reason may have been the greater autonomy of principals, which resulted from the restructuring of the school system in the summer of 2007. As a result, principals were less responsive to NYCDOE's efforts to encourage them to respond. While our total sample sizes remained fairly substantial, it is difficult to know whether and to what extent this decline in participation affected our findings.

Because estimates of the distribution of at-risk students across schools in each survey cohort were used in preparing the sample, we used post-stratification weighting of the survey responses to ensure that the actual distribution of at-risk students was properly represented in the survey analyses. For each of the four survey cohorts, a set of weights was developed for each subcohort: at risk and retained, at risk but promoted, and not at risk. We examined the distribution of each subcohort in each grade-range stratum across schools, dividing the distribution into meaningful subsets. The responding students in each subset were then weighted to represent the full subcohort population in that subset of schools. Because all eligible students in a sampled school were included in the sample, all responding students in the same subcohort within a school were assigned the same weight.

### Data on School Characteristics

NYCDOE provided school-level data on all schools with a 3rd or 5th grade for the 2002–2003 through 2007–2008 school years. These data included school classification, student characteristics, and school performance variables. School classification information included location descriptors, such as district and region, as well as grade span and dates of operation. School-level student descriptors included student enrollment, student attendance, the racial/ethnic composition of the student body, and the number and percentage of students who were in special education, had ELL status, or were eligible for free or reduced-price lunch. The student descriptors were available school-wide and by grade. Available school performance information included AYP status under NCLB and Title 1 status.

### Longitudinally Linked Student Data

The overall purpose of the student achievement analysis was to understand differences in the outcomes of students who were in need of services at the beginning of the promotion-policy year, students identified as at risk of being retained during the policy year, and retained students relative to their peers who were not at risk or who were at risk but promoted. Our study allowed us to follow these students longitudinally. We also specifically considered whether participation in SPA or SSA helped improve outcomes.

We tracked three cohorts of 5th graders held to the promotion policy (2004–2005, 2005–2006, and 2006–2007—the P1, P2, and P3 cohorts, respectively) and a comparison cohort (2003–2004—the C0 cohort) that was not held to the policy.[5] NYCDOE provided data on these cohorts for the 2001–2002 through 2007–2008 school years, when applicable.[6] The individual student data included the following:

- demographics (gender, race/ethnicity, ELL status, special education status, prior retention status, school, grade, free-lunch status, attendance information)
- performance on the spring ELA and mathematics assessments (both scale scores and proficiency-level scores in prior years, the cohort year, and future years, through 2007–2008)

---

[5] We also tracked outcomes for two 3rd-grade cohorts held to the policy (2003–2004 and 2004–2005) and one comparison cohort (2002–2003). Note that many students in the 3rd-grade cohorts were also students in the 5th-grade cohorts.

[6] Limited 2008–2009 information was provided at the beginning of that school year for the purpose of administering the fall 2008 student survey.

  - performance on summer assessments and portfolio performance (where relevant) for students failing the proximal-year spring assessments and so at risk of retention
- final promotion disposition and means of promotion (meeting standards on the spring or summer assessments, spring portfolio, August portfolio, or appeal)
- SPA and SSA data for those eligible for these intervention services:
  - SPA data (enrollment and SPA site for the three policy cohorts, and individual attendance data for P2 and P3 cohorts)
  - SSA data (attendance and SSA site for the C0, P2, and P3 cohorts).[7]

We now turn to a description of the methods used to analyze the data and address the major research questions underpinning the study. Table 4.4 maps the research questions to the data and methods used to address them.

## Methods

This section provides a brief overview of the methods we used to analyze the implementation of the promotion policy and the effects of key elements of the promotion policy on student achievement. More detailed explanations are provided later in this monograph, when we present results of the analyses.

### Analysis of Implementation

To analyze implementation, we compared the city's guidance on implementing the promotion policy with how the policy was actually enacted by regions, districts, and schools, using data from our case studies and administrator surveys, as well as descriptive data provided by NYCDOE on the total numbers of SPAs and SSAs, the average number of schools served, enrollment, and attendance. We coded the notes and transcripts from interviews, focus groups, and observations along the dimensions outlined in the conceptual framework. We examined how the ability of the school to provide services, the types of strategies used to improve student achievement, priorities for investing school resources, and factors identified as barriers and enablers of policy implementation varied by the percentage of students needing services or at risk of retention. We used simple cross-tabulations of data and Pearson's chi-square test to determine whether these relationships were statistically significant. We then integrated findings from different data sources and across years to identify overarching themes regarding the nature, quality, perceived impact, and potential barriers and enablers of policy implementation.

---

7 Recall that eligible students in the comparison cohort were offered services through summer school, while students in the policy cohorts were offered services through the SSAs. Note that individual SPA and SSA attendance data were unavailable for the P1 cohort.

**Table 4.4**
**Mapping Research Questions to Data and Methods**

| Research Questions | Data | Methods |
| --- | --- | --- |
| 1. What types of supports did schools, SPAs/Saturday schools, and SSAs provide to students?<br>What were the factors that enabled or hindered the implementation of the services provided to students?<br>To what extent, if any, were student outcomes linked to the types of schools and to the strategies used to improve students' achievement? | Data collected from administrator surveys, case studies, and interviews<br>Administrator survey data linked to student-achievement data<br>School-level data linked to student-achievement data | Qualitative research methods (coding of focus groups, interviews, and observations)<br>Document review<br>Cross-tabulations of weighted survey data and statistical tests for differences in means and proportions<br>Logistic regression models to explore links between students' ability to score Level 2 on the spring assessment and school characteristics and school strategies<br>Ordinary least squares (OLS) regression models to explore links between student performance and school characteristics and school strategies |
| 2. What were the trends in student achievement for students held to the policy?<br>Which students were affected by the promotion policy, and what were their 5th-grade achievement and promotion outcomes?<br>How did retained and promoted students perform in future grades? | Longitudinally linked data on student achievement and student characteristics | Descriptive statistics |
| 3. Relative to comparable groups of students, how did 5th graders needing services at the beginning of the year perform on the spring assessments?<br>What was the relationship between SPA attendance and performance on the spring assessments? | Longitudinally linked data on student achievement and student characteristics, including SPA attendance | Propensity score weighting using data from the comparison cohort not subject to the policy to define appropriate control groups<br>Doubly robust regression models to estimate the net effect of the "treatment" on students' spring outcomes<br>Generalized additive mixed models to examine the effect of SPA attendance on students' spring outcomes |

**Table 4.4—Continued**

| Research Questions | Data | Methods |
| --- | --- | --- |
| 4. Relative to comparable groups of students, how did students at risk of retention perform on the summer assessments? | Longitudinally linked data on student achievement and student characteristics, including SSA attendance | Propensity score weighting using data from the comparison cohort not subject to the policy to define appropriate control groups |
| What was the relationship between SSA attendance and performance on summer assessments? | | Doubly robust regression models to estimate the net effect of the "treatment" on students' summer outcomes (relative to the prior summer program) |
| | | Generalized additive mixed models (GAMM) to examine the effect of SSA attendance on students' summer outcomes |
| 5. Relative to comparable groups of students, what were the future academic outcomes of at-risk students? | Longitudinally linked data on student achievement and student characteristics | Propensity score weighting using data from the comparison cohort not subject to the policy to define appropriate control groups |
| What is the impact of the promotion policy on at-risk students over time? | | Doubly robust regression models to estimate the net effect of the "treatment" on students' future outcomes |
| | | RDD models to examine the effect of retention on students' future outcomes |
| 6. What were the socioemotional outcomes of students subject to the promotion policy? | Student surveys linked to student characteristics and student-achievement data | Pairwise comparisons of weighted means and statistical tests of significance of differences in means |
| For students at risk of retention, what was the relationship between socioemotional status and meeting promotion standards via the summer assessment or portfolio review? | | OLS linear regression or logistic regression, where appropriate, with standard errors adjusted to account for grade range stratification and clustering at the school level |

NOTE: As mentioned earlier, the study addressed seven major research questions. The seventh research question focused on lessons learned regarding promotion policy design and implementation from states and districts other than NYC that had implemented similar test-based promotion policies. The results of that effort are detailed in Marsh et al. (2009). This monograph focuses on the six research questions presented in the table but does include the important findings from Marsh et al. (2009).

We use logistic regression models to explore the association between school supports for students needing services and students' ability to pass the spring assessments. We also use OLS regression to explore the association between school supports and student achievement more broadly. We originally considered estimating school-level outcome measures, such as the percentage of students needing services who met the promotion standards on the spring assessments. However, the small sample sizes of students needing services or at risk in many of these schools meant that the school outcome measures were extremely sensitive to one or two students making (or not making) the cutoff.

### Analysis of the Effects of Supportive Interventions and Retention on Academic Outcomes

Short of a randomized experiment, the best methods for producing causal inferences are those that employ quasi-experimental designs comparing outcomes in the treatment group to those in a plausibly similar control group. Given that we had observational data on NYC students, modeling the effects of retention and supportive interventions—in particular, SPAs and SSAs—became challenging. Therefore, we used two quasi-experimental methods, RDD and doubly robust regression, to exploit available comparison groups in evaluating the impact of the promotion policy on proximal and future outcomes. To complement these designs, we used GAMM to explore the relationships between attendance at supportive interventions and proximal outcomes. These methods are discussed in more detail later in this chapter. Technical details are provided in Appendix A.

First, we discuss the issue of same-age versus same-grade comparisons and offer a rationale for our decision to use same-grade comparisons.

**Same-Age Versus Same-Grade Comparisons.** When considering the effects of the policy on future outcomes, an additional challenge arises in terms of how best to identify an appropriate comparison outcome relative to the outcomes of those being treated. In selecting a comparison outcome for the retained students, we sought to determine how these students would have performed in the absence of treatment under the policy—in this case, retention. Both same-age and same-grade comparison strategies are found in the retention literature, and they inform different variants of the counterfactual. The same-age strategy compares the outcomes of retained students to their expected outcomes had they not been retained, at a particular number of years after the promotion decision. In other words, it assesses whether retained students are performing better at a fixed point in time than would have been the case if they had not been retained. Of course, after the promotion decision, retained and promoted students would not be in the same grade; thus, same-age comparisons require outcome measures that are vertically equated across grades such that scores from adjacent grades may be interpreted on a common scale. Alternatively, the same-grade strategy compares the outcomes of retained students to their expected outcomes had they not been retained, but at a particular number of grades beyond the promotion decision. In other words, it assesses whether retained students are performing better in a given grade than would have been the case if they had not been retained. Retained students would, of course, reach a future grade one year later than if they not been retained; thus, same-grade comparison requires that the outcome measures in the grade being compared are equated across years. Each of these alternative comparison strategies addresses an interesting and different question.

From a data-availability standpoint, same-age comparisons were not feasible our analyses. As discussed, the city assessment program, which was vertically scaled, ceased administering spring assessments after 2005, when cohort P1 was in 5th grade. The state spring assessment program that replaced it in 2006 was not vertically scaled; however, scores in each grade were equated across years and fully support same-grade comparisons through 8th grade.

From an application standpoint, we believe that the counterfactual supported by the same-grade comparison is consistent with the theory of action on which the retention treatment is designed. The retention treatment was put in place as part of the promotion policy so that students not meeting the 5th-grade learning standards would have an additional year to learn those standards and, thus, be better prepared to succeed in 6th grade and beyond. A same-age comparison ignores this extra year of preparation that the treatment is specifically designed to provide, while a same-grade comparison explicitly accounts for the extra year of preparation. Thus, we find the same-grade strategy to be appropriate for the research question.

Earlier in this monograph, we mentioned that the rationale behind promotion policies is that repetition of the grade will give students an additional year to master the academic content that they failed to master the previous year and that, "by catching up on prerequisite skills, students should be less at risk for failure when they go on to the next grade" (Shepard and Smith, 1990, p. 84). Presumably, the ultimate goal is to ensure that the retained student is better prepared at the end of his or her schooling to be a productive member of society—this is a same-grade question. In our analyses, we explored whether retained students were better prepared at an interim grade level. The lack of improvement over expected performance in the absence of retention in the interim grade yields little hope of future positive effects from retention, while evidence of interim grade improvement establishes the possibility that retained students may be better prepared at the end of their academic careers.

**Regression Discontinuity Design Models.** An RDD (Imbens and Lemieux, 2008; Shadish, Cook, and Campbell, 2002) compares the relationship between an assignment variable (such as a 5th-grade spring assessment scale score) and an outcome variable (such as the 6th-grade assessment score) for subjects above and below an assignment threshold that determines "treatment" status (such as mandated summer school). Thus, students who score below the threshold are exposed to the intervention, while those who score above are not. Van der Klaauw (2002) showed that if treatment is conditioned on meeting this known threshold, analyzing individuals in the neighborhood of the threshold for selection is similar to a random-assignment design. This approach exploits the fact that subjects included in the comparison and the intervention groups differ only with respect to the assignment variable (and any other variable correlated with it); thus, one controls for the confounding factors by contrasting marginal participants to marginal nonparticipants, where the term *marginal* refers to those units close to the threshold for selection (Battistin and Rettore, 2002). The difference in outcomes between these two groups indicates the mean impact of the intervention locally with respect to the threshold for selection that can be attributed to the intervention or treatment, given that the two groups are very similar prior to the treatment. Cook (2008) summarizes the advantages and disadvantages of RDD versus a randomized experiment. RDD is "less statistically powerful; it involves less transparent assumptions about functional form; its implementation is less well empirically understood; and methods for improving its implementation are less developed" (p. 652). However, the strength of RDD is that it can be used in a particular set of circumstances in which a randomized experiment might not be feasible, and in these circumstances, RDD is superior to all known quasi-experimental methods. In fact, Lee and Lemieux (2009, p. 3) show that if individuals "are unable to precisely manipulate the forcing variable, a consequence of this is that the variation in treatment near the threshold is randomized as though from a randomized experiment."

In our study, there were two policy treatments that lent themselves to the RD design and helped us determine whether the theory of change worked as intended. For each, differences in assessment-level scores that mapped to specific scale-score threshold values on ELA and mathematics assessments implied an increased probability of receiving the treatment:[8]

1. Students scoring below threshold on either 5th-grade spring assessment were classified as at risk of retention and had an increased probability of attending an SSA for additional summer instruction.
2. Students scoring below threshold on the 5th-grade summer assessment had a large increase in the probability of retention.

The basic idea behind the RDD approach is that assignment to the treatment is determined by the value of a predictor being on either side of a fixed threshold (Hahn, Todd, and Van der Klaauw, 2001). In the "sharp" RDD, the assignment to treatment is completely determined by the forcing variable, so all units with a covariate value

[8] Available assessment results from both the city and state used a raw-score-to-scale-score conversion to produce the student scale scores. This design allowed for a one-to-one correspondence between each level score and a range of scale scores, such that the transition between level scores may be identified on the scale. To the extent that scoring Level 1 versus Level 2 corresponds to an increased probability of treatment, the scale score may then be used as an assignment variable in an RDD.

below a given threshold value are assigned to the treatment group and participation is mandatory; all others are in the control group and are not eligible for treatment. In the "fuzzy" RDD, the probability of receiving treatment need not change from 0 to 1 at the threshold, so some individuals who fall below the threshold do not participate in the treatment and some who fall above the threshold do participate. However, scoring below the threshold implies an increased probability of treatment (Imbens and Lemieux, 2008; Hahn, Todd, and Van der Klaauw, 2001). In this case, we identified the treatment effect for the treated students below the threshold by dividing the effect of treatment eligibility on the outcome at the threshold value by the effect of treatment eligibility on the probability of receiving the treatment at the threshold value.

In each of the cases we analyzed, not all students below the threshold received the treatment. For example, a student may avoid being retained by demonstrating Level 2 proficiency via a portfolio. Thus, having an assignment variable score on the treatment side of the threshold does not imply that a student will receive the treatment with absolute certainty. To account for this, we used a fuzzy RDD approach to model the effect of the two treatments.

**Propensity Score Models and Doubly Robust Regression.** Several other studies have used propensity score methods to examine the effects of retention (Hong and Raudenbush, 2005; Hong and Yu, 2007, 2008; Wu, West, and Hughes, 2008). The first three used data from the Early Childhood Longitudinal Study, Kindergarten Cohort, for a large sample of kindergartners in schools with and without retention policies. Each of these studies used hierarchical logistic regression models to estimate a student's probability of being retained as a function of pretreatment covariates and then divided the sample into several strata on the basis of the logit of the propensity score. Hong and Raudenbush (2005) used a similar technique to obtain a comparison group of nonretention schools for their treatment group of retention schools. Wu, West, and Hughes (2008) used data from a sample of at-risk students in three districts in Texas who took independent mathematics and reading assessments over a period of three years, as well as data collected from parents, teachers, and peers. From this group, the authors were able to closely match 97 pairs of retained and promoted students based on their propensity scores using optimal matching procedures. This was a case in which they actually matched one to one.

Our propensity scoring models are somewhat different from the ones used by these previous studies. We did not use stratification or one-to-one matching but instead used a propensity score weighting approach to estimate the effect of retention on student achievement. Propensity score weighting methods reweight the untreated subjects so that the distribution of their characteristics matches the distribution of characteristics of the treated subjects. These models first estimate $P(\text{treated} \mid \mathbf{x}_i)$, the probability that a subject with observed features $\mathbf{X}_i$ would be in the treated group. McCaffrey, Ridgeway, and Morral (2004) show that, by assigning each untreated subject *i* a weight equal to $P(\text{treated} \mid \mathbf{x}_i) / \left(1 - P(\text{treated} \mid \mathbf{x}_i)\right)$, the distribution of the characteristics of the untreated group would be the same as those of the treated group. Note that, for untreated subjects with characteristics that are atypical of treated subjects, $P(\text{treated} \mid \mathbf{x}_i)$ would be near 0 and would receive a weight near 0. On the other hand, untreated subjects with features typical of the treated subjects would receive larger weights. The advantage of this approach over regression models is that it is far easier to diagnose and correct model misspecification in the propensity score models simply by checking whether the treated and untreated groups appear balanced on all the observed features.

A second advantage of propensity score analysis is that the outcomes do not influence how the model adjusts for balance on the observed features. Since the computation of the propensity scores does not use the outcomes, the method avoids the bias that analysts can introduce by fitting a variety of regression models to the outcome. We then weight each case by the propensity score weight, assigning each treated subject a weight of 1 and each control subject a weight $w_i = P(\text{treated} \mid \mathbf{x}_i) / \left(1 - P(\text{treated} \mid \mathbf{x}_i)\right)$ and then estimate a weighted regression model that also includes individual- and school-level covariates and an indicator variable denoting whether the student was from the treated cohort or the comparison cohort. The inclusion of covariates makes the treatment effect estimate "doubly robust" in the sense that if either the propensity score model or the regression model is correct, then the treatment effect estimator will be consistent and the bias will be small (Bang and Robins, 2005). Note that the counterfactual we are using here is "How would the treated students have performed, had they not been treated?" Thus, we are interested in the treatment effect on the *treated*.

**Modeling the Relationship Between Intensity of Treatment and Student Outcomes.** Available data on SPA and SSA participation allowed us to explore the impact of intensity of participation among the treated students. For example, research on out-of-school programs suggests that benefits from after-school participation are related to the frequency and duration of participation, but the extent to which students must participate before positive outcomes are demonstrated is unknown (Baker and Witt, 1996; Bodilly and Beckett, 2005; Gilman, 2001). Furthermore, it is unclear whether the relationships between participation and outcomes are linear, curvilinear, or represent some other functional form (Chaput, 2004). We examined the relationship between program participation and outcomes using data on number of days attended as a proxy for a measure of intensity of SPA and SSA participation and examined how that intensity relates to outcomes. We used a three-level GAMM to model the effect of intensity of treatment on outcomes. This is similar to a three-level hierarchical linear model of student achievement in which SPAs are nested within districts and regions,[9] except that GAMM allows the relationship between attendance and outcomes throughout the range of possible attendance values to be nonlinear. This enabled us to identify potential change points in the relationship between attendance and outcome. For

[9] During the period analyzed, the regions were responsible for the SPA and SSA programs.

example, suppose that there exists some minimum treatment intensity that is necessary before any increase in outcome is realized. If three or fewer sessions are not associated with any increase, and if there are incremental increases beyond that, then the true marginal relationship would be flat until the third session, with an increasing slope thereafter. This would not be captured in the more traditional linear mixed model. Such information might be valuable to program administrators in designing these types of intervention programs.

We should note that there is a natural hierarchy with respect to the internal validity of the three analytical methods used here. RDD and doubly robust regression are both quasi-experimental methods and attempt to equalize the selection bias present in treatment and comparison groups so as to yield unbiased estimates of treatment effects. As noted earlier, RDD is superior to all quasi-experimental methods with respect to internal validity. Doubly robust regression methods cannot control for unobserved confounders, while RDD models do. Indeed, in Chapter Eight, we discuss the need to adjust our doubly robust regression estimate to account for the presence of city and state policies that were enacted concurrently with the promotion policy. However, RDD models are limited in their generalizability in that they provide only a local average treatment effect (i.e., the effect of the treatment on subjects who are close to the treatment threshold), whereas doubly robust regression methods provide an estimate of treatment effects generalized to all subjects who would typically be subject to the treatment—the so-called average treatment effect on the treated (Wooldridge, 2001; McCaffrey, Ridgeway, and Morral, 2004). Thus, the two methods are complementary. On the other hand, the data do not support inferences regarding causal effects from the GAMM analyses. Because the attendance data are observational and students self-select their level of participation, the GAMM analyses represent an exploratory technique that examines the relationship between intensity of treatment and a given outcome, as well as any nonlinearities that may exist in that relationship.

### Analyzing Differences in Students' Socioemotional Attitudes and Their Association with Future Achievement

To address the question of whether students who were at risk but promoted, at risk and retained, and not at risk differed in their attitudes toward school and in self-confidence regarding ELA and mathematics, we conducted weighted pairwise comparisons for each subgroup and for each survey year and tested whether the differences were statistically significant. We adjusted the standard errors to account for grade-range stratification and clustering at the school level, and we adjusted the p-value to account for the multiple comparisons we conducted.

We also explored models that included both student- and school-level covariates. We conducted OLS linear regression or logistic regression, where appropriate, and again adjusted the standard errors to account for grade-range stratification and clustering at the school level. We also compared the responses of retained students to those of different groups of at-risk promoted students (for example, those who were promoted by passing the summer assessments, through the spring or August portfolio review, or by an appeal to district officials) to see whether there were systematic differences in attitudes.

## Study Limitations

Our study has several limitations. First, in response to the requirements of NCLB, in spring 2008, New York State began administering state assessments in grades 3 through 8 in ELA and mathematics, replacing the spring assessments previously administered by the city in these two subjects in grades 3, 5, 6, and 7 and by the state in grades 4 and 8. The new state assessments were not equated to years prior to 2006, and a separate standard-setting procedure was used to establish the proficiency-level scores. Thus, both the scale and level scores of students from the consecutive spring assessments in 2005 and 2006 are not comparable. This restricted our ability to compare longitudinal outcomes over time for some cohorts in certain years, particularly in using the comparison cohort as a control group for cohorts that took the state assessment.

Second, a major limitation of using test scores as outcome measures was that test scores are imperfect measures of student knowledge and skills. As Koretz (2002, pp. 754–755) points out,

> Users of test scores often assume that scores are direct and unambiguous measures of student achievement, much as the price of a commodity is unambiguous. An increase in test score, for example, is typically taken as clear evidence that students are learning more. Scores on most achievement tests, however, are only limited measures of the latent construct of interest, which is some aspect of student proficiency. As measures of these constructs, test scores are generally incomplete, and they are fallible in two senses: in the traditional statistical sense that they include measurement error, and in the sense that they are vulnerable to corruption or inflation.

CTB/McGraw-Hill, the vendor for the New York State assessments, provides comprehensive reports on test design and development (see NYSED, undated), permitting some confidence regarding the validity of the test itself.[10] However, a number

---

[10] We relied on the assumption that all scale and level scores provided were reasonable and consistent measures of achievement over time. Technical documentation about the assessments indicates that level scores are equated within grade and across years. However, the number of questions a student had to answer to score Level 2 on the state assessment declined over time, which increased the probability that a student could "guess" his or her way into Level 2. For the grade levels in our study years (2006–2008), this probability remained quite low—less than 1 percent, with one exception (7th-grade ELA in 2008, which was 5 percent). In 2009, the probability of scoring Level 2 by "guessing" increased, but this occurred after the period of our study.

of other concerns remain, particularly vulnerability to inflation due to teaching methods focused solely on test preparation. While we recognize that test scores are subject to measurement error, they are the best available uniform measure for tracking student outcomes and are used for accountability purposes. Obviously, our study shares this limitation with all studies that use test scores as outcome measures.

Third, we needed to account for the relative influences of the city, state, and national contexts on student outcomes. The study was intended to evaluate NYC's 5th-grade promotion policy and to estimate the effects of retention as well as the supportive interventions offered to at-risk students on current and future student outcomes. The promotion policy was only one piece of a larger set of reforms adopted by the city under the Children First initiative, which, in turn, was enacted against the larger backdrop of a whole series of reforms undertaken by New York State and the pressures engendered by the federal NCLB mandates. These factors make the question of attribution enormously challenging. As we show in subsequent chapters, these reforms appeared to be associated with a large increase in assessment scores in both NYC and the rest of the state between 2006 and 2008. This is particularly true of grades 5 and 7—grades that are the focus of our proximal and future outcome analyses. This is especially problematic when analyzing the outcomes for retained students who took the higher-grade assessments a year behind their cohorts. If performance within grade is increasing over time, the performance of retained students will receive a boost from this general upward trend, regardless of any effect that inheres in the act of retention alone, by being a member of a subsequent cohort. We attempt to adjust for the upward trajectory in overall test scores to estimate the "net" effect of the promotion policy on student outcomes. If we have not been entirely successful in this adjustment, then the effects we estimate may still reflect some of these confounding influences. Because of this, we emphasize the bounds on our estimated effects.

Fourth, as mentioned earlier, because of the timing of the start of our study, we faced several constraints in fielding surveys of SPA and SSA administrators. In the first year, because of the late start of the study, we had to field three administrator surveys simultaneously, which resulted in low response rates on the SPA and SSA surveys. In the second year, due to the restructuring in the summer of 2007, we were unable to field an SSA survey because contact information for SSA administrators was no longer centralized and readily accessible. In the third year, because principals had responsibility for SPAs and SSAs by the fall of 2008, the principal survey fielded that year had to be amended to ask for information on both in-school supports and programs offered on Saturdays and during the summer. To mitigate respondent burden, the amount of data we were able to collect was necessarily more limited.

Fifth, the measures of implementation of different types of strategies and supports were limited in several ways. They were largely from self-reports on surveys, reinforced by observations and teacher focus groups, in a limited number of case-study sites. In addition, because of budgetary constraints, we were unable to field teacher surveys that could have shed more light on actual classroom practices.

Finally, while not a limitation of the study itself, it is important to remember that this study focused on short-term outcomes of the policy and did not address its long-term costs or benefits. The policy has not been in place long enough to determine whether short-term effects persist over time or whether, as prior research has found, retained students have an increased probability of dropping out. For the 5th-grade students retained under the policy, we are limited to two to three years from the retention decision for all cohorts. Ultimately, it is important to understand the relative benefits and costs associated with implementing a promotion policy and holding students back a year from both a societal and individual point of view. In this monograph, we examine the effects of retention primarily from a student's point of view, tracking the academic and socioemotional outcomes of students held to the policy. The larger societal costs and benefits, such as whether the benefits of AIS and an additional year of retention outweigh their costs, are not addressed. For example, Eide and Goldhaber (2005), focusing strictly on the economic consequences of retention, concluded that the overall longer-term benefits from retention fall far short of covering the costs to society in terms of financing an extra year of education and loss of future earnings. While the policy has not been in place long enough to address these larger systemic costs and benefits, without positive short-term benefits accruing to students, it is unlikely that positive long-term benefits will result at either the societal or individual level. Thus, it is important to examine short-term results—as we did in our study—to see whether there is any possibility of longer-term benefits.

Despite these limitations, the results make important contributions to the body of literature on the effects of grade retention. We used quasi-experimental methods to evaluate the effects of supportive interventions and retention on current and future outcomes, which permitted greater confidence in the results. We provide evidence—albeit limited—that certain types of instructional strategies appear to hold promise for helping struggling students and that the conditions of learning and teaching in some schools may require special attention and resources to level the playing field for students in these schools.

## Summary

Our study focused on (1) the implementation of the 5th-grade promotion policy, (2) current and future academic and socioemotional outcomes of four cohorts of 5th graders, (3) the effects of various supports provided both in school and through Saturday and summer programs and retention on student outcomes, and (4) lessons learned regarding the design and implementation of promotion policies. This monograph focuses on the first three topics. We used data from a variety of sources—case studies, surveys of administrators and students, longitudinally linked student data, and interviews with leaders of support agencies—and a mix of qualitative and quantitative methodologies to analyze these issues. Marsh et al. (2009), the second report in the series of reports documenting RAND's evaluation of the NYC 5th-grade promotion policy, addresses the fourth topic. To draw out lessons learned regarding design and implementation, we analyzed data from several interviews with state and district officials in a sample of states with promotion policies similar to NYC's.

CHAPTER FIVE

# School-Provided Support for Students: Academic Intervention Services

*Jennifer Sloan McCombs, Scott Naftel, Gina Schuyler Ikemoto, Catherine DiMartino, and Daniel Gershwin*

The NYCDOE 5th-grade promotion policy calls for schools to identify students who are at risk of being retained under the promotion policy and provide them with additional AIS to prepare them to meet the criteria for being promoted to the 6th grade. Schools are also expected to notify parents regarding their child's status and prepare appeals portfolios for students who score at Level 1 in ELA or mathematics on the spring assessment.

As mentioned earlier, one of the key components of the promotion policy was early identification of students who were at risk of retention and the provision of AIS to these students to help them meet promotion standards. To accomplish this, NYCDOE expected schools to form AIS teams, headed by AIS leaders, to identify students needing services at the beginning of the school year, recommend appropriate intervention services, and monitor the progress of individual students over the course of the year. NYCDOE leaders envisioned a process in which the team assessed each student's level of need and classified him or her according to three tiers. Tier 1 students were those who had been identified as needing services but who had needs that could be sufficiently addressed using differentiated instruction in the regular classroom. Tier 2 students struggled in particular areas, such as decoding or memory strategies, and would benefit from additional intervention outside regular classroom instruction. Tier 3 students had needs that could not be sufficiently addressed by differentiated instruction or intervention services and therefore needed to be evaluated for special education. While NYCDOE leaders did not necessarily expect that all schools would explicitly use this tiered system, they did expect schools to consider student needs and to prioritize the provision of services according to their level of need.

As described in Chapter Three, prior to June 2007, regions monitored and supported AIS leaders and service provision. In 2007–2008, principals assumed greater autonomy and authority over whether and how to provide in-school and out-of-school services to students who needed them. While the SSOs supported schools in areas requested by the principal, SSOs did not have a monitoring role and could not require schools to engage in certain activities.

To set the context for AIS provision, this chapter first describes the characteristics of 5th-grade schools and provides data on all students needing services and how those students were distributed across schools, using NYCDOE administrative data. It then examines schools' implementation of the promotion policy, including the identification and monitoring of students needing services, support services provided to students, support for and hindrances to implementation, and educators' perceptions about the promotion policy. The analysis is based on three years of quantitative and qualitative data, including principal surveys (spring 2006 and spring 2008), an AIS leader survey (spring 2007), and school site visits (seven in spring 2006, 10 in spring 2007, and 23 in spring 2008). Two of these years matched our policy cohorts, which were followed in our achievement analyses: Our spring 2006 survey matches our P2 cohort (2005–2006), and our spring 2007 survey matches our P3 cohort (2006–2007).

For ease of exposition in this chapter, when presenting survey results, we refer to each school year by its spring year (when the survey was administered). For example, the 2006–2007 school year is referenced as 2007. We describe trends using specific indicators when possible. The final section explores the relationship between school context and the use of various support strategies and students' performance on the spring assessments. Appendix B provides supplementary tables for this chapter.

## Profile of NYC Schools with a 5th Grade

In school year 2004–2005, the year in which the promotion policy was first implemented at the 5th-grade level, there were 707 schools with a 5th grade in NYC. These schools were organized into 10 regions and 32 districts. As shown in Table 5.1, the majority of the schools with a 5th grade—80 percent—were traditional elementary schools, covering grades prekindergarten (PK) through 5 or PK–6. Some of these schools covered slightly different grade spans (for example, grades 1–5, 3–6, or 4–5). Another 15 percent spanned grades PK–9. A small number (5 percent) encompassed grades 5–9.[1]

Overall, enrollment in these schools varied from a low of 118 students to a high of 1,980. NYC schools are disproportionately high-poverty and high-minority. On average, about two-thirds (65 percent) of the students in schools with a 5th grade were eligible for free or reduced-price lunch (traditionally used as a proxy for the poverty rate of the school). The percentage of economically disadvantaged students in schools varied from 5 percent to a high of 100 percent, and the median was 78 percent (i.e., in half of the schools, the percentage of economically disadvantaged students was equal to or greater than 78 percent). On average, about 16 percent of students were non-Hispanic white, 33 percent were black, 38 percent were Hispanic, and 12 percent were Asian. However, demographics differed markedly by region and district, and often within districts as well. For example, in the 42 schools in District 31, the percentage

[1] Some schools were in the process of transition—they were becoming middle schools with no 5th grade. However, during the 2004–2005 school year, they still included a few 5th graders and thus are included in the table.

**Table 5.1**
**Grade Span of Schools with a 5th Grade, 2004–2005**

| Low Grade | High Grade | Number of Schools | Percentage of Total |
|---|---|---|---|
| PK–5 | 3–6 | 561 | 79.7 |
| 5–8 | 7–9 | 36 | 5.1 |
| PK–4 | 7–9 | 102 | 14.5 |
| PK–4 | 10–12 | 5 | 0.7 |
| 5–8 | 10–12 | 0 | 0.0 |
| Total | — | 704 | 100.0 |

of students eligible for free or reduced-price lunch ranged from 16 to 99 percent, with an average of 47 percent.

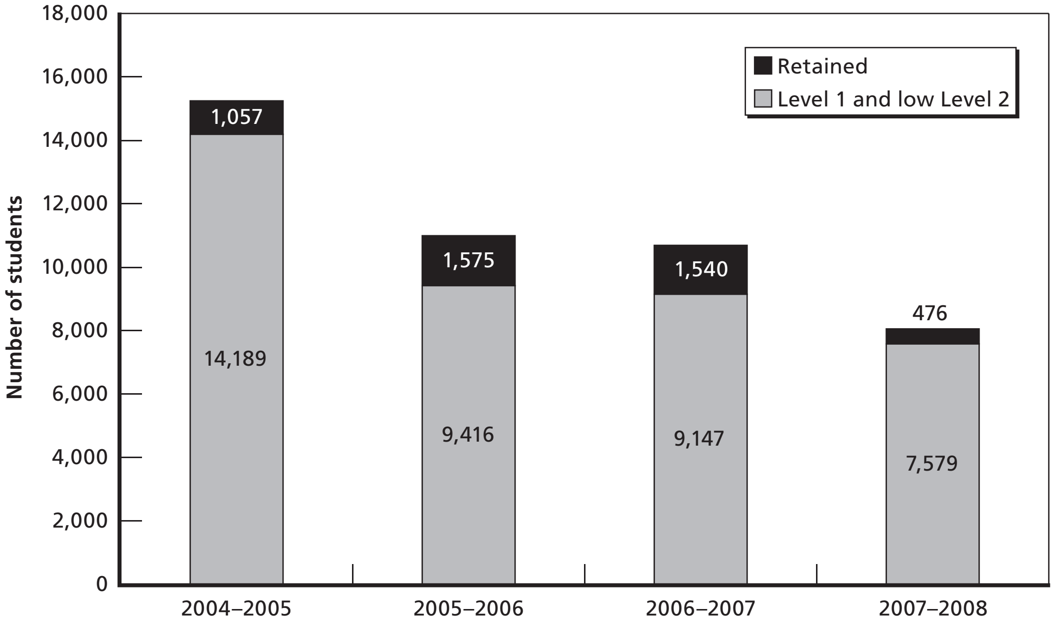
About 12 percent of students were classified as ELLs, although this varied from 0 to 60 percent across the schools. On average, 11 percent of the school populations were special education students with IEPs, with a range of 1 percent to 43 percent. Student attendance and stability rates were high—well over 90 percent across all schools.

## Students Needing Services Across Schools

According to NYCDOE policy, schools should, at a minimum, provide AIS to 5th-grade students who score at Level 1 or low Level 2 on the prior year's assessment, as well as to students retained in grade. In the first year of implementation of the promotion policy (2004–2005), the number of priority students needing services in any subject (retained, Level 1, and low Level 2) was more than 15,000 (Figure 5.1). In 2005–2006, the first year of our study, the number of students needing services declined to approximately 11,000. The number of students needing services held steady in 2006–2007 and then fell again to approximately 8,000 students in 2007–2008. In addition, the number of retained students in 5th grade also declined over time, from more than 1,500 in 2005–2006 and 2006–2007 to 476 in 2007–2008.

However, students needing services were not evenly distributed across schools with a 5th grade. Some schools enrolled a much higher percentage of these students than other schools. For instance, in 2005–2006, the percentage of students needing services in a given school ranged from 0 to 80 percent of the 5th-grade population (Figure 5.2). The median percentage needing services was 12 percent. We found a similar range (0 to 91 percent) and median (13 percent) in 2006–2007. In 2007–2008, the range across schools was smaller (0 to 58 percent), and the median was lower (9 percent).

**Figure 5.1**
**Number of Students Needing Academic Intervention Services, by Year**

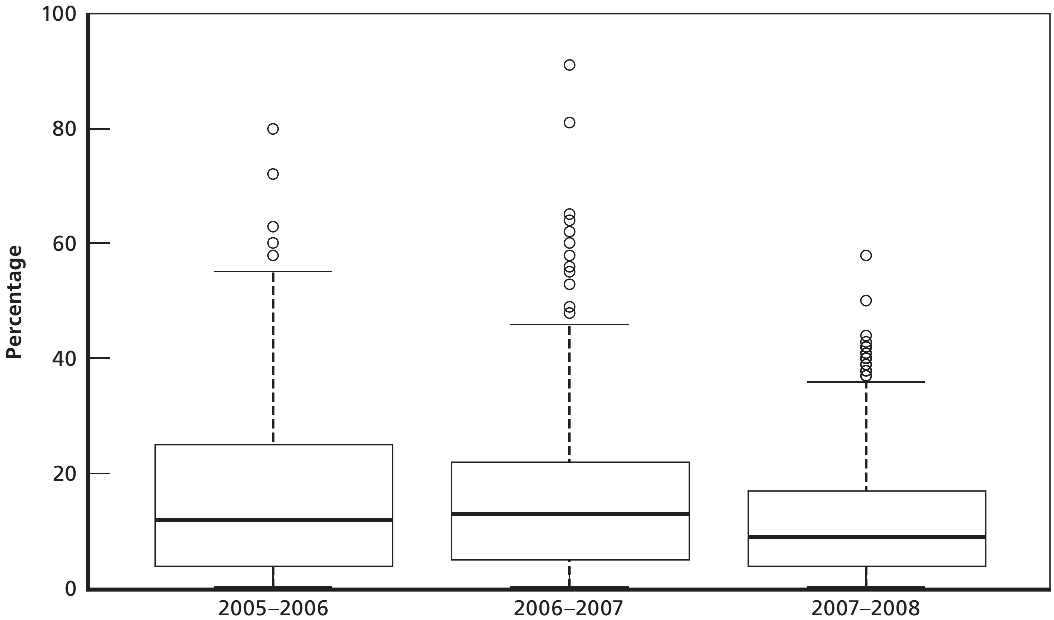



RAND *MG894-5.1*

We hypothesized that schools' abilities to provide services to students and students' experiences with additional services might vary based on the percentage of students needing AIS as a proportion of the school population. That is, implementation of AIS might be different for a school in which 70 percent of 5th-grade students require services than would be the case for a school in which only 5 percent need services. We used the distributions shown in Figure 5.2 to determine the approximate quartiles of the three distributions. Although the quartiles were somewhat different, they were sufficiently close to allow us to use common quartile thresholds for the three surveys. Schools were categorized into three categories based on the percentage of 5th-grade students in the school requiring AIS: low-need schools (lowest quartile, with ≤ 7 percent of 5th-grade students needing services), medium-need schools (the middle half of the distribution, defined as schools with 7–25 percent of 5th-grade students needing services), and high-need schools (highest quartile, with > 25 percent of 5th-grade students needing services).

As expected, high-need schools faced a more challenging school context and more demands on their resources. As a group, they typically had a somewhat greater proportion of students needing specialized instruction who might not have been held to the promotion policy but would need additional services—special education and ELL

**Figure 5.2**
**School-Level Percentages of 5th Graders Needing Services at the Beginning of the School Year**



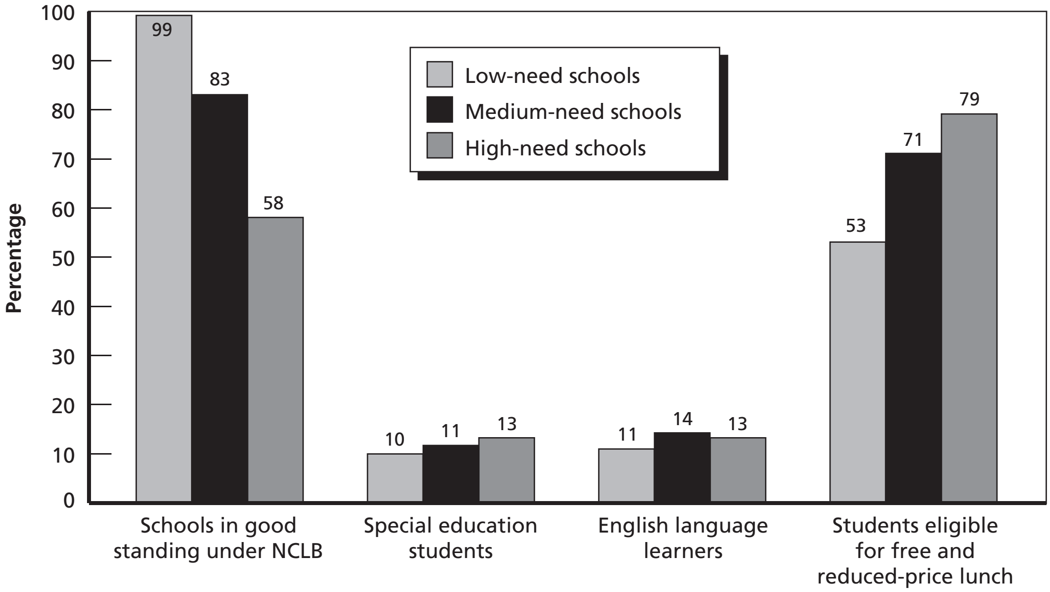

RAND *MG894-5.2*

students (Figure 5.3). Further, compared to other schools, high-need schools enrolled greater percentages of students from low-income families and were far less likely to be in good standing under NCLB.[2]

We now discuss the findings from our surveys and case studies on the implementation of various components of the promotion policy. Because of space limitations and the somewhat different focus of the three administrator surveys over the three years, the findings are often drawn from the years in which the particular questions were asked, and the tables display the survey years from which the data were drawn. If a particular question or item was not asked in a given year, we use "NA," or not applicable, to indicate that this was the case. It is important to remember that, while case-study data provide rich, descriptive representations, they are not generalizable and, thus, are not representative of the population of schools with a 5th grade.

2 Schools in good standing under NCLB have met both test-participation requirements and a performance requirement (the annual measurable objective) for every accountability group in every tested grade level. Schools that failed to meet these targets in the past and were designated as needing improvement must meet these targets for two consecutive years to be in good standing.

**Figure 5.3**
**Characteristics of School Populations, by Need Status of School, 2005–2006**




RAND *MG894-5.3*

## Selection and Monitoring of Students

### Process for Selection and Monitoring of Students for AIS

Almost all schools, regardless of need status, reported having an AIS team to select students for and monitor the provision of AIS (see Table B.1 in Appendix B). In our case studies, we found that these teams tended to meet on a weekly or monthly basis. When selecting students for AIS, the vast majority of respondents (more than 95 percent) reported relying on state assessments and teacher recommendations to a moderate or great extent (see Table B.2). Parent requests were also considered, albeit to a lesser degree.

In our case studies, we found that the majority of schools identified students needing services at the beginning of the year based on the previous year's performance. Some schools reported making their initial AIS rosters in the spring of the previous year following the release of the 4th-grade test scores. Then, in the fall, these schools administered follow-up assessments (for example, Running Records) to assess students' growth and learning needs. A few schools reported that the identification of students eligible for intervention services occurred all year long, with students continuously being added to and subtracted from the AIS roster. These schools reported using the results of teacher recommendations or periodic or interim assessments (tests given occasionally during the year to track students' progress toward meeting grade-level standards) to update AIS rosters. The majority of school officials with whom we spoke stressed that AIS placement was a fluid process, though many principals and AIS leaders noted that students were much more likely to be added over the course of the year than to be removed from the roster.

Once students were identified as needing services, almost all (97 percent or more) principals in 2006 and 2008 reported sending letters to parents or guardians and holding in-person meetings with them to explain the promotion policy and discuss their children's service needs. The vast majority of principals also reported discussing ways in which families could support the child's learning.

In terms of monitoring student progress throughout the year, schools leaders reported relying on a variety of data, such as interim assessments, teacher-created tests, student work, and progress reports from AIS teachers (see Table B.3). Almost all the school leaders (96–99 percent) reported that student work was moderately or very helpful in monitoring student progress. Letter grades or grade point averages (GPAs) were reported as less helpful than other sources of data.

In 2006 and 2007, schools were required to administer the interim assessment provided by the city, and the majority of respondents reported that these assessments were useful to a moderate or great extent in tracking students' progress (see Table B.3). In 2008, schools were able to select the interim/periodic assessment (e.g., Scantron Performance Series, Acuity Predictive Assessments, own designs) that they wished to use. Of those using a particular assessment, almost all principals reported using those results to a moderate or great extent in monitoring AIS.

In our case studies, AIS providers and teachers reported that the results received throughout the school year informed them about students' learning strengths and weaknesses and influenced their teaching. Many principals reported that, compared to prior years, AIS provision had improved in part because staff had been using data to a greater extent to inform teaching and learning.

### Students Targeted for Services

Consistent with NYCDOE's expectations, schools appeared to focus AIS on the lowest-performing students. More than 85 percent of AIS leaders in 2007 and principals in 2008 reported that all their Level 1 5th-grade students received AIS in ELA (Table 5.2).[3] These leaders were slightly less likely to report that all Level 1 students received services in mathematics (82–83 percent). The majority of schools also served all low Level 2 students in ELA and mathematics.

Schools seemed to provide extra assistance to as many students as they had the capacity to serve. Compared to 2007, in 2008, a greater percentage of schools extended

[3] In 2006, we asked about combined services for ELA and mathematics. Approximately 78 percent of principals reported serving all Level 1 students, 64 percent reported serving all low Level 2 students, and 30 percent reported serving all mid- or high Level 2 students.

**Table 5.2**
**Percentage of Respondents Reporting That None or All 5th-Grade Students Received AIS in ELA and Mathematics, by Level**

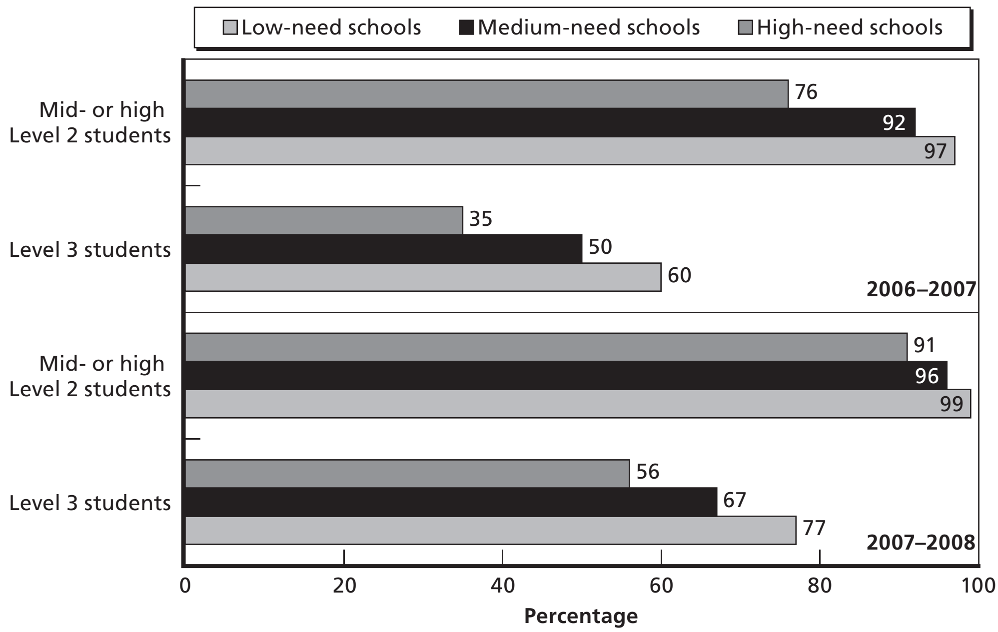
| | 2007 | | 2008 | |
|---|---|---|---|---|
| **Subject and Performance Level** | **None** | **All** | **None** | **All** |
| ELA | | | | |
| Level 1 | 3.0 | 85.9 | 2.6 | 87.9 |
| Low Level 2 | 3.0 | 56.3 | 0.0 | 69.6 |
| Mid- or high Level 2 | 12.2 | 19.3 | 4.7 | 38.6 |
| Level 3 | 53.7 | 2.1 | 33.9 | 5.5 |
| Mathematics | | | | |
| Level 1 | 4.4 | 81.7 | 4.6 | 83.2 |
| Low Level 2 | 5.1 | 52.5 | 2.7 | 64.5 |
| Mid- or high Level 2 | 14.8 | 17.6 | 9.4 | 34.1 |
| Level 3 | 55.1 | 2.0 | 39.1 | 6.0 |

services to mid- or high Level 2 students and Level 3 students. It is likely that schools were able to do so as a result of the sharp decline reported earlier in the number of students who were retained or classified as Level 1 or low Level 2 at the beginning of the school year.

Further, in 2007, schools with low percentages of students in need of services were significantly more likely than high-need schools to serve students who had scored above Level 1. While 73 percent of low-need schools served *all* low Level 2 students in ELA, and 64 percent did so in mathematics, only 40–42 percent of high-need schools served *all* low Level 2 students in ELA and mathematics (not shown). Low-need schools were also more likely than high-need schools to extend services and serve a few high Level 2 students and Level 3 students (Figure 5.4). For instance, 97 percent of low-need schools extended services to a few mid- or high Level 2 students in ELA, compared to 76 percent of high-need schools. In our case-study visits, school personnel sometimes referred to these students as "pushables" or "slippables": students (high Level 2) who could be "pushed" up to Level 3 proficiency, or Level 3 students who could potentially "slip" down to Level 2 without additional support during the school year.

In 2008, with fewer 5th-grade students in need of services throughout the city, a significantly greater percentage of high-need schools extended services to mid- or high Level 2 students, and all schools, regardless of need status, were more likely to serve at least a few more Level 3 students than they did the prior year.

**Figure 5.4**
**Percentage of Schools Providing In-School AIS in ELA to at Least a Few Students at Various Achievement Levels, by Need Status of School**




RAND *MG894-5.4*

In our case studies, we also found that schools served the maximum number of students possible. For instance, one school we visited in 2007 only had nine or 10 students who were determined to be in need of services at the beginning of the year. Because they had capacity to serve more students, the school provided AIS to 16–18 students, including some Level 3 students who were perceived by teachers as in danger of slipping to Level 2.

## Academic Intervention Service Provision

### Providers

In all three years, schools relied on a variety of personnel to provide AIS. On average, survey respondents reported that five different types of school staff provided at least some AIS to students. While many people provided AIS, including coaches, administrators, aides, and parent volunteers, the majority of schools reported relying to the greatest extent on reading and mathematics specialists and AIS leaders to provide AIS. However, at some schools, certain staffing resources did not exist. A smaller percentage of schools had reading coaches, mathematics coaches, and AIS leaders in 2008 than in the prior year. Schools were more likely to have reading specialists than mathematics specialists in both years. Overall, high-need schools were significantly more likely than low-need schools to have a variety of personnel resources, and they were significantly more likely to have mathematics specialists, reading coaches, and mathematics coaches (see Figure B.1 in Appendix B). In the open-ended question on the survey, a few principals in low-need schools noted that they had a limited budget for AIS. As one principal wrote, “Being a non–Title I school, my budget is extremely limited. I never know if I will have enough money in my budget to afford AIS teachers.”

### Types of Academic Intervention Service Programs

Schools provided AIS in a number of settings. In all three years, almost all schools reported serving students in need of services through small-group tutoring during the school day (Table 5.3). Similarly, almost all schools served students needing services through the 37.5-minute program, which was adopted in 2007. Because the 37.5-minute program was created to provide additional help and tutoring for struggling students, this is not a surprising finding. Fewer schools were able to offer one-on-one tutoring during the school day—between 64 and 80 percent, depending on the year. Only a small percentage of schools chose to operate an intersession program to provide additional instruction to students during school breaks, and this percentage declined in 2008 (from 44 to 30 percent).

Of schools providing these services to in-need students, the percentage of school officials reporting being able to serve *all* in-need students, rather than a few or some, increased between 2007 and 2008 (Table 5.4). For example, in 2008, 24 percent of schools offering one-on-one tutoring reported that *all* their in-need students received this type of help, compared to 15 percent in 2006 and 8 percent in 2007. Similarly, 72 percent of schools providing small-group tutoring during the school day reported serving *all* in-need students, compared to 53 percent in 2006 and 43 percent in 2007.

However, in our case-study visits, we found that the form and frequency of these services varied. Some schools used “push-in” strategies, in which an intervention teacher provide additional in-class support to a group of students, other schools preferred to have AIS providers take students out of the classroom to provide individual

**Table 5.3**
**Services Available to In-Need 5th-Grade Students**

| Service | 2006 | 2007 | 2008 |
|---|---|---|---|
| One-on-one tutoring with a specialist during the school day | 80.1 | 63.8 | 73.2 |
| Small-group tutoring with a specialist during the school day | 99.4 | 97.9 | 98.8 |
| 37.5-minute program | NA | 99.7 | 99.7 |
| Intersession | NA | 44.0 | 30.1 |

**Table 5.4**
**Percentage of Schools Offering Services to All In-Need 5th-Grade Students**

| Service | 2006 | 2007 | 2008 |
|---|---|---|---|
| One-on-one tutoring with a specialist during the school day | 14.9 | 7.7 | 23.6 |
| Small-group tutoring with a specialist during the school day | 53.3 | 43.2 | 71.6 |
| 37.5-minute program | NA | 71.8 | 83.4 |
| Intersession | NA | 8.2 | 27.5 |

or small-group instruction in another setting, and some schools used both methods of AIS provision.

The frequency of the services reported by providers and teachers varied across schools. In our 2008 visits, we found that services were typically offered according to the following schedules:

- four to five times per week (10 schools)
- two to three times per week (nine schools)
- once a week (one school)
- occasionally or not at all (three schools).

Across all three years, respondents described consistency of service provision as a problem. Across our case-study schools, services were offered less often than intended because providers were pulled away to cover classes, grade assessments, or proctor tests for students with modifications. Many teachers reported that providers missed at least one session with students per week and that, during periods of peak testing, no in-school AIS was provided. As one AIS leader noted, "The major challenge is staffing and schedule. . . . AIS people get pulled out quite often for different purposes such as testing, make-up testing, covering absent teachers' classrooms, and paperwork."

In terms of services offered to students outside the school day, while the percentage of schools offering intersession decreased over time, those schools that offered intersession had better attendance, with 28 percent in 2008 reporting that all their in-need students attended versus 8 percent in 2007.

Many schools also offered students the opportunity to participate in before- or after-school programming, which may or may not have an academic focus. Among our 2008 case-study schools, the majority provided after-school programming other than the 37.5-minute period. At a few of these schools, the after-school program was a federally funded supplemental education service program that provided tutoring to struggling students. Other schools ran their own programs and provided preparation for upcoming state ELA or mathematics tests. Still other programs offered enrichment activities. For instance, one school offered dance and knitting classes as well as a mathematics team during its after-school program. Most of these programs were open to all students, and very few specifically targeted at-risk students.

### Focus of Intervention Services

In 2007, we gathered additional data on the focus of the intervention services provided during and outside the school day. We found some differences in the focus of services provided to students, depending on whether they were offered during or outside the school day (Table 5.5). AIS leaders were more likely to report that services focused on working with a specific intervention program to a moderate or great extent during the school day (71 percent) than outside the school day (34 percent). On the other hand, teaching test-taking strategies was a focus to a moderate or great extent both during the school day (74 percent) and outside the school day (69 percent).

According to teachers in our case-study schools, test preparation was a major focus of the 37.5 minutes. Of the schools that focused on test preparation, half used professionally compiled test-preparation materials, mostly from Kaplan, while the other half used teacher-generated test-preparation materials. At many schools, teachers described the 37.5 minutes as being aligned with the testing calendar. In other words, from October to January, the 37.5 minutes focused on preparing students for the upcoming state assessment in ELA and, from January to March, on preparing students for the upcoming state mathematics assessment. After the tests, teachers at some schools reported concentrating more on students' individualized learning needs during the 37.5-minute period, while teachers at other schools tried to incorporate games and other activities to make the extra time more fun for the students.

Table 5.6 shows that, compared to AIS leaders in low-need schools, AIS leaders in high-need schools were significantly more likely to report that in-school services focused to a moderate or great extent on working with a specific intervention program (77 percent versus 56 percent). Also, they were less likely to report that these services

**Table 5.5**
**Focus of Academic Intervention Services During and Outside the School Day, 2007**

| Service | During the School Day | Outside the School Day |
|---|---|---|
| Working with a specific intervention program | 71.0 | 34.3 |
| Teaching students test-taking strategies | 74.4 | 68.9 |
| Tutoring students on class work | 53.7 | 39.9 |
| General remediation | 67.9 | 55.0 |
| Pre-teaching content to students | NA | 21.2 |

**Table 5.6**
**Focus of Academic Intervention Services During and Outside the School Day, by Need Status of School, 2007**

| | During the School Day | | Outside the School Day | |
|---|---|---|---|---|
| Service | Low Need | High Need | Low Need | High Need |
| Working with a specific intervention program | 55.6 | 75.7 | 13.2 | 33.1 |
| Teaching students test-taking strategies | 70.3 | 69.1 | 59.3 | 68.2 |
| Tutoring students on class work | 60.7 | 51.5 | 36.8 | 27.2 |
| General remediation | 75.0 | 61.8 | 50.4 | 48.0 |
| Pre-teaching content to students | NA | NA | 19.7 | 20.6 |

focused on general remediation to a moderate or great extent (62 percent versus 75 percent). Similarly, high-need schools were more likely than low-need schools to focus out-of-school services on specific intervention programs (33 percent versus 13 percent).

### Perceived Effectiveness

Overall, the vast majority of principals and AIS leaders using the various supports rated them as moderately or very effective in improving the academic performance of in-need students. More than 95 percent of respondents rated one-on-one and small-group tutoring as moderately or very effective (Table 5.7). AIS leaders in 2007 were less likely than principals in 2008 to rate intersession as being moderately or very effective (60 percent versus 81 percent). This may not be surprising, considering the likelihood that only principals who thought intersession was an effective intervention would choose to fund it.

More than 80 percent of AIS leaders and principals rated the 37.5-minute program as moderately or very effective, though teachers interviewed for the case studies were not as sanguine about the program. While we visited a few schools in which the program had a clear purpose and there was a smooth transition between the school day and the 37.5-minute period, teachers in the majority of the schools we visited reported that their school struggled to implement it. In a few schools, teachers reported that it was a "logistical nightmare" to get students into the classroom and engaged. For instance, one teacher noted that "a lot of time is wasted in transition [between dismissal and the start of the 37.5 minutes]. If the students get 15 minutes of instruction during the 37.5 minutes, we're lucky." In other schools, the program's effectiveness was undermined because of low student participation. One AIS leader noted,

> All of the at-risk children are invited to attend 37.5 minutes, but only a small number of them actually attended. Many children won't stay for 37.5 minutes

**Table 5.7**
**Percentage of Administrators in Schools Offering the Service Who Perceived It to Be Moderately or Very Effective**

| Service | 2006 | 2007 | 2008 |
|---|---|---|---|
| Before- or after-school program | 98.0 | 89.7 | 90.2 |
| One-on-one tutoring with a specialist during the school day | 96.6 | 94.8 | 95.9 |
| Small-group tutoring with a specialist during the school day | 97.8 | 99.6 | 98.6 |
| 37.5-minute program | NA | 80.8 | 84.6 |
| Intersession | NA | 59.3 | 80.5 |

> because their parents won't let them. Some families have multiple kids, and some of their kids have to stay for the 37.5 minutes and some don't. And parents won't have one kid stay while others don't. I think you either have to give all students 37.5 minutes or none to anyone.

## Additional School Strategies to Improve Student Performance

In addition to AIS, schools used a variety of other strategies to improve the academic performance of in-need 5th-grade students. Interestingly, the use of various student- and teacher-placement strategies increased significantly over time (Table 5.8). In 2006, 46 percent of principals reported assigning the best teachers to the 5th grade, while 62 percent reported doing so in 2008. Similarly, we found increases in the percentage of principals who reported assigning the best 5th-grade teachers to 5th-grade students with the weakest skills (50 percent versus 64 percent), placing in-need 5th-grade students into smaller classes (42 percent versus 49 percent), and grouping retained 5th-grade students together for part of the day (29 percent versus 37 percent). Teacher-assignment strategies were rated as moderately or very effective by more than 90 percent of schools that used those strategies. In schools that grouped retained students together for part of the day, the percentage of administrators rating this strategy as moderately or very effective increased over time—from 77 percent in 2006 to 90 percent in 2008.

One principal wrote about the decisions he or she was able to make given the increased autonomy granted to principals in 2008:

> By being permitted to hire F-Status [regular, part-time] personnel to tutor Level 1 students and to reduce class size by using experienced personnel, we were able to reduce our Level 1 population to numbers never before seen at my school.

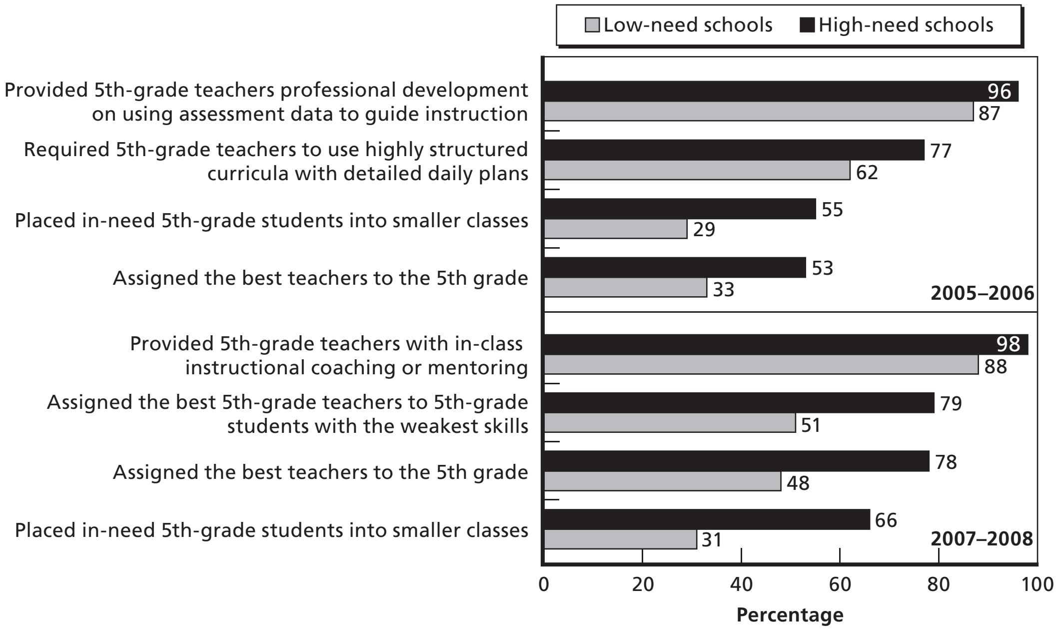
**Table 5.8**
**Principals' Reports of Strategies Used to Improve the Academic Performance of In-Need 5th-Grade Students**

| Strategy | 2006 | 2008 |
|---|---|---|
| Assigned the best teachers to the 5th grade | 46.3 | 61.7 |
| Assigned the best 5th-grade teachers to 5th-grade students with the weakest skills | 49.9 | 63.8 |
| Placed in-need 5th-grade students into smaller classes | 42.0 | 49.2 |
| Grouped 5th-grade retained students together for instruction for the entire day | 12.7 | 18.8 |
| Grouped 5th-grade retained students together for instruction for part of the day | 28.7 | 37.0 |
| Provided 5th-grade teachers in-class, instructional coaching or mentoring | 92.9 | 92.2 |
| Provided 5th-grade teachers professional development on using assessment data to guide instruction | 98.4 | 99.7 |
| Provided 5th-grade teachers professional development on differentiated instruction | 97.0 | 99.1 |
| Required 5th-grade teachers to use highly structured curricula with detailed daily plans | 75.6 | 74.9 |
| Required 5th-grade teachers to use materials that taught test-taking strategies | 97.2 | 94.3 |

Other types of strategies were used by a vast majority of schools across all three years, such as providing 5th-grade teachers with professional development on differentiated instruction or using assessment data to guide instruction and requiring 5th-grade teachers to use materials that teach test-taking strategies. The vast majority of administrators in schools using these strategies rated these strategies as moderately or very effective (see Table B.5 in Appendix B).

Principals in high-need schools employed certain strategies with greater frequency than those in low-need schools (Figure 5.5). For instance, principals of high-need schools were significantly more likely than principals of low-need schools to report assigning the best teachers to the 5th grade and placing in-need students in smaller classes. In 2008, these principals were also more likely to report assigning the best 5th-grade teachers to students with the weakest skills (79 percent versus 51 percent).

**Figure 5.5**
**Principals' Reports About Adopting Certain Strategies in 5th Grade, by Need Status of School**




RAND *MG894-5.5*

## Support for Promotion Policy Implementation

Across the three years, the majority of principals and AIS leaders believed that their region or SSO provided a number of useful supports (Table 5.9). About 90 percent of principals in 2006 and AIS leaders in 2007 reported that the region or district clearly communicated the rules and regulations regarding the 5th-grade promotion policy. More than 70 percent of principals reported that their region or SSO provided them with high-quality professional development focused on supporting in-need students and their teachers and that it had provided their school with the necessary guidance on selecting curriculum materials to support in-need students.

Interestingly, AIS leaders in 2007 were less likely than principals in 2006 to agree that the region provided high-quality professional development focused on teaching in-need students. AIS leaders were also more likely than principals in 2006 to report that the region or district required too much paperwork that took time away from serving students. During school visits, some AIS leaders noted that excessive paperwork, particularly PIP forms, hindered their ability to help in-need students.

**Table 5.9**
**Percentage of Respondents Agreeing or Strongly Agreeing About Support Received in Implementing the 5th-Grade Promotion Policy, from Region or District in 2006 and 2007 and from School Support Organization in 2008**

| Support | 2006 | 2007 | 2008 |
|---|---|---|---|
| Provided the school with necessary guidance on selecting curriculum materials to support in-need students | 80.3 | 82.2 | 74.8 |
| Provided principals with high-quality professional development focused on supporting in-need students and their teachers | 71.3 | NA | 81.2 |
| Provided teachers with high-quality professional development focused on teaching in-need students | 65.2 | 51.5 | 69.7 |
| Clearly communicated the rules and regulations regarding the 5th-grade promotion policy | 91.5 | 89.7 | NA |
| Required too much documentation and paperwork, taking time away from serving students | 60.2 | 74.6 | NA |
| Helped the school get the resources it needed to support in-need students | 64.0 | 61.5 | 74.0 |
| Provided useful assistance in analyzing student data | 63.9 | 59.9 | 89.6 |
| Provided information in a timely fashion | 64.3 | 62.7 | NA |
| Provided useful support for AIS leaders | NA | NA | 73.9 |
| Provided me, the AIS leader, with high-quality professional development focused on supporting in-need students and their teachers | NA | 90.3 | NA |

NOTE: The 2007 percentages do not include AIS leaders in empowerment schools.

As a group, principals reported varying levels of support in certain areas from the region and SSO. Principals were more likely to agree that the SSO (versus the region) provided the following supports:

- high-quality professional development for principals focused on supporting in-need students and their teachers (81 percent versus 71 percent)
- help with obtaining the resources needed to support in-need students (74 percent versus 64 percent)
- useful assistance in analyzing student data (90 percent versus 64 percent).

This increase in support for data analysis also coincided with the creation of a new district position called the “senior achievement facilitator.” As a key member of NYCDOE’s new Inquiry Team initiative, the senior achievement facilitator assisted schools in their understanding and use of achievement data. Every school was assigned a senior achievement facilitator.

On the other hand, a smaller percentage of principals reported receiving guidance on selecting curriculum materials to support in-need students from the SSO (75 percent) than from the region (82 percent and 84 percent of principals and AIS leaders, respectively).

On the whole, principals seemed more satisfied with the school-requested type of assistance provided by the SSOs than with the region-directed support; however, in our case-study visits in 2008, many AIS leaders noted the absence of monthly meetings, limited exposure to curricular resources for in-need students, and a lack of networking opportunities with other AIS leaders that were previously organized by the region. One AIS leader commented, "It'd be nice to have a little bit more [professional development]—hear what's new and see what's out there." Given that many AIS leaders mentioned the importance of the strong support and guidance they received from the region, particularly from the RDISs, it is not surprising that they would experience the transition to the SSO structure differently than principals.

In particular, new AIS leaders reported struggling due to a lack of support. Three schools that we visited in 2008 had new AIS leaders who reported that AIS had deteriorated due to the loss of experienced school-based staff. Specifically, they noted that seasoned AIS leaders either retired or left the school, taking their knowledge about AIS with them.

## Hindrances to Improving the Performance of In-Need 5th-Grade Students

Overall, the majority of school administrators did not report significant hindrances to improving the performance of in-need 5th-grade students (Table 5.10). One area cited as a moderate or great hindrance by the majority of administrators was lack of teacher time or ability to properly differentiate instruction in the classroom (55–56 percent). More than 60 percent of AIS leaders reported that other demands that kept AIS team members or intervention teachers from working with students during the school day as scheduled were a moderate or great hindrance in 2006; however, a smaller percentage of principals (44 percent) reported that this was a problem in 2008. While it is possible that this situation improved, perhaps the finding is the result of the decline in the number of in-need students. It is also possible that principals were less aware of the problem.

While many areas were not considered substantial hindrances by the majority of schools, respondents in high-need schools were much more likely than those in low-need schools to report that a given factor was a moderate or great hindrance

**Table 5.10**
**Percentage of School Leaders Reporting Factors as a Moderate or Great Hindrance to Improving the Performance of In-Need 5th-Grade Students**

| Factor | 2006 | 2007 | 2008 |
|---|---|---|---|
| Other demands that kept AIS team members or intervention teachers from working with students during the school day as scheduled | NA | 62.8 | 44.2 |
| Lack of teacher time or ability to properly differentiate instruction in the classroom | 55.0 | 56.0 | NA |
| Large numbers of retained students who prevented proper attention from being paid to in-need students | 16.6 | 12.5 | 11.1 |
| Large numbers of at-risk students in other grades affected by the promotion policy (i.e., 3 or 7) | 23.3 | 24.9 | 14.5 |
| Lack of cooperation from parents | 34.5 | 40.9 | 34.8 |
| Low student participation in out-of-school programs | NA | 32.4 | 33.6 |
| Large class size | 48.0 | 46.0 | 36.1 |
| Lack of time for teacher planning and professional development | 58.7 | 39.4 | 44.3 |
| Student attendance | 29.2 | 24.8 | 26.1 |

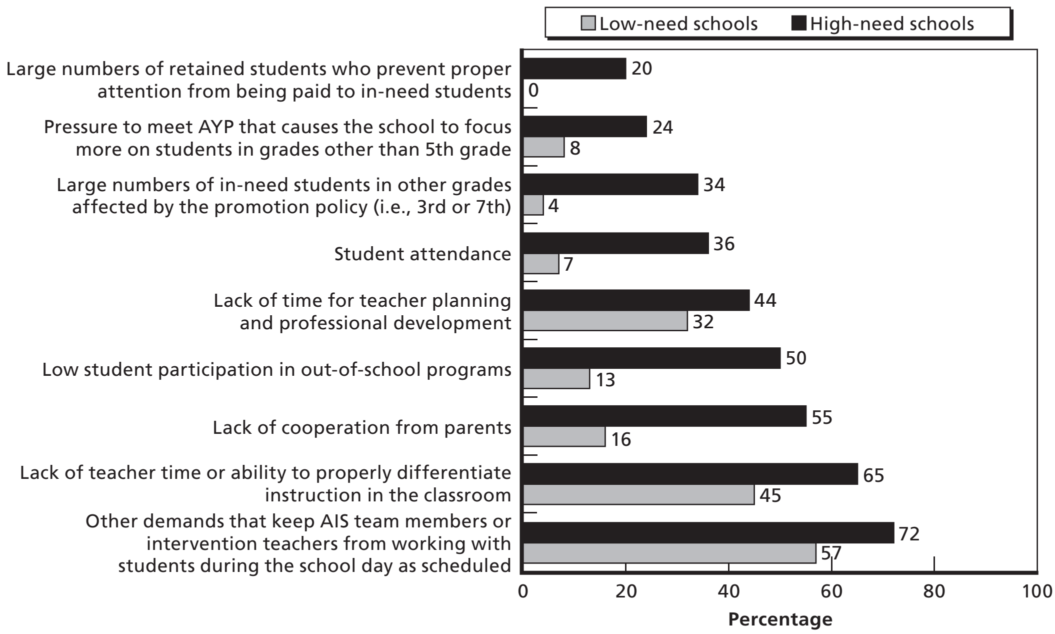
NOTE: The question about hindrances was worded differently in 2005–2006. It asked about hindrances to improving student performance generally rather than in implementing the promotion policy specifically.

(Figure 5.6). For example, in 2007, other demands on the AIS team or intervention teachers and a lack of teacher time or ability to properly differentiate instruction in the classroom ranked as the top two hindrances mentioned by AIS leaders in schools with high percentages of in-need students, with more than 60 percent of respondents mentioning these as moderate or great hindrances.

Three other factors were a particular problem for high-need schools: lack of cooperation from parents (55 percent), low student participation in out-of-school programs (50 percent), and student attendance (36 percent), but these were not cited as moderate or great hindrances for the vast majority of low-need schools, in which only 7–16 percent of respondents mentioned them as such.

Trends for 2008 are similar, with one notable exception. In 2008, principals in low-need schools were significantly more likely than high-need schools to report that large class size was a moderate or great hindrance (51 percent versus 22 percent). This may reflect principals' decisions to dedicate more resources to maintaining smaller class size in high-need schools.

**Figure 5.6**
**Extent to Which Factors Were a Hindrance to Improving the Performance of In-Need 5th-Grade Students, by Need Status of School, 2007**




RAND *MG894-5.6*

## Perceptions Regarding Student Retention and the Promotion Policy

Study participants tended to have mixed views of retention in general. Survey respondents were almost evenly split as to whether holding students back in grade improves their long-term chances of success. An overwhelming majority (88–91 percent) of respondents agreed or strongly agreed that retention would have different impacts on different students. A smaller majority (61–65 percent) agreed that most students end up with stronger skills after repeating a grade. At the same time, 67–70 percent agreed that retention negatively affects students' self-esteem. Given that the vast majority of respondents agreed that retention affects different students in different ways, the range of opinions regarding the specific impacts of retention is not surprising. There were no significant differences in the responses based on the percentage of students needing services in the school.

Respondents tended to be positive about many aspects of NYC's promotion policy (Table 5.11). Approximately three-quarters of AIS leaders and principals agreed that the policy focused their schools' instruction efforts in a positive way and that it had made parents and guardians more concerned about their children's progress. Approximately 49–63 percent also agreed that it had provided their school with additional resources to help serve low-achieving students. For instance, on the open-ended question soliciting additional comments, two AIS leaders wrote about the benefits of the policy:

> Since the 3rd grade and the 5th grade students have to meet certain criteria we have focused in on at-risk children in every grade. I am very much in favor of leaving back children who do not meet the criteria. The additional years that a child sometimes needs in his elementary years are crucial for that child in doing better in all the future years of his education.

> I think the policy is fair. Having an AIS team and requiring the team to write PIPs for at-risk students has made us more accountable. We analyze data more, create viable programs and monitor progress regularly. The students have been positively affected. For the past 3 years that we have been doing this, students have improved their skills by participating in AIS programs. The city has raised expectations for students and schools. I think this is a good thing.

However, the majority of respondents thought that the promotion policy relied too heavily on state assessment scores and, interestingly, that the policy made it more difficult to retain students who should be retained but passed the test. In our case-study visits, some respondents noted that, because other criteria were not considered as part of the promotion decision, the policy ended up discouraging student effort during the regular school year and reduced teachers' ability to motivate students to work hard. For example, respondents wrote,

> A student's promotion to the next grade, regardless of grade, should be based on multiple criteria. A single test should not be the determining factor, but rather the test score in conjunction with attendance and performance throughout the year. I have seen many students who do nothing all year and then [get] promoted because they scored a "2" on the state tests, and [I have seen] students who have numerous unexcused absences also be promoted.

> Give teachers more power in determining if students are promoted and use multiple criteria in addition to test scores for student promotion. . . . [S]ome students are promoted based on test scores when their classroom performance is not up to par and others are left back for failing [the] test due to testing anxiety.

Not surprisingly, respondents in high-need schools were significantly more likely to mention the burden created by overage students and the reduction in teacher authority. For example, 69 percent of AIS leaders in these schools agreed that the policy imposed a burden in dealing with overage students, compared to 28 percent of AIS leaders in schools with low percentages of in-need students. In our case-study visits, some respondents were concerned about the impact of the policy on students who were

**Table 5.11**
**Perceptions Regarding New York City's Promotion Policy**

| Perception | 2006 | 2007 |
|---|---|---|
| It has made parents/guardians more concerned about their child's progress. | 79.5 | 74.5 |
| It has focused the school's instructional efforts in positive ways. | 76.7 | 73.7 |
| It has provided the school with additional resources to help serve low-achieving students | 48.8 | 62.6 |
| It makes it more difficult to retain students who should be retained but passed the test. | 55.5 | 68.0 |
| It has diverted resources to grades affected by the promotion policy (3, 5, and 7) at the expense of other grades. | 46.5 | 38.7 |
| It has resulted in positive outcomes for students. | 63.7 | 59.5 |
| It has created additional administrative burdens for the school. | 62.9 | 56.4 |
| It has created burdens in dealing with overage students. | 68.8 | 57.2 |
| It has an appeals process that is arbitrary. | 49.1 | 53.2 |
| It has diverted resources to in-need students at the expense of higher-achieving students. | 50.6 | 48.6 |
| It is unfair to the students who are held to the policy. | 20.8 | 25.7 |
| It relies too much on state assessment scores, which are not good measures of students' progress. | 66.2 | 75.9 |
| It reduces teachers' authority over students. | NA | 49.9 |

multiple holdovers and much older than the other students in their grade. For example, one teacher noted that she had taught the same student in 5th-grade summer school for three years in a row and was concerned that the policy was doing more harm to this student than good. Another AIS survey respondent wrote,

> There are students who cannot "make the grade." We have yet to address what to do with such students when they are two to three years older (and bigger) than the other elementary students. These students should have alternative placements. We have so few of these students (two), they stand out. This lowers their self-esteem and brings about behavioral issues.

A few respondents believed that many of these students should have IEPs but did not because of parent resistance. In fact, some students who were retained in grade were referred for special education services during the retained year. Between 5 and 16 percent of students retained in 5th grade over the three years were not held to the policy the following year because they were categorized as special education students.

## Exploring the Relationships Among School Context, Implementation Factors, and Students' Outcomes on the Spring Assessments

Given the variation in school conditions and strategies used to assist students, we were interested in examining whether these factors were associated with better student achievement in ELA and mathematics. We conducted a number of exploratory models by year and subject:

- Using logistic regression models, we examined whether certain school conditions were associated with in-need students' ability to pass the spring assessments, controlling for students' prior achievement (all in-need students in NYC schools).
- For students in the schools for which we had survey responses, we examined the following:
  - the relationship among particular school services, strategies, or contextual factors and the ability of in-need students to pass the spring assessments, controlling for students' prior achievement (logistic regression models)
  - the relationship among school context and implementation factors and *in-need* students' performance, as measured by the scale score on the spring assessments (OLS regression models)
  - the relationship among school context and implementation factors and *all* students' performance, as measured by the scale score on the spring assessments (OLS regression models).

School context variables in the models included the percentage of at-risk students, the percentage of students not held to policy, and NCLB status. From our survey data, we selected a small number of AIS implementation factors that showed some variability across schools and might be considered to influence student achievement. AIS implementation factors modeled included whether the school reported serving all students needing services; providing one-on-one tutoring to some or all of students needing services; placing in-need students in reduced-size classes; and using special teacher assignments, such as assigning the best teachers to 5th grade or assigning the best 5th-grade teachers to 5th-grade students with the weakest skills. Note that because all these variables are derived from survey data, they capture the use of these interventions and strategies but do not capture the quality of their implementation. As such, they are weak measures of implementation.

We also modeled contextual factors derived from our surveys that could influence the implementation of strategies, including indexes of *perceived* hindrances, the level of support received from the region or SSOs, teacher quality, and principal leadership (where modeling was possible).

All models included robust standard errors to correct for the fact that students are nested within schools. Details of these modeling efforts and definitions of the variables used in the modeling are provided in Appendix B.

Because these models were exploratory and explained only a small amount of the variation in student achievement,[4] we examined whether there were consistent patterns over time with respect to the association between different factors and student outcomes. Two factors were consistently and significantly related to better student outcomes across the models:

- *Lower percentage of students needing services.* Across the models, students in schools with lower percentages of students needing services performed better on assessments, after controlling for prior achievement, compared to peers in schools with higher percentages of students needing services. This is not surprising, considering that our earlier findings suggested that the capacity to provide AIS was related to the percentage of students needing services in the school.
- *One-on-one tutoring in a subject (particularly mathematics).* One-on-one tutoring in mathematics was positively and significantly related to students' spring achievement in 2006 and 2007. This finding is consistent with prior research, which suggests that one-on-one tutoring is an effective form of remediation (Wasik and Slavin, 1993).

Perceived hindrances and support provided by the region or district were negatively related to student outcomes in a number of the models. That the level of support provided by the region was negatively related to student achievement can be partly explained by the fact that—as we found in our interviews—regions targeted the weakest schools with the most intensive support. While the additional support was intended to bolster schools' capacity, these results suggest that it may not have been sufficient to overcome the school context effects.

## Summary

In 2006, the first year of our study, there was a substantial decrease in the number of students needing services—both those scoring at Level 1 or low Level 2 on the prior year's assessment and those retained in the 5th grade.

Students needing services were not evenly distributed across schools. Controlling for the prior year's achievement and other factors, students in need of services in schools with greater percentages of students needing services were less likely to meet the promotion threshold than their peers in schools with smaller percentages of students needing services.

[4] The adjusted $R^2$—the proportion of the variation in the dependent variable that is explained by the set of independent variables in the model, adjusted for the number of independent variables used in the model—was between 0.05 and 0.15 for the OLS models.

Schools seemed to serve as many students as they had the capacity to serve. That is, students' access to services was more a function of school capacity than student need. In 2008, a greater percentage of schools extended services to mid- or high Level 2 students and Level 3 students than in 2007, presumably because of the decline in the number of students needing services. High-need schools were less likely to serve all students needing services, compared to other schools.

The majority of schools with reading or mathematics specialists and AIS leaders reported primarily relying on these individuals to provide AIS. Schools were more likely to have reading specialists than mathematics specialists and high-need schools were significantly more likely than low-need schools to have mathematics specialists, reading coaches, and mathematics coaches.

In all three years, almost all schools reported serving students in need of services through small-group tutoring during the school day. Fewer schools were able to offer one-on-one tutoring during the school day (64–80 percent). Our modeling results suggest that being in a school that offered one-on-one tutoring in mathematics to some or all students needing services increased students' probability of passing the spring mathematics assessment. Results for ELA one-on-one tutoring were promising in some models as well.

While in-school services were valuable to students, our case-study visits suggest that services may not have been offered consistently when AIS providers were pulled away for other duties. AIS leaders were more likely to report that services focused on working with a specific intervention program to a moderate or great extent during the school day than outside the school day. However, teaching test-taking strategies was a focus to a moderate or great extent both during the school day and outside the school day.

Schools used a variety of other strategies in addition to AIS to improve the academic performance of in-need 5th-grade students. Interestingly, the use of various student and teacher placement strategies increased significantly in 2008, when principals had more authority to make these decisions. Although principals felt that these strategies were quite effective in improving student performance, they did not appear to have a measurable effect in our exploratory models.

Across the three years, the majority of respondents believed that their region or SSO provided a number of useful supports. On the whole, principals seemed more positive about the type of assistance provided by the SSOs, which was more demand-driven than the region-directed support. However, in our case-study visits in 2008, many AIS leaders noted missing some opportunities that regions provided to all AIS leaders, such as monthly meetings, regular exposure to curricular resources for in-need students, and networking opportunities.

While the majority of schools did not list many major hindrances to improving student achievement, respondents in high-need schools were much more likely to report a given factor as a moderate or great hindrance than respondents in low-need schools. Further, our exploratory models suggest that, controlling for other factors, schools facing higher levels of perceived hindrances had lower levels of student achievement.

Study participants tended to have mixed views of retention in general, and an overwhelming majority (88–91 percent) of respondents agreed or strongly agreed that retention has different impacts on different students. Respondents tended to be more positive about NYC's promotion policy, and approximately three-quarters of AIS leaders and principals agreed that the policy focused their schools' instruction efforts in a positive way and that it had made parents and guardians more concerned about their children's progress. However, the majority of respondents thought that the promotion policy relied too heavily on state assessment scores and, interestingly, that the policy made it more difficult to retain students who should be retained but passed the test.

CHAPTER SIX

# Implementation of the Policy: Saturday and Summer Schools

*Gina Schuyler Ikemoto, Jennifer Sloan McCombs, Catherine DiMartino, and Scott Naftel*

As originally implemented, two key components of the 5th-grade promotion policy were the SPAs and SSAs, through which struggling students were provided with additional instructional time and supports. The SPAs were intended to provide students needing services with between 17 and 24 Saturday sessions beginning in October that included three hours of ELA and mathematics instruction. As described in Chapter Three, by 2007–2008, SPAs were no longer centrally administered; instead, principals decided whether to offer a Saturday school program in their schools. SSAs served students who failed the spring assessments and thus were at risk of retention in the 5th grade. SSAs provided four-and-a-half hours of ELA and mathematics instruction four days per week for six weeks during the summer. Each of these programs was designed to have a low student-teacher ratio (15 to 1) in order to provide students with the benefits of a smaller class size and the opportunity for more individualized instruction.

To the extent that these special support programs are related to student outcomes, it is important to understand their content and structure as implemented. This chapter describes the design and implementation of SPAs and SSAs over time and the factors that helped or hindered implementation. Our analysis is based on a variety of data sources: NYC administrative data from SPAs and SSAs; SPA administrator survey data; SSA administrator survey data; principal survey data; and data from our case studies of SPAs, SSAs, and schools. Unfortunately, our data varied in terms of both source and coverage across time: NYC administrative data for SPAs and SSAs were available for 2005–2006 and 2006–2007 but not for 2007–2008; we had SPA survey and case-study data for 2006–2007 but not for 2005–2006;[1] and we had SSA survey and case-study data for summer 2006 but only case-study data for summer 2007.[2] Finally, our SPA and SSA data for 2007–2008 came from the principal survey (because principals were responsible for these programs in that year).[3]

1 The study began in March 2006, when the SPA sessions had already ended for the school year.

2 NYCDOE was unable to provide contact information for the summer 2007 SSA administrators, partly because SSAs were no longer centrally administered.

3 To minimize the response burden on principals, we asked only a limited set of questions about the Saturday and summer school programs being offered by these schools.

## Saturday Preparatory Academies

NYCDOE established the SPA program in the 2004–2005 school year to provide struggling 5th-grade students with additional instruction during the school year in order to help them gain the skills they needed to be promoted. The number of sessions offered declined over time—from 24 sessions in 2005 to 19 sessions in 2005–2006 and 17 sessions in 2005–2007. Across all years, sessions were to include free breakfast and lunch, three hours of instruction, and an hour of either arts or physical education enrichment. SPAs were to have small classes of no more than 15 students, allowing teachers to provide intensive intervention and instruction based on students' individual needs. Regions were responsible for selecting and providing the curriculum, professional development for teachers, and additional supports or specialized staff. As described earlier, in 2007–2008, NYCDOE stopped requiring SPAs and instead allowed individual principals to decide whether to offer a Saturday school program. This section discusses the population of the SPAs in 2005–2006 and 2006–2007, the implementation of SPAs in 2006–2007, and the implementation of Saturday schools in 2007–2008. For ease of exposition, we refer to each school year by its spring year, when surveys and case studies were conducted (e.g., 2005 for school year 2004–2005); however, SPAs were offered during the fall and winter of the school years as well.

### Profile of Saturday Preparatory Academies and Students

In 2006, 156 SPAs served 9,936 5th-grade students from 590 schools (Table 6.1). On average, each SPA served three schools. Approximately 20 percent of SPAs were "standalone" SPAs, serving students from only one school; 31 percent of SPAs served students from two to three schools; and 49 percent served students from four or more schools. The number of 5th-grade students attending a given SPA ranged from one to 221 students. The median number of students served was 46.

In 2007, there was a decline in the number of students attending SPAs to only 4,880 students from 493 schools. However, the number of SPA sites remained stable (about 150). Not surprisingly, the median number of students served per site declined to 39 students. SPAs in 2007 were somewhat more likely to serve students from only one school, compared to the prior year (29 percent versus 20 percent).

In 2008, about half (49 percent) of principals reported that their school offered a Saturday program that 5th-grade students could attend. Principals of schools with medium or high percentages of students needing services were significantly more likely to offer a Saturday program than were principals of schools with low percentages of in-need students (55–56 percent compared to 24 percent). Of the 23 schools we visited in spring 2008, 13 offered some type of Saturday program. At the 10 schools that did not offer Saturday programs, principals perceived the cost of running the program as outweighing its benefits to students, especially because attendance tended to be so low. The schools that discontinued SPAs decided that other strategies, such as additional

**Table 6.1**
**Selected Characteristics of SPAs and Their Students**

| Characteristic | 2006 | 2007 |
|---|---|---|
| Number of SPAs | 156 | 151 |
| Number of schools sending students | 590 | 493 |
| Percentage of SPAs serving students from | | |
| 1 school | 19.9 | 28.5 |
| 2–3 schools | 30.8 | 33.1 |
| 4–5 schools | 27.6 | 21.9 |
| 6+ schools | 21.8 | 16.6 |
| Median number of students attending SPA (range) | 56.0 (1–221) | 39.0 (1–190) |
| Median number sessions attended (range) | 7 (1–19) | 7 (1–17) |
| Total number of students attending SPA | 9,936 | 6,281 |
| Number of students held to policy attending SPA | 7,066 | 4,880 |
| Number of in-need students attending SPA | 3,294 | 3,004 |
| Number of not-in-need students attending SPA | 3,872 | 1,876 |
| Number of students not held to policy attending SPA | 2,870 | 2,130 |

AIS providers or reduced class sizes, would be more effective in supporting students who were at risk of being retained.

### Student Participation

While SPAs were meant to serve students needing of services at the beginning of the year, most SPAs served other students as well. NYC administrative data show that, in 2006, SPAs served 2,870 students not held to policy (e.g., special education and ELL students). These students accounted for 29 percent of total SPA enrollment. Also, SPAs enrolled a significant percentage of general education students who were not classified as in need of services. In 2006 and 2007, in-need students accounted for 47 percent and 62 percent of the SPA student population, respectively. Fewer than 30 percent of the approximately 12,000 in-need students in 2006 and 2007 participated in SPAs. Studies of other voluntary tutoring or assistance programs offered outside of the school day, such as supplemental education services provided under NCLB, have found similarly low take-up rates for these services (Sunderman and Kim, 2004; GAO, 2006; Vernez et al., 2009).

More than 70 percent of SPA administrators responding to the 2007 survey reported that their SPA served general and special education students who needed additional help but were not prioritized as in need of services, and 15 percent of SPA administrators reported serving higher-achieving (Level 3 or 4) students as well. For example, a few SPA administrators interviewed during our case-study visits reported that some principals sent additional students whom they believed would benefit from extra assistance to SPAs. One principal reportedly sent Level 3 students because his school wanted these students to move up to Level 4. In some cases, the SPA administrators were not expecting these students and had to "scramble" to find additional resources to provide services for these extra students.

While SPAs enrolled a more varied group of students than intended or expected, the majority of administrators in 2007 reported that enrolling these students was not a hindrance to the SPA's ability to assist students and improve their performance. NYCDOE projected student enrollment for the SPAs based on the number of Level 1 students in feeder schools associated with the SPAs. However, because many invited students did not attend, on average, SPAs had lower-than-projected enrollment (by about five students), even after feeder schools had sent students not technically at risk to the SPA. However, the range of the discrepancy between projected and actual enrollment varied considerably across the SPAs—from enrolling 105 students fewer than projected to 85 students more than projected.

On average, students attended about seven sessions in 2006 and 2007. Thus, median attendance rates were fairly low—37 percent and 41 percent in the two years, respectively. In our case studies, we found that class sizes tended to be quite small—often five or fewer students were in the classes we observed—particularly because of high levels of absenteeism, and some teachers in SPAs with very small class sizes noted that this enabled them to provide individual support to students.

In 2008, we found that schools with Saturday sessions tended to be even more inclusive. Almost all (98 percent) invited any general education student needing extra help to attend, and 91 percent invited special education students to Saturday sessions as well. In our case studies, all of the schools we visited that provided Saturday school invited all students in the school to attend. The majority of Saturday schools seemed to continue the earlier SPA focus on small class sizes, and 70 percent of principals operating a Saturday school reported that the average class size was 15 students or fewer. We do not have attendance data for this year.

### Schedule

Our survey and case-study data suggest that SPAs citywide followed a fairly consistent schedule that had been mandated by NYCDOE. In 2006, 19 five-hour sessions were held on Saturdays between October and March. In 2007, however, the program started late due to some delays in administrative decisionmaking. To make up the time, additional sessions were provided during the holiday breaks.

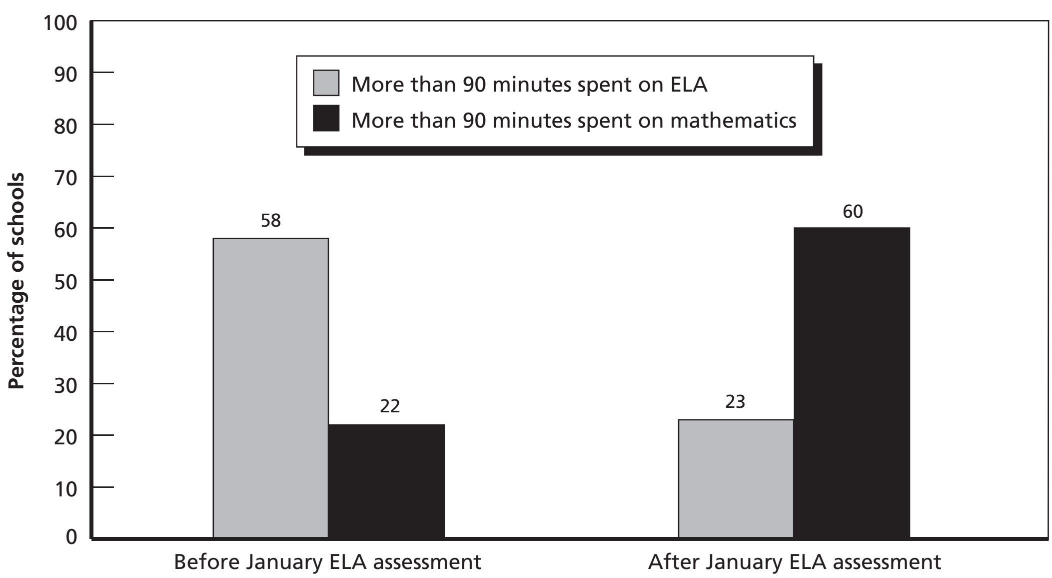
As required, sessions included free breakfast and lunch and at least three hours of instruction. In 2007, 70 percent of SPA administrators reported that all 5th-grade students participated in enrichment activities at their SPAs, and another 17 percent reported that some students participated. A small percentage (17 percent) did not provide an enrichment period. The most common activity was physical education (83 percent). Arts activities were offered in 39 percent of the SPAs, and socioemotional development programs were offered in 20 percent of SPAs. Additional AIS options were provided to at least some 5th-grade students in 79 percent of the SPAs. About half of the SPAs provided these services during ELA or mathematics instruction time.

Many SPAs focused more on ELA before the ELA assessment in January and then spent more time on mathematics after January, in preparation for the spring mathematics assessment (Figure 6.1).

For instance, 58 percent of administrators reported spending more than 90 minutes of instruction time on ELA before the January ELA assessment, while only 23 percent of administrators reported doing so after January; and the reverse was true for mathematics.

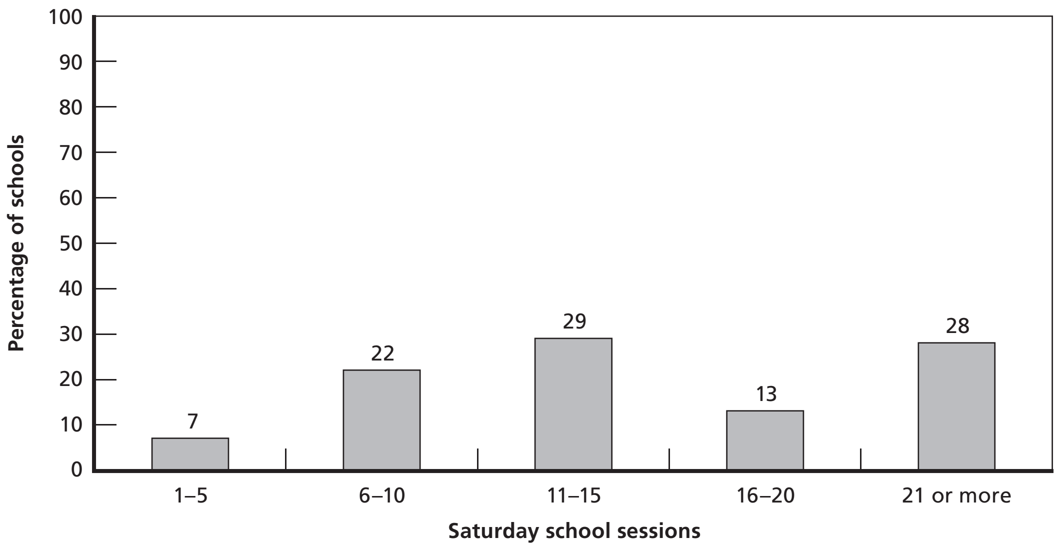
Not surprisingly, we found variations in Saturday programming in 2008. The number of sessions that principals reported offering ranged from one to 30. A small percentage (7 percent) offered one to five Saturday school sessions, while a larger percentage (28 percent) offered 21 or more (Figure 6.2). In the case-study schools that were operating Saturday programs, half structured their schedule after the SPAs and ran their programs from October to March. The other half of these sites ran their programs a few weeks before the ELA and mathematics state assessments.

**Figure 6.1**
**Time Spent on ELA and Mathematics Before and After ELA Assessment, 2007**




RAND *MG894-6.1*

**Figure 6.2**
**Number of Sessions Offered by Schools Operating a Saturday Program, 2008**




RAND *MG894-6.2*

The vast majority of these programs operated in either three-hour (59 percent) or four-hour (30 percent) sessions, while a small minority were held for two hours (4 percent) or five or six hours (7 percent). Approximately half (53 percent) of principals reported that their Saturday program did not include enrichment activities, and only 25 percent offered enrichment activities for all students attending the Saturday program.

## Staffing

Overall, the vast majority of administrators surveyed in 2007 reported being very satisfied with the quality of teachers in their school in terms of their attitude and knowledge. Note that NYCDOE hired NYC teachers to staff SPAs, so these teachers were qualified professionals and familiar with NYC's standards and school context. More than 90 percent of SPA administrators agreed or strongly agreed that teachers at the SPA

- had a "can-do" attitude
- worked hard to help their students succeed
- really cared about their students
- were knowledgeable about the 5th-grade curriculum
- were knowledgeable about the subjects they were assigned to teach
- were knowledgeable about strategies for struggling students.

In our case-study SPAs, half the administrators used such terms as "highly qualified," "very satisfied," "excellent teachers," "very lucky," and "very happy" when responding to questions about the quality of the SPA teachers.

In addition to teachers, some SPAs had specialists who worked with students or teachers. Approximately 17 percent of administrators reported that their SPA had a Wilson teacher,[4] and 23 percent reported having another ELA intervention teacher. Very few SPAs (only four) reported having a Wilson teacher *and* another ELA intervention teacher. Almost one-third of SPAs had a mathematics intervention teacher. SPA administrators were more likely to report having a mathematics coach (27 percent) than an ELA coach (11 percent), but the majority of SPAs did not have either a mathematics coach or an ELA coach (67 percent).

Well over half of SPAs reported having teacher aides (57 percent), and 37 percent of SPAs reported that teachers used teacher aides or paraprofessionals to a moderate or great extent in the classroom (not shown). Forty-five percent of administrators reported having a parent liaison, and almost 60 percent had a guidance counselor, though these positions may have been filled on a part-time basis.[5]

In 2008, principals were more likely to have intervention teachers and coaches involved in their Saturday programs than in the SPAs in 2007. For instance, 68 percent reported having an ELA intervention teacher and 61 percent reported having a mathematics intervention teacher at the Saturday program. We did not ask principals whether these positions were used to supplement regular classroom instruction, however. It could be that principals asked intervention teachers and coaches in their schools to teach at Saturday schools, instead of assigning regular classroom teachers to this task.

### Curriculum

In the first two years of our study, the regions selected the curriculum materials for the SPAs. Although the materials selected for the SPAs differed somewhat by region, administrators across the regions mentioned using Great Source Reading Advantage, Great Leaps, and/or iOpeners for the ELA curriculum, and McGraw-Hill, Macmillan, Breakaway, and/or Navigator for the mathematics curriculum. Administrators overwhelmingly ranked both the ELA and mathematics curricula highly, with between 82 and 98 percent agreeing or strongly agreeing that the curricula met the needs of the majority of students identified as at risk of being retained, that they were aligned with the New York State subject-content standards and the state 5th-grade subject test, and that they were effective in improving student achievement. However, between

[4] The Wilson Reading System is a research-based program designed for students who struggle with decoding and spelling. Teachers trained through the program are referred to as "Wilson teachers."

[5] The survey also asked administrators whether they had an enrichment teacher. Note that there was a substantial difference between the percentage of administrators who reported having an enrichment teacher (24 percent) and those who reported offering enrichment activities (72 percent). Thus, administrators may have misunderstood the question, or other teachers were being used to lead enrichment activities.

67 and 70 percent agreed or strongly agreed that the curricula needed to be supplemented with additional materials, and in our case studies, teachers were more critical of the materials overall than were the administrators. For example, in half of our case-study sites, teachers commented that the literacy curriculum was too difficult for in-need students or that it was not well aligned with the state ELA test. In general, teachers were also dissatisfied with the mathematics curriculum that they had received (this was true across a variety of curricula). Criticisms included that it was not aligned with the mathematics test, that it was too difficult and too fast-paced, that materials were not broken down into usable lessons, and that it did not meet students' needs.

In 2008—when schools were allowed to make their own curriculum choices—principals in our case-study schools reported that they tended to choose test-focused curricula. Curricula used by 12 of the Saturday programs we visited included New York State Coach, Kaplan, Buckle Down, and STARS.[6] Depending on the needs of their students, some of the case-study schools providing the Saturday program in 2008 offered it for only for one subject: either mathematics or ELA. In contrast to these test-focused programs, one school offered an enrichment Saturday program.

### Professional Development

Each SPA was expected to provide 12 hours of in-school professional development in 2007. Although 84 percent of SPA administrators reported doing so, the amount of time varied considerably, ranging from two to 25 hours, with a median of approximately 10 hours. SPA administrators reported that the content tended to focus on introducing or reviewing the ELA and mathematics curricula, differentiating instruction, discussing processes for portfolios and/or grade reports, and teaching test-taking strategies. However, many case-study teachers noted that the SPA-provided professional development time tended to be used for administrative tasks, such as preparing portfolios, or as teacher preparation time. Some teachers expressed a desire to have more specific hands-on training, such as coaches modeling lessons. A few teachers reported that, while the professional development was helpful for teachers who were new to the SPA program, it was not as helpful for experienced teachers. In addition, they reported that at least five of the regions provided professional development before the start of the SPA session to review programs. Some teachers could not attend, however, because of the late start of the SPA session and the late notice of hiring decisions. In 2008, SSOs did not provide professional development support for Saturday school, and we did not collect data about in-service professional development opportunities.

### Communication Between SPAs and Schools

Communication between SPAs and schools tended to be sporadic or ineffective. NYCDOE intended for schools sending students to SPAs to provide the SPA with a

[6] Details regarding perceptions of usefulness of the curricula were not collected in 2008.

portfolio for each student that could assist the SPA in understanding and addressing individual student needs. In return, SPAs were to update schools about students' progress. Well over one-third of SPA administrators responding to our 2007 survey reported that teachers never received portfolios for the majority of their students. None of the 10 SPAs we visited in 2007 had received portfolios. Survey respondents indicated that—when they did receive portfolios—the portfolios contained a range of information, including student work, teacher notes, and results from teacher-made assessments, but only 55 percent of the administrators reported that they contained assessment information, such as results from city interim assessments or state tests.

In 2007, almost all the administrators reported calling or sending emails to schools regarding student attendance, and a little over 70 percent reported doing so with respect to student performance. The majority of administrators reported providing information to schools through student intervention folders or portfolios and through written reports of student progress. However, a significant minority did not attempt to communicate with the home schools regarding student performance or progress. For example, 16 percent of the administrators did not communicate at all with the home schools regarding student performance via phone or email, and 31 percent did not provide written reports of student progress. There was substantial overlap between these two categories, suggesting that some administrators failed to provide any reports on student progress or performance.

Even when SPAs reported communicating about student progress, in some cases the information did not reach the AIS leader. In 2007, 38 percent of AIS leaders responded that they did not know about communication between the SPA and the school, while another 34 percent disagreed or strongly disagreed that there was frequent communication between the SPA and the school.

### Perceived Effectiveness of SPAs

When asked about the perceived effectiveness of SPAs in 2006 and 2007, a substantial minority of principals (20–28 percent) and AIS leaders (47–48 percent) reported not knowing about the effectiveness of the programs (Table 6.2). Given that communication between schools and SPAs was not always effective, this may not be surprising. Interestingly, in 2007, the percentage of AIS leaders reporting not knowing was significantly higher in schools that did not host an SPA than in schools that did (61 percent versus approximately 27 percent). However, the majority of respondents who felt able to comment on the effectiveness of SPAs agreed that they were helpful in improving students' skills in reading, mathematics, and test-taking.

It is not surprising that the vast majority of principals (90 percent or more) who held Saturday school programs in 2008 agreed that the program helped students improve their skills in reading, mathematics, and test-taking. It is likely that principals who believed that Saturday school would be effective were the ones who decided to provide Saturday school. In addition, the principals were much more closely connected

**Table 6.2**
**Percentage of Administrators Agreeing About Aspects of SPA and Saturday Programs' Effectiveness**

| | 2006 | | 2007 | | 2008 | |
|---|---|---|---|---|---|---|
| Statement | Don't Know | Agree or Strongly Agree | Don't Know | Agree or Strongly Agree | Don't Know | Agree or Strongly Agree |
| Attending the Saturday program helped students improve their academic skills in reading | 19.9 | 68.9 | 47.0 | 40.3 | 1.1 | 91.7 |
| Attending the Saturday program helped students improve their academic skills in mathematics | 20.3 | 66.7 | 47.8 | 40.0 | 1.7 | 90.3 |
| The Saturday program helped students gain test-taking skills | 28.2 | 61.8 | 47.1 | 43.1 | 0.0 | 91.3 |

to the Saturday program; thus, very few principals reported that they did not know about the effectiveness of Saturday school.

## Summer Success Academies

Students at risk of retention under the promotion policy were invited to attend SSAs and take a summer assessment in August. SSAs operated as part of a larger summer school that served other grade levels. While similar in many ways to typical summer school programs, SSA programs were provided with additional resources that permitted small class sizes and the hiring of additional intervention teachers. In summer 2006, NYCDOE established certain system-wide requirements for the SSAs, including that each SSA must operate four days per week for six weeks during the summer. Each SSA day included 4.5 hours of instruction, plus a half-hour for lunch. Class size was capped at 15 students, and each SSA was required to provide instruction in ELA and mathematics, to have at least one Wilson teacher, to hold a parent-teacher conference, and to provide teachers with eight hours of in-service professional development. Each region had discretion over curriculum, similarly to the SPAs, as well as pre-service professional development provided to teachers and additional supports or specialized staff provided to the SSAs.

Beginning in summer 2007, principals, rather than the regions, became responsible for organizing SSAs for their students. While principals were provided greater flexibility in running SSAs, NYCDOE mandated the provision of summer school services to students at risk of being retained and a lower student-teacher ratio for these students. NYCDOE encouraged, but did not explicitly require, SSAs to include other program components, including the Wilson program, professional development for teachers, and parent-teacher conferences. As a result of the increased autonomy, the structure and content of SSAs became more varied in 2007 as principals adapted the program to align with the context of their schools.

### Profile of Summer Success Academies

In summer 2006, 4,029 students at risk of being retained under the promotion policy attended an SSA; this number dropped to 2,251 in summer 2007 (Table 6.3). In summer 2006, a total of 270 SSAs provided services to students from 610 schools; in summer 2007, the number of SSAs had increased to 293, although the number of schools from which they drew their students was considerably smaller: 523. Given this trend, it is not surprising that the percentage of "stand-alone" SSAs—those serving only students from the host school—increased between 2006 and 2007 (24 percent versus 33 percent).

SSAs enrolled between one and 51 5th-grade students at risk of being retained under the policy in summer 2006, with a median of 14. The median number of 5th-grade students at risk of being retained dropped significantly, to five, in summer 2007. Data for summer 2008 were unavailable.

### Student Participation

Summer schools also enrolled a number of 5th-grade students who were not held to the policy as well as students who were held to the policy but were not at risk of being retained in grade (not shown). For instance, in summer 2006, 3,120 students who were not at risk of retention attended SSAs, and 868 did so in 2007. In our summer 2006 case

**Table 6.3**
**Selected Characteristics of SSAs and Their Students**

| Characteristic | Summer 2006 | Summer 2007 |
|---|---|---|
| Number of at-risk students held to the policy attending | 4,131 | 1,685 |
| Number of SSAs | 263 | 261 |
| Percentage of SSAs serving students from | | |
| 1 school | 30.4 | 36.4 |
| 2–3 schools | 54.8 | 47.5 |
| 4–5 schools | 12.2 | 14.6 |
| 6+ schools | 2.7 | 0.5 |
| Median number of students attending SSA (range) | 14 (1–51) | 5 (1–32) |
| Median number sessions attended (range) | 16 (1–21) | 18 (1–22) |

studies, we found that the majority of the 5th-grade SSA students had been mandated to attend. Because the state had not established performance levels for the new state assessment prior to summer school,[7] NYCDOE had to establish temporary cutoff scores to determine who would go to summer school. In an effort to ensure that all Level 1 students would be mandated to attend summer school, these temporary cutoff scores included students who eventually were found to have scored low Level 2 on the state assessment. In addition, we found that a few of the 20 SSAs we visited also enrolled a considerable number of nonmandated students, and, in one school, half of the 5th-grade students were nonmandated. Nonmandated students were referred to as "enrichment" students, or "NIs" (needs improvement) and were typically attending because their principal or teacher had recommended that they receive extra help during the summer, despite the fact that they had met the promotional criteria. In one region, almost half of the 5th-grade students at the SSAs we visited were enrichment students, while in other regions serving nonmandated 5th graders, the enrichment students generally represented between 5 and 20 percent of the 5th-grade population. Although class sizes were supposed to be capped at 15, the majority of the SSAs we visited in 2006 had class sizes between 15 and 18 students, and we observed class sizes as high as 30 students in one SSA. We noted that the SSAs with larger class sizes tended to be those that enrolled a significant number of "enrichment" students who were not held to the promotion policy.

In summer 2007, significantly fewer students were mandated to attend SSAs than in 2006. This occurred in part because NYCDOE had formalized its performance levels and in part because student test scores had improved, resulting in fewer students being mandated to attend SSA. Nonetheless, SSA programs continued to vary in terms of the types of students they invited to attend. Five of the SSAs focused exclusively or almost exclusively on serving mandated students. The others reported serving a mix of both mandated and nonmandated students. One SSA reported that only 15 percent of its 5th-grade students were mandated and that the program was open to students who had received a low Level 2 score on the state tests. One SSA determined how many slots would be needed for mandated students and allocated the remaining slots to allow each feeder school (i.e., schools sending students to the school that hosted the SSA program) to invite nonmandated students. Administrators and teachers in two SSAs specifically reported that their SSAs enrolled students who had scored a Level 1 on the state ELA or mathematics test but were not at risk of retention under the policy due to their IEP or ELL status.

In general, the SSAs we visited in 2007 were more successful in maintaining small class sizes than those that we visited in 2006. In 2007, most SSAs we visited maintained class sizes of seven to 15 students, while two SSAs had class sizes that ranged from 16 to 20 students. Respondents did not provide explanations for the

---

7 The establishment of cutoff scores in 2006 is discussed further in Chapter Eleven.

greater prevalence of small classes, but it is likely that this was at least partly driven by the decreased numbers of at-risk students mandated to attend SSAs.

Student attendance remained consistent across 2006 and 2007. The median number of sessions attended by students at risk of being retained under the policy was 16 and 18, respectively, in the two years, meaning that half the students attended more than 16 sessions in 2006 and more than 18 sessions in 2007 (between 76 and 81 percent of the sessions offered), and half attended less frequently.

#### Schedule

In summer 2006, all case-study and survey respondents reported following the centralized schedule for SSAs. They all operated from Monday through Thursday, July 5 through August 9, from 8:30 a.m. to 1:30 p.m., with lunch provided during the last half hour. All of the schools we visited also provided breakfast from 8:00 a.m. to 8:30 a.m. The 4.5 hours of instructional time generally included at least 1.5 hours of ELA, 1.5 hours of mathematics, and one hour of intervention. SSAs varied markedly in terms of how they used the last half hour of instructional time. Only five of the 20 SSAs we visited offered a half hour of enrichment to students on a daily basis, as suggested by NYCDOE. Four of the SSAs provided enrichment to students two or three days per week, and 11 sites did not offer a formal enrichment program. These sites offered additional instruction in ELA, mathematics, or intervention instead.

Although NYCDOE issued similar guidelines regarding SSA schedules for summer 2007, we found that program schedules and durations varied across the 10 SSAs that we visited. Seven of the 10 SSAs implemented at least four hours of classroom instruction for 19 instructional days, but three SSAs chose to deviate from this schedule. The most marked difference was an SSA that condensed its schedule into 12 days of instruction, during which students received approximately five hours of classroom instruction, including a two-hour intervention period. Two other SSAs chose to shorten their instructional day, providing three hours of classroom instruction. They reported that these changes were intended to improve student and teacher participation and attendance.

#### Staffing

In summer 2006, NYCDOE required each SSA to have at least one Wilson teacher in addition to classroom teachers. In our surveys, 90 percent of SSA administrators reported having a Wilson teacher (not shown). However, four of the 20 sites we visited were not providing the Wilson program to 5th-grade students, primarily because the Wilson teacher had left the SSA early in the session. SSAs were significantly less likely to have a mathematics intervention teacher (23 percent), even though the majority of students were sent to the SSA because they failed the mathematics assessment. Approximately one-third of SSAs reported having an ELA or mathematics coach, and some SSAs had teacher aides, teaching fellows, or summer interns to help in classrooms as well.

### Curriculum

In 2006, the regions mandated the curricula used in the SSAs. With the exception of one region, curricula were characterized as highly scripted and followed a set pacing schedule. Many teachers responded positively to the scripted curricula, given that the short duration of summer school limited their preparation time. The majority of supervisors responding to our survey in 2006 agreed or strongly agreed that the ELA and mathematics curricula met the needs of the majority of students, were well aligned with the New York State content standards and the NYC summer assessment, and were effective in improving student achievement. In 2008, principal survey respondents were similarly positive about the curricula, with 76 percent reporting that a "lack of high-quality curriculum materials to meet student needs" was not a hindrance to their summer school's ability to improve 5th-grade students' academic performance.

Teachers in our case-study sites tended to provide more negative or mixed reviews of the curriculum materials, often because they thought the curricula were not aligned to their students' needs. Other teacher criticisms were that the curricula failed to cover the range of topics addressed by the exam or that the pacing was too fast, particularly for students who were significantly below grade level.

Even though schools were given more autonomy to choose curriculum materials in 2007, many SSA administrators indicated that they were using the same materials as in 2006 because there was insufficient time to thoughtfully research and choose a different curriculum. Due to our more limited data on 2008 SSAs, we do not know whether autonomy was used to exercise more changes in the curriculum in 2008.

### Intervention

In 2006, almost all site-visit classroom teachers taught an in-class intervention period that used materials that differed from the core SSA curriculum. During in-class intervention, teachers focused on test preparation or used other supplemental materials or series provided by the region (e.g., Kaplan, Rewards, iOpeners). Teachers generally tended to be pleased with the intervention materials. In contrast, in 2007, only two SSAs that we visited implemented a planned intervention period. Instead, five SSAs scheduled a "test-preparation" or "test-sophistication" period, during which they used additional curricula—such as Test Ready, Kaplan, Princeton Review, or Finish Line—to provide students with test-taking skills and strategies.

### Time for Professional Development

In 2006, SSAs typically provided pre-service and in-service professional development for teachers. In general, pre-service professional development involved either one or one-and-a-half days of curriculum training, often delivered by the regions. The other portion of the pre-service professional development was generally administrative in nature and took place at the SSA. During this time, teachers typically received classroom assignments and lists and set up their classrooms. Also in 2006, per NYCDOE requirements, all the SSAs provided eight hours of in-service training to teachers, which tended to be administrative in nature (e.g., how to assemble the promotion portfolios).

In 2007, SSAs continued to provide teachers with pre-service professional development. However, it tended to be two half-day sessions, which was approximately half of the two days provided in 2006. Furthermore, in-service professional development was, in many cases, no longer provided. This was not surprising, given that NYCDOE no longer required SSAs to provide professional development. Without such paid time for meeting and planning, some administrators noted that it was more difficult to communicate with teachers.

### Portfolios

Per NYCDOE guidelines, schools were supposed to send students' promotion portfolios to the SSA that their students attended. The portfolios were intended to help SSAs understand the specific needs of individual students so that they could differentiate instruction accordingly. However, only 9 percent of the administrators surveyed in 2006 reported that the 5th-grade teachers received the majority of their students' portfolios on or before the first day of SSA, and 68 percent reported that teachers did not receive this information until after the first week of SSA. Both case-study and survey respondents indicated that the portfolios varied in terms of content and completeness. While the city-mandated portfolio assessments should have been in all student promotion portfolios, only 68 percent of SSA administrators reported that these assessments were included in all the students' portfolios.

## Factors Perceived to Affect Implementation of SPAs and SSAs

### Factors Hindering Implementation

**Poor Attendance.** About 65 percent of SPA administrators in 2006 reported poor attendance as a moderate or great hindrance to efforts to improve students' academic performance. In our case-study visits, the majority of administrators interviewed reported low attendance, particularly during vacations, and two noted that attendance dropped off after the ELA test in January. For example, attendance in the SPAs that we visited during the week of February break was typically below 50 percent. Teachers hypothesized that parents either were taking their children away for the break or had to work during the week and could not arrange a child-care solution that permitted them to send their child to SPA for half a day (for example, children who were staying with grandparents in another part of the city that week might not be able to attend).

A few respondents thought that distance from the SPA and transportation issues contributed to low attendance among students who were not attending SPA at their home school. On the other hand, one administrator listed the provision of bus service as one of the most important reasons that her SPA had high student attendance. On the survey, administrators of stand-alone SPAs were significantly more likely to report higher attendance rates than were administrators of host SPAs, which served multiple schools. Administrative data do show slightly better attendance at stand-alone SPAs than host SPAs. We also heard anecdotally that changes to the program structure in 2008 helped improve attendance. For example, one principal noted,

> In previous years students participated in Regional Saturday programs. These programs lasted months and attendance dropped during the course. Parents in my school prefer the shorter intense program. They are able to schedule their time better and make sure the children attend the sessions. This has proven to be more effective than a program that runs every Saturday over several months.

Poor attendance was also reported to be a problem for SSAs, but to a lesser extent than for SPAs. For example, 20 percent of SSA administrators—compared to 65 percent of SPA administrators—reported that poor attendance was a moderate or great hindrance to efforts to improve students' academic performance. Attendance may have been less of a problem for SSAs because the schedule was easier for parents logistically or, perhaps more likely, because the stakes were higher for students invited to attend an SSA. By summertime, parents and students knew that students would be retained unless they improved their performance on the summer test.

**Late Materials.** The timeliness of receipt of curriculum materials was clearly an issue for SPAs in 2007. A majority (62–66 percent) of the administrators reported that the materials did not arrive in time to allow teachers to become familiar with the content and prepared to use them with students. All 10 of our case-study SPAs reported that the curriculum materials (for either one subject or both) arrived late. As of February—when the program was about to end—six of the 10 SPAs still had not received the materials they needed or had just received them.

During the 2006 site visits, the majority of SSAs also reported at least one problem in receiving curriculum materials and classroom supplies. While almost all sites reported receiving their literacy materials on time, the mathematics materials arrived late or were missing at a majority of sites. Eight SSAs that year received at least one set of curriculum materials late; six reported not receiving part of their materials (e.g., promised intervention programs, leveled readers); five reported insufficient quantities of materials (e.g., mathematics workbooks) and had to rely on the teacher to make copies; and six reported that they did not obtain classroom supplies before the start of the SSA. Not surprisingly, these experiences varied across regions.

**Lack of Communication and Coordination.** In June 2007, SSA administrators reported that they were hindered by the fact that NYCDOE and the regions failed to provide adequate time and information for them to plan the SSAs. A few months before the start of the 2006 SSAs, authority for designing and planning SSAs shifted from the region to principals. This was a new responsibility for principals and site administrators, and they reported that they often did not have the logistical information they needed to make decisions in a timely manner. In the words of one site supervisor,

> [The SSA set-up] was a challenge this year because the region was disappearing. Normally you would get emails at the beginning of April for setting up the program, the ordering of materials, curriculum, et cetera—paced at a normal rate for a human being. This year at the end of May, everything [for SSA set-up] went into top speed. We didn't know about funding, curriculum, all of the stuff [you need to know to set up the program]. A lot of that fell on our principals and me.

As a result, this SSA reportedly did not have time to order materials and instead relied on materials that the school had on hand from the regular school year and previous SSAs. This issue continued to be a problem in 2008, when principals reported that "delays in notification about funding levels and logistical information for summer school" was one of the biggest hindrances to SSA implementation, with 41 percent reporting this to be a moderate or great hindrance.

While SPAs did not report any communication problems with the regions, they did report problems with poor communication between host and home schools. In 2007, four of the 10 SPA administrators interviewed were not associated with the host school, and two of these administrators reported that this caused major problems that affected the quality of instruction for students. In particular, they were unable to find needed materials or use copiers to generate needed materials. The other two administrators reported that they were lucky in that they already had strong relationships or were friends with the host principal. On the other hand, when administrators were associated with the host school, they were able to easily use copiers and gain access to other school resources. Even when relations were good with host schools, teachers often reported that their students did not have access to materials—such as leveled readers and manipulatives—that they might have otherwise had if they were teaching in their own classrooms. In many cases, teachers were teaching in different schools, and these materials were too bulky to transport back and forth.

#### Factors Enabling Implementation

**Administrator Knowledge and Experience.** In the later years of the study, SPAs and SSAs were reportedly enabled by prior knowledge and experience implementing these programs. For example, eight of the 10 SSA supervisors interviewed in 2007 had overseen SSAs in the past. These supervisors said that they could more efficiently run the programs after their first year. Teachers similarly reported that the programs were more efficient when the administrator was familiar with the program and the host school.

**Small Class Sizes.** When class sizes were small, teachers reported that they enabled implementation. SSA teachers interviewed in 2007 were particularly positive about their class sizes. Eight of the 10 SSAs reported class sizes of 13 or fewer students. Although teachers from two of these schools preferred to have even smaller class sizes, the teachers from the two schools with class sizes of 15 to 20 students did not view these larger class sizes as problematic. Teachers said that the smaller class sizes enabled them to better manage student engagement and better attend to individual students' needs. SPA teachers were similarly positive about their class sizes.

**Teacher Knowledge and Experience.** In 2007, SSA interviewees said that teacher experience and background enabled implementation. Seven of the 10 site supervisors reported being satisfied or very pleased with teacher quality. One site supervisor attributed the quality of teachers to the fact that principals were allowed to make hiring decisions and give preference to teachers who taught at the host or sending schools: "I think it is a good thing because teachers know the students." This policy differed from that in 2006, when the region typically assigned teachers to sites.

In 2007, SPA administrators were similarly positive, with eight of 10 administrators reporting that they were satisfied with the quality of teachers. Half the administrators seemed particularly pleased, using such terms as "highly qualified," "very satisfied," "excellent teachers," "very lucky," and "very happy" when responding to questions about the quality of their teachers.

However, 41 percent of principal survey respondents in 2008 reported that a "lack of teacher knowledge regarding how to support struggling learners" and "lack of teacher experience in teaching 5th grade" were moderate or great hindrances for their summer programs.[8] Almost 60 percent of principals said that "union policies that require the hiring of particular teachers for the summer program" were a moderate or great hindrance, suggesting that they would have preferred to hire different teachers. This finding was interesting because—as noted in the previous paragraph—principals were given more input into the hiring process in the later years of our study.

## Summary

Participation and attendance in the voluntary SPAs tended to be low, and SPAs served a larger population than simply students who were in need of services. This was due, in part, to low participation rates among in-need students. After the city stopped man-

---

[8] We did not ask principals questions about teacher quality in Saturday programs in 2008, when more than half of the principals decided not to hold a summer program.

dating SPA in 2007–2008, about half of 5th-grade schools decided to offer a Saturday program to students, which was typically open to all students in the school.

Participation and attendance rates were relatively high in the SSA program. As the number of students at risk of retention declined over time throughout the city, the number of students attending SSAs also declined.

Both SSAs and SPAs (when offered) consistently focused instructional time on ELA and mathematics and tried to maintain small class sizes of 15 or fewer students. As principals were given more autonomy, both programs began to vary more in terms of schedule and types of students invited to attend. The SPAs also varied a great deal in terms of the number of sessions offered. Both programs offered less pre-service and in-service professional development for teachers over time.

The implementation of both types of programs was hindered by common problems, including poor student attendance (for SPAs), late materials, and lack of communication and coordination. The programs were also affected by several enablers, including knowledgeable and experienced administrators and small class sizes. SSA administrators also noted that teacher knowledge and experience was an enabler. Overall, issues hindering implementation seem to have improved over time, perhaps because the increased autonomy allowed sites to address them more effectively or because the number of students requiring services declined.

CHAPTER SEVEN

# Performance of 5th Graders in New York City and Overall Performance Trends in New York State

*Sheila Nataraj Kirby, Scott Naftel, Jennifer Sloan McCombs, Daniel Gershwin, and Al Crego*

This chapter examines trends in student performance prior to the policy (C0 cohort), under the policy (P1, P2, and P3 cohorts), and across the state as a whole.[1] The results presented here are purely descriptive: Our statistical models that estimate the effect of the promotion policy and its components on student achievement are presented in Chapters Eight and Nine. This chapter answers key descriptive questions regarding the performance of students under the policy: How many students were held to the policy? How many students were "in need" of services? How did students perform on the spring assessments? How did students at risk of retention perform on the summer assessments? What were students' final promotion outcomes? How did at-risk students perform in future years? In addition, the chapter compares the performance trends in NYC with those in the rest of New York State in order to situate the trends in NYC against the broader state context and provide a more nuanced perspective on the city findings.

The chapter is divided into two major sections—the first presents data on the composition, status, and performance of 5th-grade students in our four NYC analysis cohorts; the second presents data on performance trends in grades 3–8 from 2005–2006 through 2007–2008, comparing NYC students with those in the rest of New York. Appendix C provides supporting tables and figures for this chapter.

## Performance of the NYC 5th-Grade Cohorts

### Students Who Were Subject to the Promotion Policy

Under the regulations, some students are exempt from the promotion policy: (1) ELL students who have been in the school system for three or fewer years; (2) students with IEPs, generally referred to as special education students; and (3) students attending District 84, the administrative district for the city's charter schools, in which all students are exempt from the promotion policy.

[1] Appendix E provides similar data for the 3rd-grade cohorts.

Across the three policy cohorts, 75–78 percent of students were held to the promotion policy. We constructed a held-to-policy indicator for the comparison cohort by using the criteria in the previous paragraph. Students who met one of these criteria were coded as not held to policy. Because 2–3 percent of students in the policy cohorts were not held to policy for reasons other than those stated (e.g., ELL waiver), we are slightly overestimating the percentage students who would have been held to policy in the comparison cohort.

Table 7.1 shows the total number of students in our four cohorts, the percentage of students who were exempt for various reasons, and the percentage held to policy.

Boys,[2] Hispanic students, students eligible for free or reduced-price lunch, and those with more frequent absences from school (16 or more days in a school year) tended to be exempt from the policy at higher rates than their counterparts.

### Overall Performance on the 5th-Grade Spring Assessments

Before presenting the figures that map the progress of students and provide summary information for each cohort, we present a table showing the overall performance of students in the comparison and policy cohorts on the 5th-grade spring assessments. This helps provide the context for the subsequent discussion.

In examining these results, it is important to remember that 5th graders took the city test in spring 2004 and 2005 and the state test in spring of 2006 and 2007, and

**Table 7.1**
**Students Classified as English Language Learner, Special Education, and Held to Policy, by Cohort**

| Cohort Characteristic | C0 Cohort | P1 Cohort | P2 Cohort | P3 Cohort |
|---|---|---|---|---|
| Number of students | 82,575 | 80,768 | 77,857 | 75,954 |
| Number of students held to policy | 66,544 | 63,101 | 58,695 | 57,762 |
| Percentage held to policy | 80.6 | 78.1 | 75.4 | 76.1 |
| Percentage not held to policy | 19.4 | 21.9 | 24.6 | 23.9 |
| ELL for 3 or fewer years | 4.4 | 4.1 | 4.3 | 4.1 |
| Special education | 14.0 | 14.5 | 14.8 | 14.0 |
| ELL and special education | 0.6 | 0.6 | 0.9 | 0.5 |
| In District 84 (charter schools) | 0.4 | 0.9 | 1.7 | 2.1 |
| Other | — | 1.8 | 2.9 | 3.2 |

NOTE: The comparison cohort was not held to policy. For this cohort, the number of students held to the policy was estimated using policy guidelines.

2 Boys were much more likely than girls to be classified as special education students.

**Table 7.2**
**Performance of All Held-to-Policy Students on the 5th-Grade Spring Assessments, by Subject and Cohort**

| | % Held to Policy | | | |
|---|---|---|---|---|
| **Performance Level and Subject** | **C0 Cohort** | **P1 Cohort** | **P2 Cohort** | **P3 Cohort** |
| ELA | | | | |
| Level 1 | 10.6 | 2.6 | 3.9 | 2.1 |
| Level 2 | 34.3 | 21.6 | 31.7 | 32.9 |
| Level 3–4 | 55.2 | 75.6 | 64.4 | 65.0 |
| Mathematics | | | | |
| Level 1 | 19.3 | 7.8 | 6.6 | 2.4 |
| Level 2 | 36.5 | 30.5 | 23.2 | 17.5 |
| Level 3–4 | 44.2 | 61.7 | 70.2 | 80.1 |
| Total held to policy | 66,544 | 63,101 | 58,695 | 57,762 |

NOTE: The comparison cohort was not held to the policy. For this cohort, the number of students held to the policy was estimated using policy guidelines.

these assessments are not directly comparable, as explained earlier. Thus, a change in the percentage of students scoring Level 1 between the P1 and P2 cohorts may not represent a change in the knowledge, skills, or abilities of 5th-grade students in these two cohorts.

In each policy cohort, well over 50 percent of held-to-policy students achieved a Level 3 or Level 4 score on the ELA assessment, indicating proficiency under NCLB requirements. Although nearly 11 percent of the C0 cohort scored Level 1 on the ELA assessment, future cohorts performed better, with only 2–4 percent scoring Level 1. In the P1 cohort, 76 percent of students scored Level 3 or Level 4, but this proficiency fell as students began to take the state assessment. About 65 percent of students in the P3 cohort achieved proficiency on the ELA assessment.

In mathematics, the percentage of students scoring Level 1 dropped dramatically over time, from 19 percent in the C0 cohort to 2 percent in the P3 cohort. Simultaneously, a larger percentage of students scored Level 3 or Level 4, from 44 percent in the C0 cohort to 80 percent in the P3 cohort.

Table 7.3 further disaggregates the risk of retention by subject to determine whether students were more likely to fail to meet the promotion criteria in one subject. In general, with the exception of the P3 cohort, in which the risk of retention was uniformly low, students were at greater risk in mathematics than ELA following the spring assessments.

**Table 7.3**
**Students at Risk of Retention, 5th-Grade Spring Assessments, by Subject and Cohort**

| | Percentage at Risk | | | |
|---|---|---|---|---|
| **Subject** | **C0 Cohort** | **P1 Cohort** | **P2 Cohort** | **P3 Cohort** |
| Total (any subject) | 22.4 | 8.7 | 8.9 | 3.9 |
| ELA only | 3.7 | 1.1 | 2.4 | 1.5 |
| Mathematics only | 12.4 | 6.2 | 5.2 | 1.8 |
| Both ELA and mathematics | 6.3 | 1.4 | 1.4 | 0.6 |
| Total held to policy | 66,544 | 63,101 | 58,695 | 57,762 |

NOTE: The comparison cohort was not held to the policy. For this cohort, the number of students held to the policy was estimated using policy guidelines.

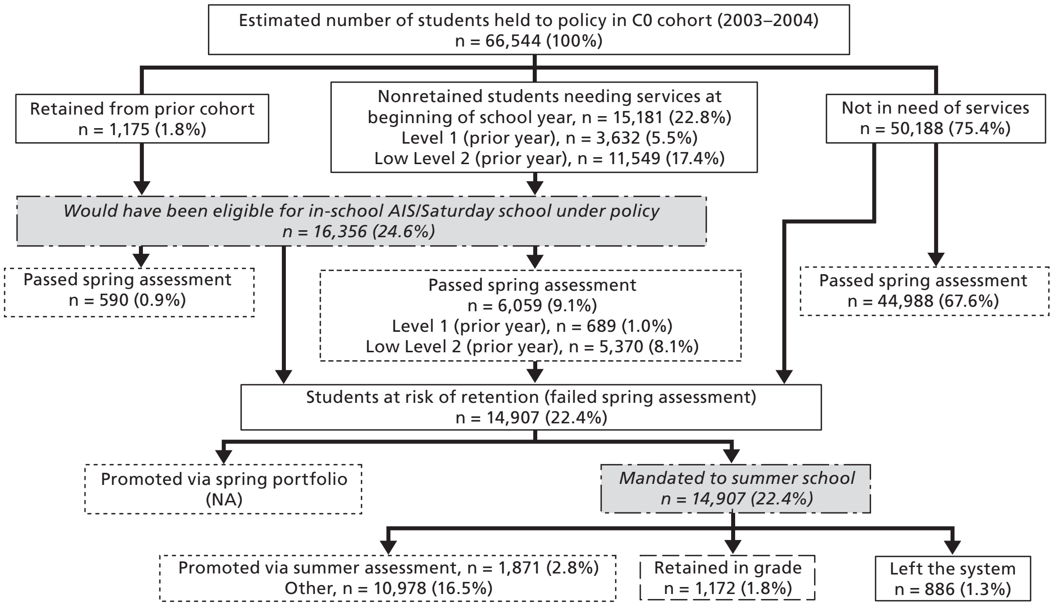
The next section provides a more detailed look at the progress of students in each cohort, with a focus on low-performing students.

### Snapshots of the 5th-Grade Cohorts

Figures 7.1 through 7.4 map the progress of students in each of our analysis cohorts from the beginning of the school year through the final 5th-grade promotion disposition. Each flowchart begins with the total number of students in that cohort who were held to policy (estimated in the case of the C0 cohort). We then show all other outcomes as a percentage of this number, allowing us to both track the cohort forward through the year and compare the relative outcomes of the four cohorts. The figures then categorize students by their "in need of services at the beginning of the school year" status. Such students are further disaggregated by their prior year performance—retained in 5th grade, scored Level 1 in ELA or mathematics on the 4th-grade assessment, or scored low Level 2 on both the ELA and mathematics 4th-grade assessment. Each group is tracked forward to determine the percentage passing the spring assessment, as measured by a score of Level 2 on both assessments. Those who failed to do so were at risk of being retained in grade. Some passed through a review of a portfolio of their work; all others were mandated to summer school. The figures then present the final promotion disposition—those promoted by different means (passing the summer assessment, review of a portfolio in August, or by appeal) or retained. These figures form the basis for our later discussion of future outcomes. To avoid confusion, students who leave the system are indicated separately in the figures because we did not have information on the future outcomes of these students. As shown in the figures, these students are a small percentage of each cohort.

**The 2003–2004 (C0) Cohort.** Figure 7.1 maps the progress of the C0 cohort. Of the 66,544 students we estimated would be held to the promotion policy in the C0 cohort, about 23 percent would have been considered in need of services at the beginning

**Figure 7.1**
**Students in the 2003–2004 (C0) Cohort Through the School Year: Need Status, Performance on Assessments, and Final Promotion Disposition**




RAND *MG894-7.1*

of the year based on their performance on the 4th-grade assessments—12 percent for ELA services, 4 percent for mathematics services, and 7 percent for both subjects (see Table C.2 in Appendix C). Overall, 19 percent of students needed ELA services and 11 percent needed mathematics services. Additionally, nearly 2 percent of the 5th-grade population was considered in need of services because they were repeating the 5th grade. Thus, one-quarter of all students would have been eligible for AIS and Saturday programs if the cohort had been subject to the promotion policy.

We found that about 78 percent of all students in the C0 cohort passed the spring assessments. Approximately 10 percent of students who would not have been designated as in-need failed the assessment, as did 60 percent of in-need students and half of retained students. The pass rate among those who had scored Level 1 on the 4th-grade assessment was dismal—19 percent.

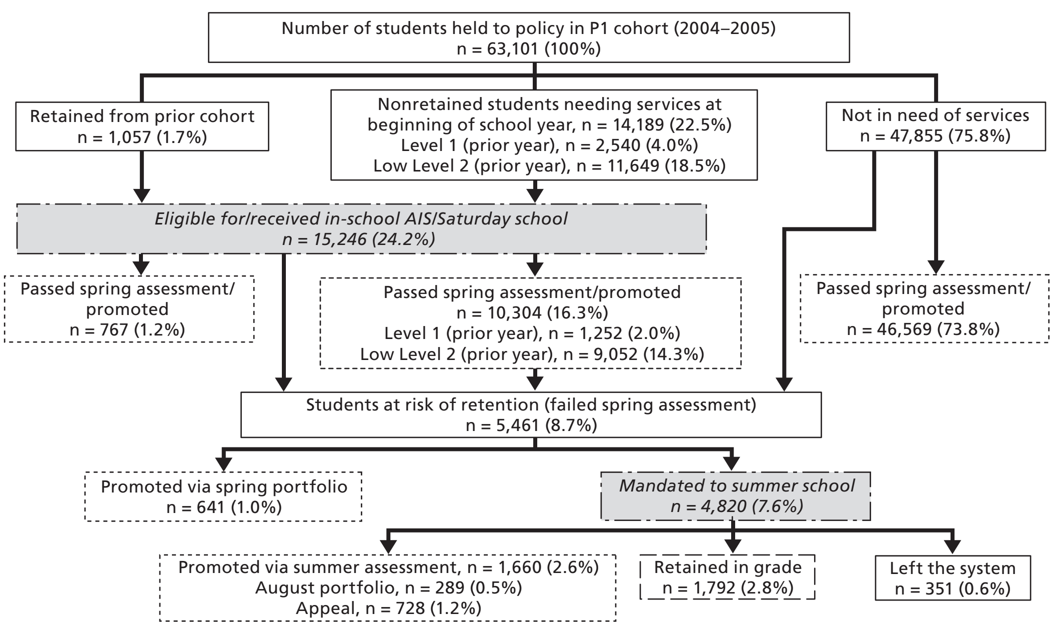
Of the 22 percent (n = ~15,000) who failed one or both spring assessments and were thus mandated to attend summer school, a considerable number of students (approximately 11,000, or 17 percent of the cohort) were promoted even though they had failed to pass the summer assessment. This represented almost three-quarters of students who would have been at risk of retention had they been subject to the promotion policy. About 3 percent of the cohort (13 percent of those at risk of retention) was promoted after passing the summer assessments. About 1,200 students, or 2 percent of the cohort, were retained in grade. This represented about 8 percent of all students who would have been at risk of being retained had the policy been in effect.

**The 2004–2005 (P1) Cohort.** As in the 2003–2004 cohort, of the 63,101 students held to the promotion policy in the 2004–2005 cohort, 23 percent were in need of services upon entering 5th grade, based on their 4th-grade assessment performance in at least one of the subjects (Figure 7.2).

Overall, 20 percent needed services in ELA and 18 percent in mathematics; 15 percent needed ELA services only, 3 percent needed mathematics services only, and 5 percent needed services in both subjects (see Table C.2 in Appendix C). An additional 2 percent of the 5th-grade population needed services because they were repeating the 5th grade. The number of retained students is smaller than indicated in the previous figure because—as we discuss later—some of the retained students were not held to the promotion policy in the retention year. Thus, 24 percent of all students in the first policy cohort were eligible to receive AIS and the Saturday sessions offered through the SPAs.

The pass rates on the spring assessments were considerably higher in the P1 cohort, and this was true across all groups. Overall, about 91 percent of the cohort passed the spring assessment. Among those needing services, 73 percent of retained students and 78 percent of those who had scored low Level 2 on the 4th-grade assessment passed the

**Figure 7.2**
**Students in the 2004–2005 (P1) Cohort Through the School Year: Need Status, Performance on Assessments, and Final Promotion Disposition**




RAND *MG894-7.2*

assessments. About half of those scoring Level 1 on the 4th-grade assessment passed—a marked improvement over the 19-percent pass rate among their counterparts in the C0 cohort.

About 9 percent of students (n = ~5,500) were at risk of retention under the policy; 1 percent of the cohort passed via the spring portfolio review, 3 percent passed the summer assessment, 0.5 percent passed the August portfolio review, and 1 percent appealed successfully. About 30 percent of students at risk of retention passed the summer assessment. Of the students at risk of retention, 13 percent were promoted via appeal, and a third were retained in grade. However, the number of students retained was low—1,792—and this accounted for fewer than 3 percent of students held to the policy. Students who were promoted via appeal or the August portfolio review tended to be older, and some had been retained in an earlier grade. For example, 12–13 percent of these students were 12 years or older in grade 5, compared to 7 percent of all students at risk of retention.

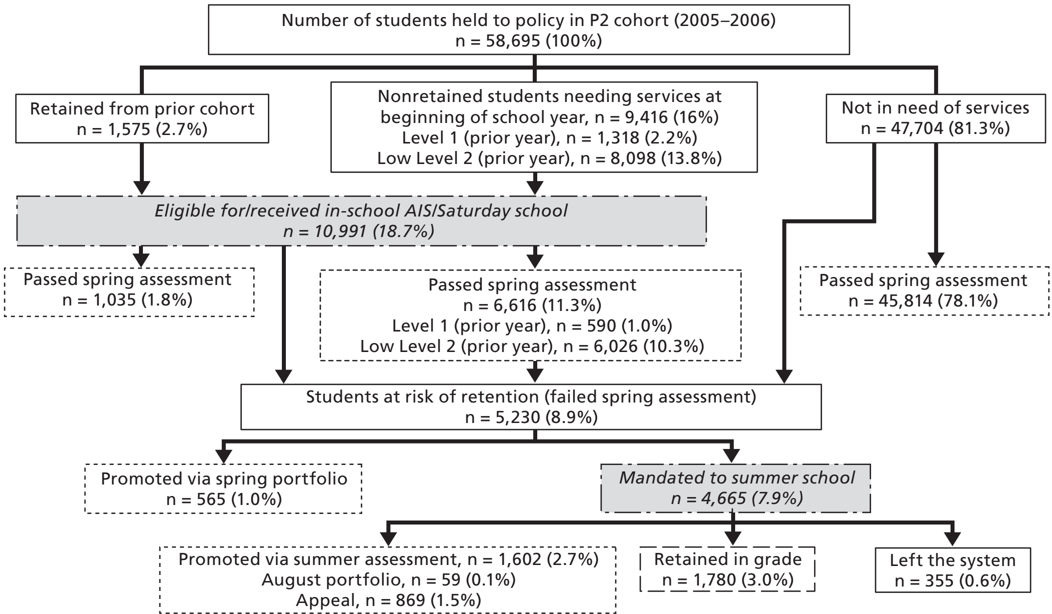
**The 2005–2006 (P2) Cohort.** In the 2005–2006 cohort—the first 5th-grade cohort to be subject to the 3rd-grade promotion policy—the percentage of students needing services at the beginning of the school year based on prior-year assessment scores dropped to 16 percent of the 58,695 held to the promotion policy (Figure 7.3). This may partly have been the result of the supportive services provided to low-performing 3rd graders under the promotion policy, and this trend of fewer in-need students continued in the subsequent cohort.

As in the P1 cohort, about 91 percent of P2 students passed the spring assessments. The pass rate among those who scored low Level 2 on the 4th-grade assessment was 74 percent, a little lower than the 78-percent pass rate of the P1 cohort. About two-thirds of retained students and half of those who scored Level 1 on the 4th-grade assessment passed the spring assessments, and these pass rates were comparable to those in the P1 cohort.

As in the P1 cohort, 9 percent of students were at risk of retention under the policy, and the patterns of promotion via different means were also very similar. About 1 percent of the cohort was promoted via the spring portfolio review, 3 percent by passing the summer assessment, and about 2 percent by appeal. Only a handful of students passed via the August portfolio review. The number and percentage of students retained in grade were also similar for the two cohorts (~1,700 students, or 3 percent of the cohort).

**The 2006–2007 (P3) Cohort.** Figure 7.4 maps the progress of students in the P3 cohort. This cohort was also subject to the 3rd-grade policy. The percentage needing services based on the 4th-grade assessments remained 16 percent for this cohort, similar to the percentage in the P2 cohort. However, there was a small uptick in the percentage of students who scored Level 1 on the prior-year assessment. For example, in this cohort, almost 7 percent of students had scored Level 1 on the 4th-grade assessment, compared to only 2 percent of the P2 cohort and 4 percent of the P1 cohort.

**Figure 7.3**
**Students in the 2005–2006 (P2) Cohort Through the School Year: Need Status, Performance on Assessments, and Final Promotion Disposition**




RAND *MG894-7.3*

The percentage needing services in ELA only dropped from 12 to 9 percent, while the percentage of students needing services in mathematics only or in both subjects rose slightly (see Table C.2). Combined with the percentage of retained students (3 percent of the cohort), the percentage needing services was close to 19 percent.

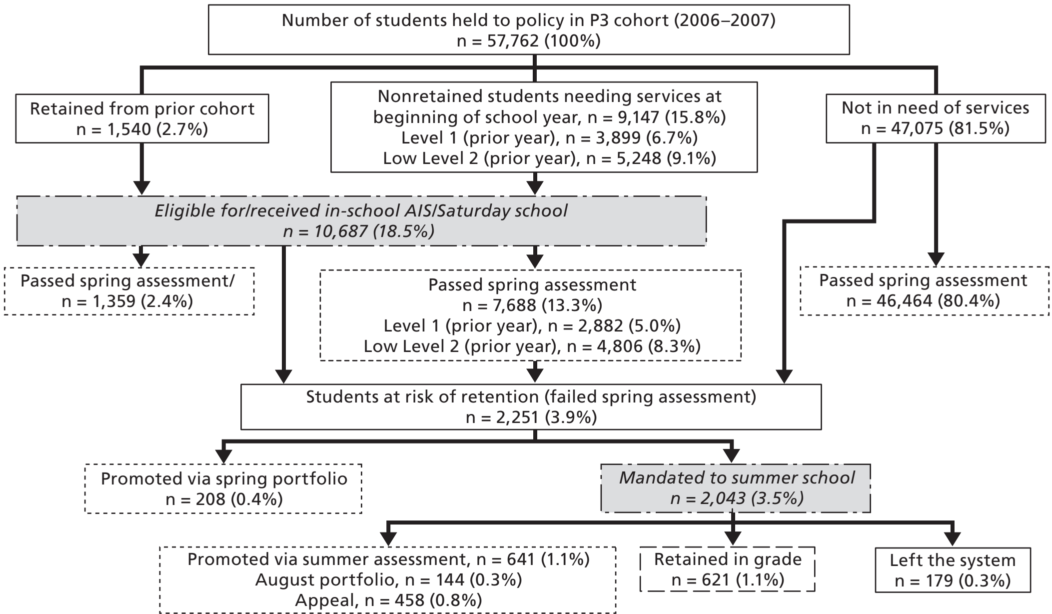
The pass rate on the spring assessment increased to 96 percent from 91 percent in the previous two cohorts. Low Level 2 students needing services increased their pass rates to 92 percent, considerably higher than the 74–78 percent in the previous two policy cohorts. The pass rates of retained students and those who scored Level 1 on the 4th-grade assessments improved substantially as well—to 88 percent and 74 percent, respectively.

About 2,200 students (4 percent of the cohort) were at risk of retention following the spring assessment. Less than 1 percent of the cohort was promoted via spring portfolio, the August portfolio, or by appeal, and about 1 percent was promoted by passing the summer assessment. Only 621 students (1 percent of the cohort) were retained in grade, accounting for 28 percent of the students at risk of retention.

## Profile of Retained Students

As shown earlier, between 2 and 3 percent of students were retained in grade, and this fell to 1 percent for the P3 cohort. Students at risk of being retained in both ELA and mathematics had the highest retention rates (60–61 percent in the first two policy

**Figure 7.4**
**Students in the 2006–2007 (P3) Cohort Through the School Year: Need Status, Performance on Assessments, and Final Promotion Disposition**




RAND *MG894-7.4*

cohorts and 46 percent in the P3 cohort; Figure 7.5). Students retained for performance in one subject were more likely to be retained due to low performance in mathematics than ELA.

Retained students tended to be disproportionately black and Hispanic (see Figure C.3), which is consistent with prior research on retention.[3] For example, Hispanic and black students made up 37 and 32 percent of the P3 cohort, respectively, but accounted for a disproportionate share of retained students (46–47 percent). Retained students also were disproportionately ELL students (see Figure C.4). For example, in the P3 cohort, while ELL students made up only 7 percent of the cohort, they accounted for 29 percent of retained students.

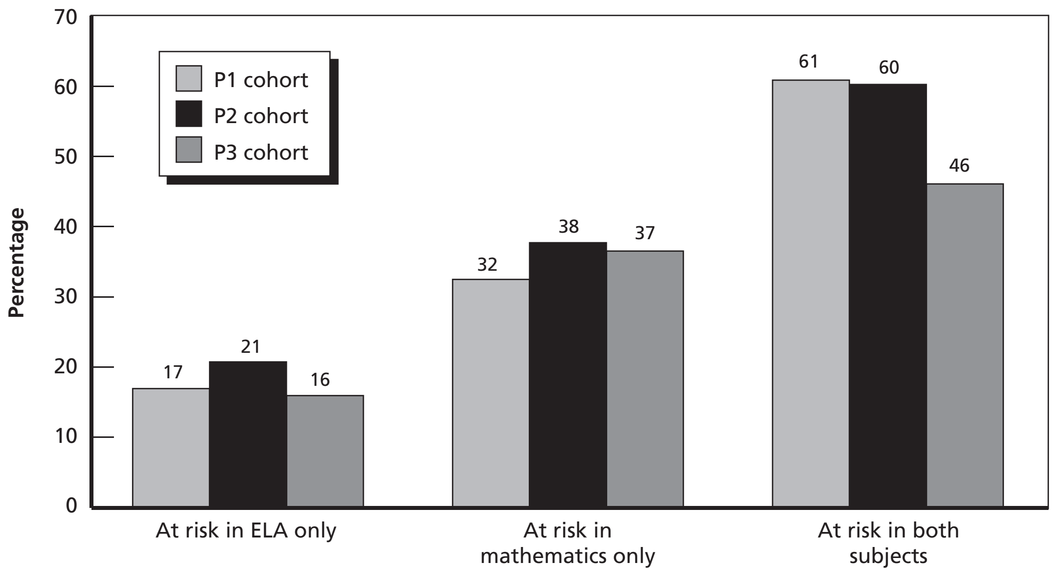
Unlike earlier studies, ours did not find that males tended to be disproportionately retained compared to females. Retained students tended to be absent from school more frequently.[4] In fact, between 46 and 54 percent of all retained students had 16 or more absences during the school year, and this was true across all four cohorts.

---

3 These differences by race/ethnicity and ELL status mirror differences among students who needed services at the beginning of the school year and those who were at risk of retention following the spring assessments (see Table C.3 and Figures C.1 and C.2 in Appendix C).

4 Students with more frequent absences in the previous school year were also overrepresented among those who needed services at the beginning of the school year and those who failed the spring assessments.

**Figure 7.5**
**Percentage of At-Risk Students Retained in Grade, by Subject and Policy Cohort**



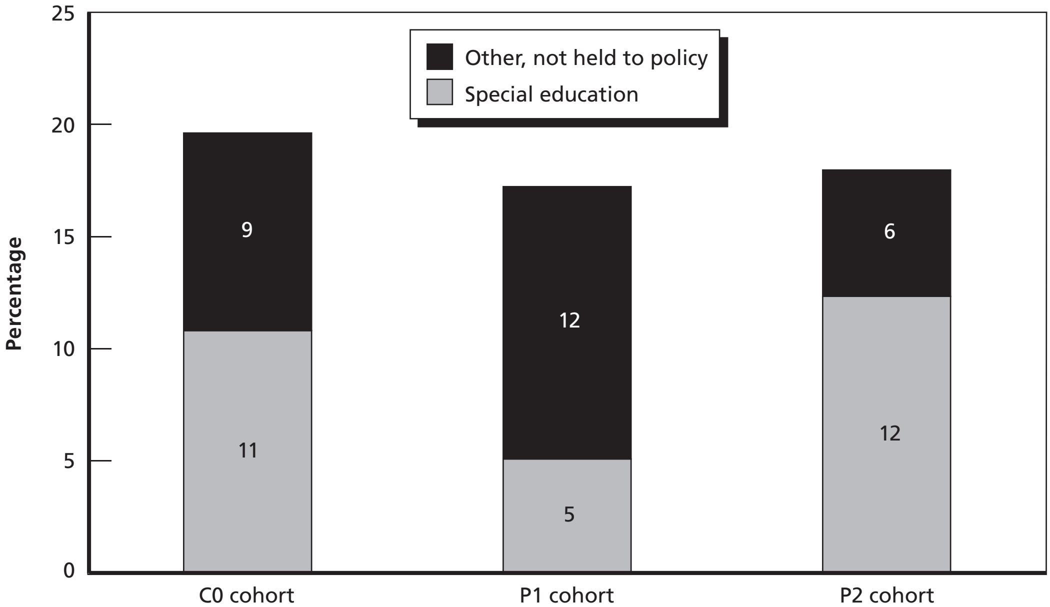

RAND *MG894-7.5*

A small proportion of retained students had already been retained in the past and so were multiple holdovers. We found that 7 percent of retained students in the P1 cohort, 11 percent in the P2 cohort, and 9 percent in the P3 cohort had previously been retained. It is likely that some of these multiple holdovers needed to be evaluated for referral to special education. An evaluation of the Chicago Public Schools promotion policy showed that, on average, 18 percent of 3rd and 6th graders retained in the 1997–1998, 1998–1999, and 1999–2000 school years were referred to special education within two years of the retention decision (Nagaoka and Roderick, 2004). We examined the held-to-policy status of the retained students during the retained year and found a similar trend, as shown in Figure 7.6.[5]

Across the three cohorts, between 17 and 20 percent of retained students were exempt from the policy the following (retention) year, and a number of these students were exempt on criteria other than those specified in the promotion policy regulations. Among retained students now exempt from the policy, 30 percent in the P1 cohort and 69 percent in the P2 cohort were newly categorized as special education students. A small percentage of students were classified as ELLs, although data indicated that they had been in ELL status for more than three years previously (and they had been held to policy the prior year). A handful of students (four students in the P1 cohort) were in

[5] We were unable to track held-to-policy status for about one-quarter of the 621 retained students in the P3 cohort—the remaining retained 469 students were held to policy the next year, and three students had special education status.

**Figure 7.6**
**Percentage of Retained Students Who Were Not Held to Policy in the Following Year, by Special Education Status and Cohort**




RAND *MG894-7.6*

District 84 (the charter school district) during the retention year and thus were exempt from the policy. The only consistent finding was that older students—most of whom were multiple holdovers—were more likely to be exempt from the policy during the retention year. For example, among retained students in the P1 cohort, 30 percent who were overage (12 years or older in grade 5) were exempt from the policy, compared to 19 percent of students who were younger.

### Performance of At-Risk Students in Later Years

The ultimate goal of the promotion policy is to help students at risk of academic failure become successful in later years by offering them a variety of supportive services during the policy year, several chances to demonstrate proficiency in ELA and mathematics, and retention in grade if they fail to meet standards so they can more fully master the promotion-gate grade material before moving on to more advanced material in subsequent grades. We now examine how retained students performed in future years. The following section discusses the future outcomes of at-risk promoted students compared to same-grade outcomes of retained students (who took the test a year later).

We previously discussed the issues that arise in comparing outcomes over time because of the switch from city- to state-administered assessments in the spring of 2006. Because of the different assessments used by the city and state, we cannot directly compare the level scores of students in two consecutive assessments if one was administered by the city and the other by the state. This restricts our ability to compare longitudinal outcomes over time for some cohorts in certain years.

### Performance of Retained Students in Later Years

**Performance on the Repeated 5th-Grade Assessments.** Table 7.4 shows how students retained for performance in a given subject performed after an additional year of instruction. It does not show outcomes for *all* retained students—only for the subset of students who failed to meet promotion standards in the subject the previous year and who remained in the NYC system. The majority of students were able to meet the promotion criteria a year later, and a significant number were able to improve their performance.

In the C0 cohort, about 16 percent of students who had been retained for ELA and 30 percent of those retained for mathematics failed to meet promotion criteria on the retention-year assessments. On the other hand, about one-quarter of those at risk in ELA and 10 percent of at risk in mathematics scored at the proficiency level in that subject.

**Table 7.4**
**Performance of Retained Students on the Repeated 5th-Grade Assessment, by Subject and Cohort**

| Cohort and Subject (Level 1 Students) | Performance Level on Repeated 5th-Grade Assessments (%) | | | Number of Students |
|---|---|---|---|---|
| | Level 1 | Level 2 | Level 3–4 | |
| C0 cohort (assessed in spring 2005)[a] | | | | |
| ELA | 16.1 | 59.1 | 24.8 | 763 |
| Mathematics | 30.0 | 59.6 | 10.4 | 1,030 |
| P1 cohort (assessed in spring 2006)[b] | | | | |
| ELA | 34.5 | 60.4 | 5.2 | 586 |
| Mathematics | 29.9 | 51.3 | 18.8 | 1,642 |
| P2 cohort (assessed in spring 2007) | | | | |
| ELA | 13.3 | 74.2 | 12.5 | 698 |
| Mathematics | 10.4 | 50.2 | 39.4 | 1,461 |
| P3 cohort (assessed in spring 2008) | | | | |
| ELA | 6.7 | 73.8 | 19.5 | 267 |
| Mathematics | 8.6 | 46.9 | 44.6 | 465 |

[a] Assessments were administered by the city.

[b] Assessment administration shifted from the city to the state. For this cohort, the original 5th-grade assessment was administered by the city, while the repeated 5th-grade assessment was administered by the state.

Recall that, for the P1 cohort, the first test was given by the city and the second by the state. We see a marked increase in the failure rate of this cohort in ELA as well as a marked decrease in the percentage of students scoring Level 3 or higher. Prior research has shown that when a new assessment is introduced to a district or state, scores initially drop and then steadily increase as teachers, administrators, and students gain familiarity with the content and format of the test (Linn, 2000; Koretz, 2002). However, in mathematics, the cohort had a rate similar to that of the previous cohort, although the percentage scoring at the proficiency level was higher (19 percent versus 10 percent). We cannot determine whether the differences are caused by actual differences in the performance of this cohort or the difference in the benchmarking procedure between the city and the state.

There was a marked improvement in the performance of retained students in the P2 cohort. For example, failure rates were between 10 and 13 percent in the two subjects, compared to 16–30 percent in the C0 cohort, when the city administered the two assessments. Thirteen percent of students scored Level 3 or higher in ELA, but proficiency rates in mathematics showed a dramatic increase, from 10 percent to 39 percent. The P3 cohort shows the same marked improvement in the performance of retained students—both in the percentages of students meeting the promotion standards and in the percentage of those who scored Level 3 or higher. For example, only 7–9 percent of retained students failed to meet the promotion standards on the repeated 5th-grade assessments, while 20 percent of students retained for ELA and close to 45 percent of those retained for mathematics increased their scores to Level 3 or higher.

**Performance on the 6th- and 7th-Grade Assessments.** Retained students from the P3 cohort were just finishing their repeated 5th-grade year in spring 2008, the last year for which we have data. The state took over testing of all grades in the spring of 2006, when retained students in the C0 cohort took the 6th-grade assessment. Thus, we can compare the relative performance of retained students in the C0, P1, and P2 cohorts on the 6th-grade assessments and the C0 and P1 cohorts on the 7th-grade assessments (Table 7.5).

The gains made by retained students during the retention year appear to have held as these students progressed to grades 6 and 7. For example, while 21 percent of retained students in the C0 cohort scored Level 1 on the 6th-grade ELA assessment, only 14 percent did so on the 7th-grade assessment. We see a decrease in failure rates in mathematics as well (from 34 percent to 18 percent) over the two grades. Compared to retained students in the C0 cohort, students from the two policy cohorts did much better in terms of both percentage passing and percentage scoring at the proficient level on the 6th- and 7th-grade assessments. For example, the percentage scoring Level 1 on the 6th-grade ELA assessment fell to 2–5 percent, while the percentage scoring Level 1 on the 6th-grade mathematics assessment fell to 24 percent among students in the P1 cohort and to 9 percent among students in the P2 cohort. Concomitantly, the percentage of retained students scoring Level 3 or higher increased over time, particularly in

**Table 7.5**
**Performance of Retained Students on the 6th- and 7th-Grade Assessments**

| Cohort and Subject | Performance on the 6th-Grade Assessments (%) | | | Number of Students | Performance on the 7th-Grade Assessments (%) | | | Number of Students |
|---|---|---|---|---|---|---|---|---|
| | Level 1 | Level 2 | Level 3–4 | | Level 1 | Level 2 | Level 3–4 | |
| C0 cohort | | | | | | | | |
| ELA | 21.3 | 70.2 | 8.5 | 624 | 13.6 | 79.4 | 7.0 | 573 |
| Mathematics | 34.4 | 51.0 | 14.6 | 837 | 18.4 | 67.7 | 13.9 | 762 |
| P1 cohort | | | | | | | | |
| ELA | 5.2 | 92.0 | 2.8 | 464 | 5.3 | 83.1 | 11.6 | 438 |
| Mathematics | 23.8 | 53.9 | 22.4 | 1,318 | 9.7 | 57.5 | 32.8 | 1,233 |
| P2 cohort | | | | | | | | |
| ELA | 2.4 | 85.6 | 12.0 | 618 | NA | NA | NA | NA |
| Mathematics | 9.0 | 48.5 | 42.5 | 1,269 | NA | NA | NA | NA |

NOTE: Shading indicates a shift in administration from city to state. For this cohort, the repeated 5th-grade assessment was administered by the city, while the 6th-grade assessment was administered by the state.

mathematics. More than 40 percent of retained students in the P2 cohort scored at the proficient level on the 6th-grade mathematics assessment.

We also found improvements in performance on the 7th-grade assessments. The percentage scoring Level 1 was 14–18 percent in the C0 cohort and 5–10 percent in the P1 cohort—considerably lower than the percentages scoring Level 1 on the 6th-grade assessments. Students in the P1 cohort also improved their proficiency rates on the 7th-grade assessments, with 12 percent scoring Level 3 or higher in ELA and 33 percent doing so in mathematics, compared to 3 percent and 22 percent, respectively, in the 6th grade. One point to note is that the 7th-grade promotion policy took effect in the spring of 2006, so some of these students may have been targeted for additional services during the 7th-grade year. Also, the general improvement in student performance over time reflects trends throughout the city and state (detailed later in this chapter).

It is notable, though, that students retained for mathematics were far more likely to achieve proficiency in that subject in future grades than were students retained for ELA. For instance, in the P1 cohort, 22 percent of students retained for mathematics reached proficiency in grade 6, and 33 percent did so in grade 7. Only 3 percent of students retained for ELA reached proficiency on the 6th-grade ELA assessment, and only 12 percent did so in grade 7.

### Performance of At-Risk Promoted Students in Later Years

The previous section examined how retained students performed over time. In this section, we examine how at-risk students who were promoted via different means (spring portfolio review, summer assessments, or appeal) performed in later grades. Students can also be promoted by portfolio review in August; however, this group was very small, so we focused on the other means of promotion. For ease of comparison, we also show same-grade trends for retained students, although it should be kept in mind that these students take the test a year later than the other student groups in their cohort.

We focus on two metrics: (1) the percentage of students in these various groups who scored Level 1 on the spring assessments in subsequent grades and (2) the percentage of students who scored Level 3 or higher—the metric used to describe proficiency for the purposes of NCLB.

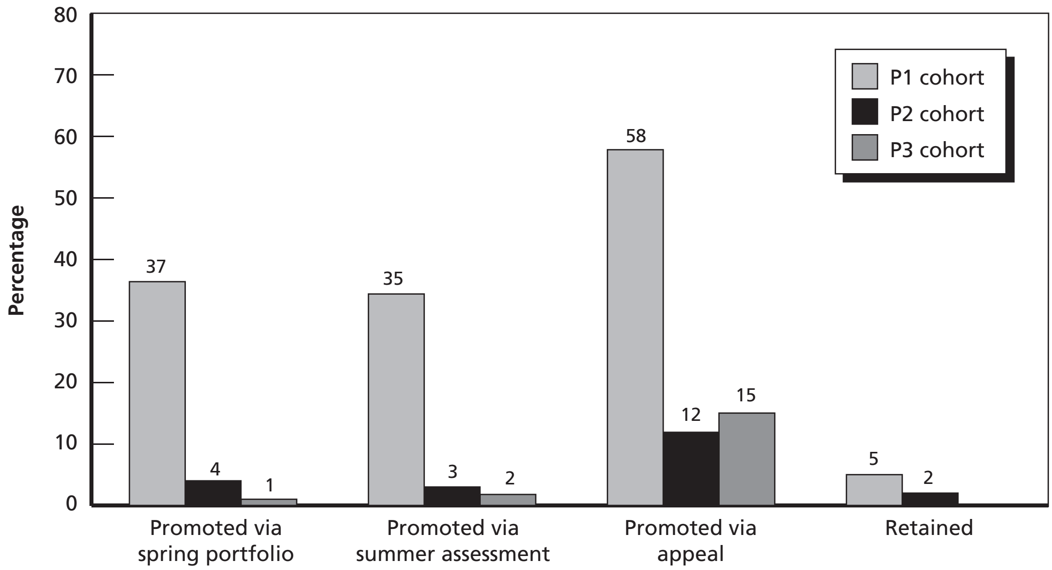
**Performance on the 6th-Grade Assessments.** For 6th-grade outcomes, we focus on the three policy cohorts because students in the C0 cohort who were promoted took a city-administered test in spring 2005, while students in other cohorts took the state test.

The percentages of students who scored Level 1 on the 6th grade in ELA fell dramatically across the three cohorts for all groups of students (Figure 7.7). For example, while 35–37 percent of students in the P1 cohort who were promoted by assessment or the spring portfolio scored Level 1 on the 6th-grade spring ELA assessment, only 1–2 percent of their counterparts in the P3 cohort did so. Students promoted via appeal tended to have poorer outcomes than other promoted or retained students. Although they experienced a marked decline in failure rates on the 6th-grade assessment, about 12–15 percent in the later cohorts scored Level 1. Retained students took the test a year later, and only 2 and 5 percent, respectively, failed to score Level 1 on the 6th-grade ELA assessment in the two cohorts for which we have data.

The percentages of at-risk promoted students scoring at the proficiency level on the 6th-grade ELA assessment remained steady (3–5 percent) over time (see Figure C.5). However, retained students increased their proficiency rates from 3 percent to 12 percent across the two cohorts—again, they took the test a year later.

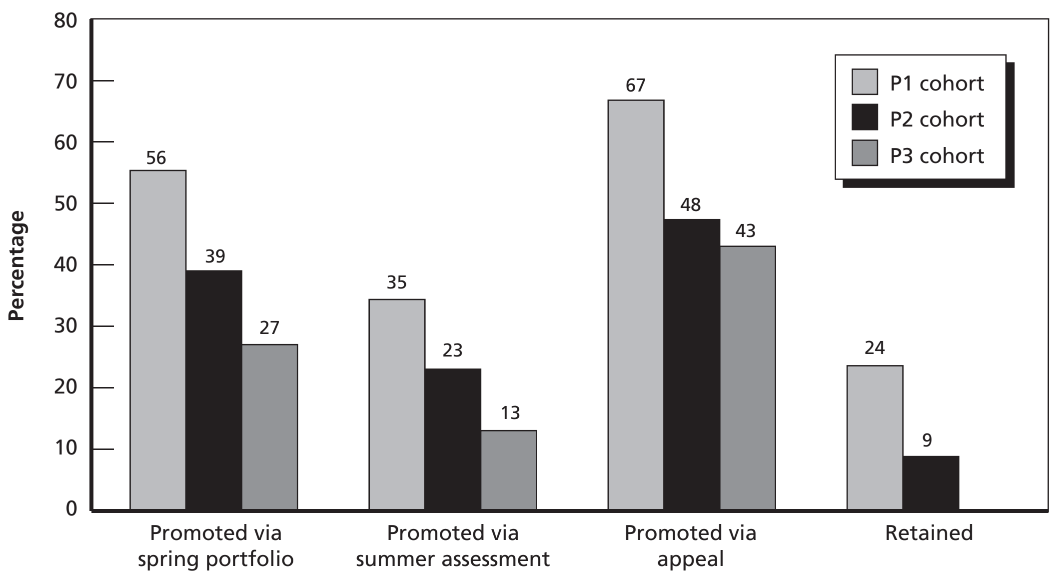
Figure 7.8 shows improvement for mathematics outcomes across the cohorts. We see the same marked decline in the percentage of at-risk promoted or retained students scoring Level 1 on the 6th-grade mathematics assessment as we had seen in ELA, although overall failure rates were higher. Thus, while 56 percent of students in the P1 cohort who were promoted via the spring portfolio scored Level 1 on the 6th-grade mathematics assessment, this was true of only 27 percent of those in the P3 cohort who were thus promoted. Again, we see that students promoted via appeal—although they also showed marked improvements across the three cohorts—do worse than other at-risk students. For instance, among students in the P3 cohort who were promoted via appeal, 43 percent scored Level 1 on the 6th-grade mathematics assessment.

**Figure 7.7**
**Percentage of At-Risk Promoted and Retained Students Scoring Level 1 on the 6th-Grade ELA Assessment**




RAND *MG894-7.7*

**Figure 7.8**
**Percentage of At-Risk Promoted and Retained Students Scoring Level 1 on the 6th-Grade Mathematics Assessment**




RAND *MG894-7.8*

Retained students taking the test after an additional year of schooling outperformed their at-risk and promoted peers.

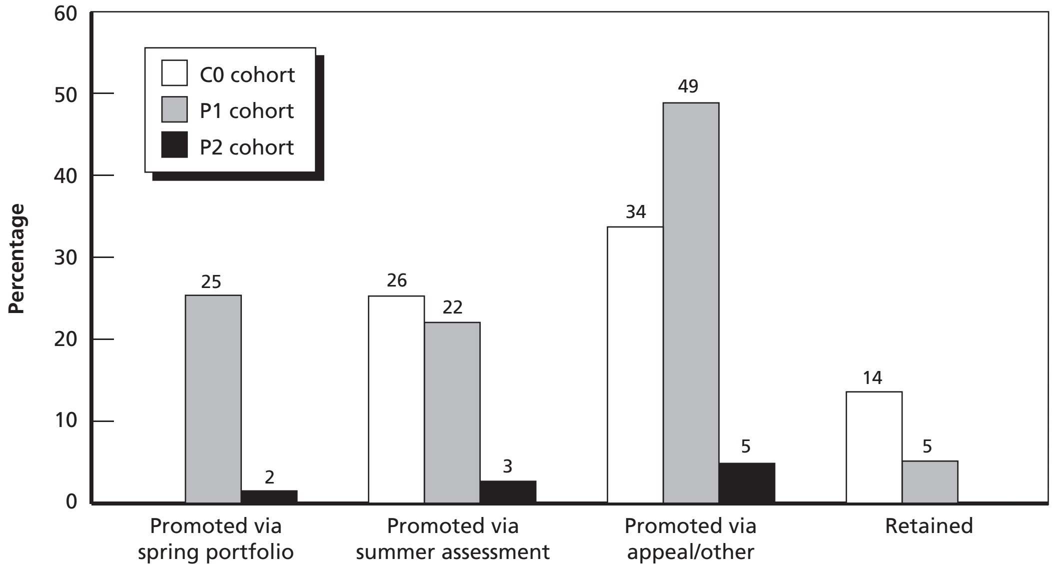
In terms of mathematics proficiency rates, all groups posted marked gains over time, particularly retained students and those promoted by summer assessment (see Figure C.6). The proficiency rates of retained students almost doubled over the two policy cohorts (22 percent to 43 percent); the percentage proficient among students promoted by summer assessments increased from 11 percent in the P1 cohort to 23 percent in the P3 cohort.

**Performance on the 7th-Grade Assessments.** For 7th-grade outcomes, we were able to examine results for students in the C0 cohort, who took the 7th-grade assessments in either spring 2006 (promoted students) or spring 2007 (retained students), in the absence of later grade retention. Both were state-administered tests. However, because that cohort was not subject to the promotion policy, there were no students who were promoted via the spring portfolio. We can also track outcomes for the P1 and P2 cohorts, though not for retained students in the latter cohort.

Figures 7.9 and 7.10 present 7th-grade ELA and mathematics outcomes for the different groups of at-risk promoted and retained students in these three cohorts—those scoring Level 1. As with the 6th-grade outcomes, consistent with city and state trends presented later in this chapter, we see marked improvements across the three cohorts and across all groups in terms of both the percentages scoring Level 1 and ELA proficiency rates. For example, while 26 percent of the students in the C0 cohort who were promoted by summer assessment scored Level 1 on the 7th-grade ELA assessment, this was true of only 3 percent of their counterparts in the P2 cohort. Retained students also showed improvement over the two years. Overall, the number scoring Level 1 on the 7th-grade ELA assessment in the P2 cohort was 5 percent or lower for all these groups of students. Proficiency rates increased from 3–4 percent to 11–17 percent among promoted students and from 7 percent to 11 percent among retained students (see Figure C.7).

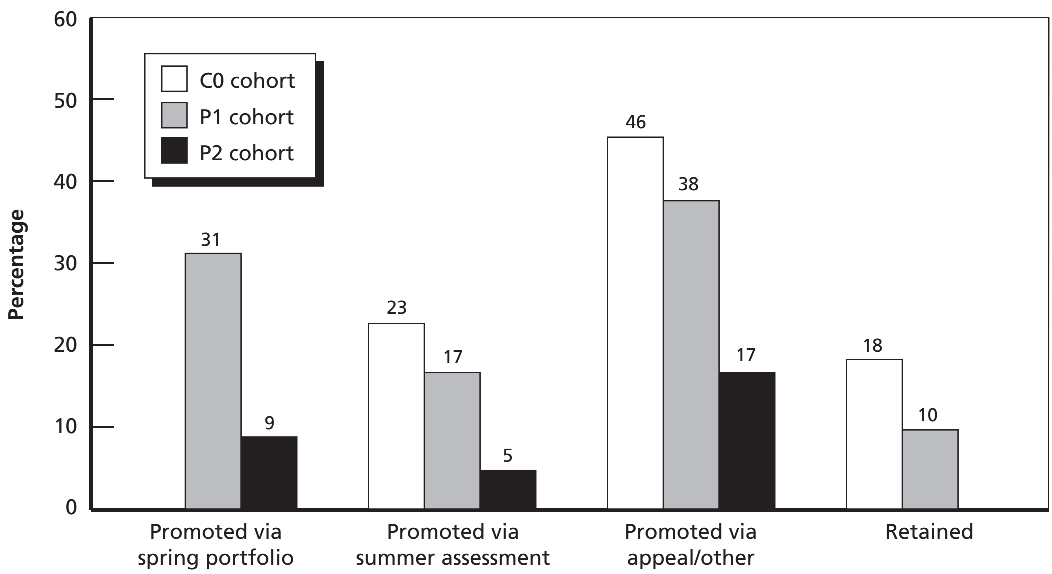
The percentages scoring Level 1 in mathematics declined markedly across the three cohorts for all groups of students (Figure 7.10). While those promoted on appeal continued to have higher failure rates than other at-risk students, there was also a sharp decline in the percentages scoring Level 1—from 46 percent to 17 percent. Retained students had the lowest failure rates on the 7th-grade mathematics assessment. Proficiency rates showed marked gains across the three cohorts for all groups (see Figure C.8). For example, proficiency rates among those promoted by meeting standards on the summer assessments rose from 9 percent for the 2003–2004 cohort to 33 percent for the 2005–2006 cohort. Promoted at-risk students—like retained students—were more likely to reach proficiency in mathematics than ELA. Again, retained students, who took the assessment a year later than their promoted peers, had the higher proficiency rates for the two years for which we have data.

**Figure 7.9**
**Percentage of At-Risk Promoted and Retained Students Scoring Level 1 on the 7th-Grade ELA Assessment**




RAND *MG894-7.9*

**Figure 7.10**
**Percentage of At-Risk Promoted and Retained Students Scoring Level 1 on the 7th-Grade Mathematics Assessment**




RAND *MG894-7.10*

One possible reason for the continued improvement over time is that the promotion policy was extended to the 7th grade in the spring of 2006. As a result, low-performing students were likely to have been identified for services at the beginning of the school year—including those retained in prior years--and many of them may have received another year of supportive services under the 7th-grade policy. However, we find similar improvements for all students across the city and state (described in the next section), suggesting that factors other than the promotion policy played a role in improving the performance of students across the state.

## Achievement Trends in Grades 3–8: New York City and the Rest of New York State

This section examines general achievement trends in NYC and the rest of the state (excluding NYC) between 2005–2006 and 2007–2008 in ELA and mathematics on the state assessment.[6] In general, student performance in NYC and the rest of the state increased over time, with the greatest increases occurring between 2006–2007 and 2007–2008. The performance of NYC students was lower than that in the rest of the state in all grades but generally followed the same trajectory of improvement.

### Performance on the Spring ELA Assessments, Grades 3–8

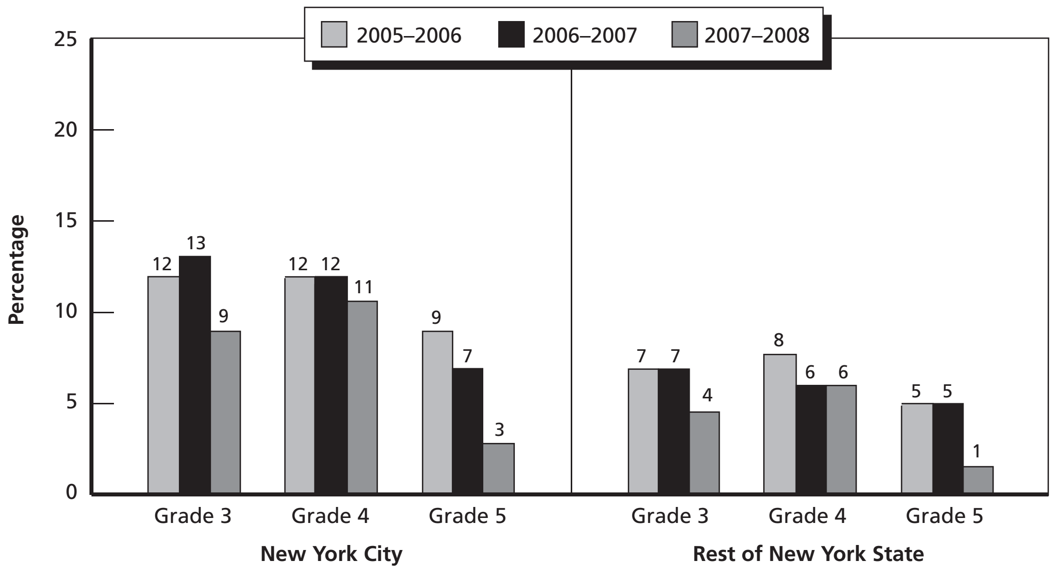
The percentage of students scoring Level 1 decreased in all grades (Figures 7.11 and 7.12):

- By 2007–2008, the percentage of students in grades 5, 6, and 7 scoring Level 1 in ELA was only 3 percent in NYC and 1 percent in the rest of the state.
- That year, the percentages of Level 1 students in grades 3, 4, and 8 were higher—8–11 percent in NYC and 4–6 percent in the rest of the state.

Concomitantly, there were large increases in the percentage of students who scored at or above the proficient level (Level 3) on the ELA assessment between 2005–2006 and 2007–2008 in grades 5, 7, and 8 (see Figures C.9 and C.10). For example, proficiency rates in grade 5 increased 12 percentage points in NYC and 9 percentage points in the rest of the state over this period, while proficiency rates among 7th graders increased 16 percentage points in NYC and 12 percentage points in the rest of the state.

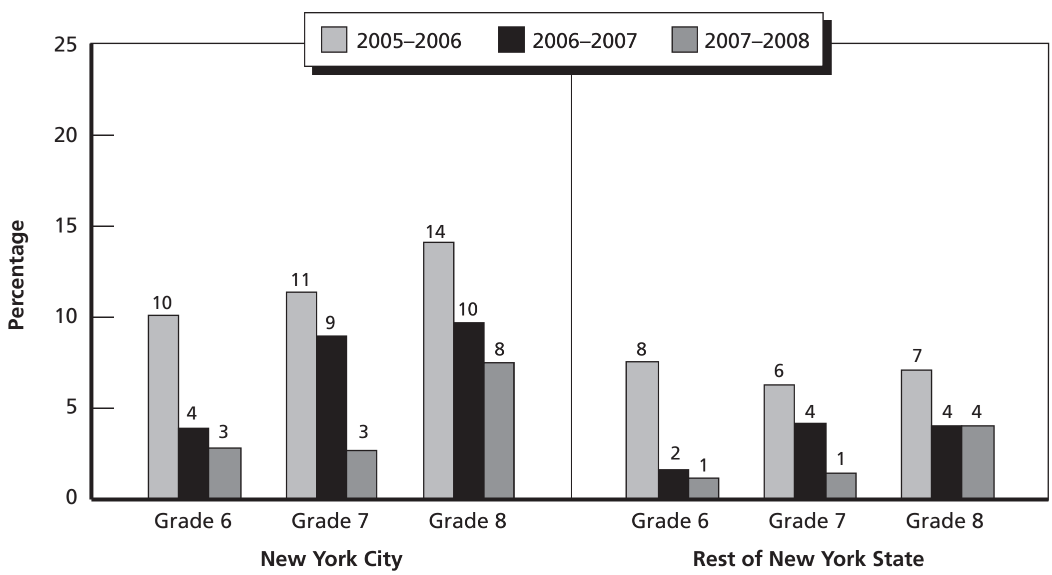
[6] Data files providing results on achievement tests by county were obtained from the NYSED Web site (NYSED, 2009). Data were available by county, thus allowing us to calculate the averages for the rest of the state by excluding the five counties that are coextensive with the boroughs of NYC—the Bronx (Bronx County), Brooklyn (Kings County), Manhattan (Manhattan or New York County), Queens (Queens County), and Staten Island (Richmond County).

**Figure 7.11**
**Percentage of Students Scoring Level 1 on the ELA Spring Assessments, Grades 3–5, NYC and Rest of State**




RAND *MG894-7.11*

**Figure 7.12**
**Percentage of Students Scoring Level 1 on the Spring ELA Assessments, Grades 6–8, NYC and Rest of State**




RAND *MG894-7.12*

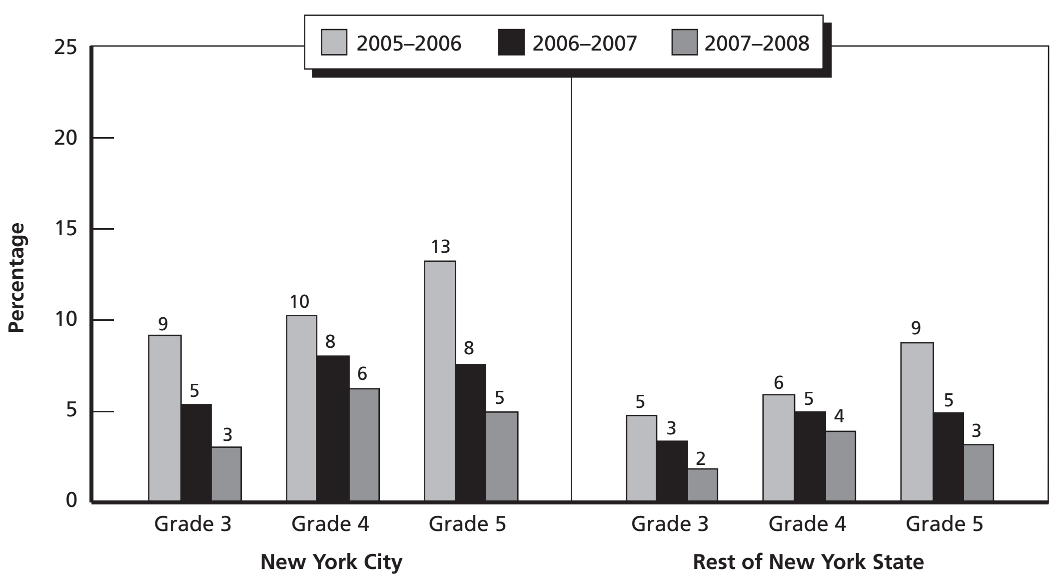
### Performance on the Spring Mathematics Assessments, Grades 3–8

NYC mathematics trends—similar to the ELA trends—tended to follow those in the rest of the state. The percentage of students scoring Level 1 on the mathematics assessment was higher in NYC than in the rest of the state, but this decreased difference over time in all grades (Figures 7.13 and 7.14). For example, in grades 7 and 8, between 2005–2006 and 2007–2008, the percentage of students scoring Level 1 on the mathematics assessment fell 13–14 percentage points.

Proficiency rates rose dramatically in all grades (see Figures C.11 and C.12), as the following three examples illustrate:

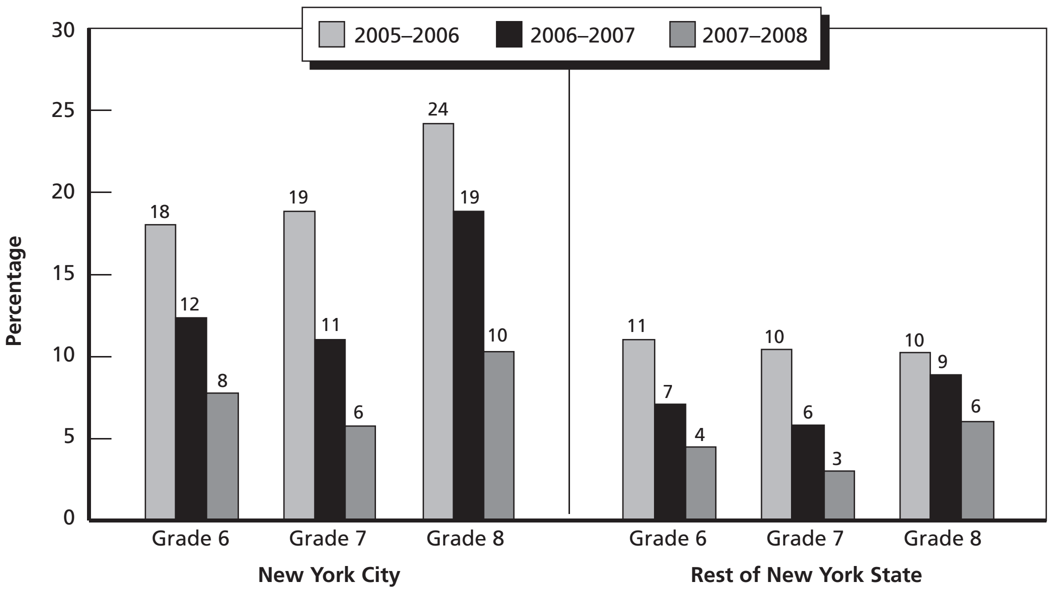
- Between 2005–2006 and 2007–2008, the percentage of students performing at the proficient level increased 9 percentage points in grade 4 (the lowest percentage-point gain across grades) and 25 percentage points in grade 7 (the highest percentage-point gain across grades).
- Trends in the rest of New York State were quite similar, and the percent proficient in grade 7 mathematics rose 22 percentage points between 2005–2006 and 2007–2008.
- Students in the lower grades had higher proficiency rates than those in higher grades. In 2007–2008, 87 percent of 3rd-grade students scored at the proficient level in mathematics, while 60 percent of 8th-grade students did so.

**Figure 7.13**
**Percentage of Students Scoring Level 1 on the Spring Mathematics Assessments, Grades 3–5, NYC and Rest of State**




RAND *MG894-7.13*

**Figure 7.14**
**Percentage of Students Scoring Level 1 on the Spring Mathematics Assessments, Grades 6–8, NYC and Rest of State**




RAND *MG894-7.14*

## Discussion

These results are impressive but raise several questions. The first concerns the rigor of the state test. To counter skepticism, NYSED states that "New York's testing system passed rigorous review" by the U.S. Department of Education; that a group of "national testing experts" comprehensively reviews the program; and that it "conducts expert independent analysis of vendor work" (NYSED, 2008, slide 25).

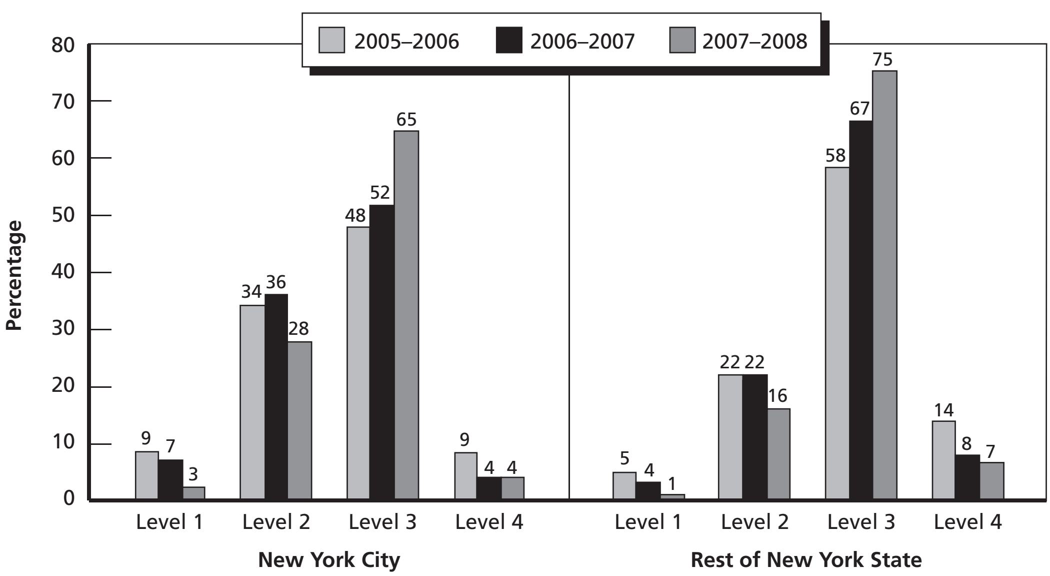
Analyzing the rigor of the test was far beyond the scope of this study, and we take the test at face value because it is the metric used by the state to measure the student progress for NCLB purposes.

The second question is what these results imply for our ability to attribute gains in NYC to the promotion policy. Earlier, we showed the marked improvements in NYC student outcomes, particularly among at-risk students in our 5th-grade cohorts. On the surface, one would tend to attribute these improvements to the promotion policy and the supports that were put in place for in-need students. However, it is clear that the storyline is much more complex. The policy was only one of a set of reforms that NYC put into place against the backdrop of the reforms enacted by the state. Since the state took over testing in 2006, there has been an across-the-board increase in the performance of students in both subjects and in almost every grade, and some of these increases have been quite substantial (Figures 7.15 and 7.16). As a result, it is difficult to attribute improvements in NYC solely to the promotion policy. On the one hand, the fact that NYC showed larger improvements may reflect the effect of the targeted services provided under the promotion policy; improvements in grades that were not subject to the promotion policy may be the result of spillover effects from the schools' increased focus on low-performing students, as a result of the policy and the pressures of NCLB and the persistence of the promotion policy's effects in later grades.

However, other mechanisms may be partly responsible for the observed improvement as well. For example, prior research has shown that large increases in test scores are associated with score inflation resulting from teachers becoming more familiar with the test and focusing on preparing students by teaching to the test (Koretz, 2008; Koretz, McCaffrey, and Hamilton, 2001; Linn, 2000). As noted earlier in Chapter Five, test-taking strategies were being emphasized in both in-school and out-of-school programs. If the tests and the standards and curriculum to which they are aligned are rigorous, there may be nothing adverse about the strategy and the results. It is beyond the mandate of this study to attribute the increasing trends in city and state results to any particular source or to evaluate the rigor of the state test.

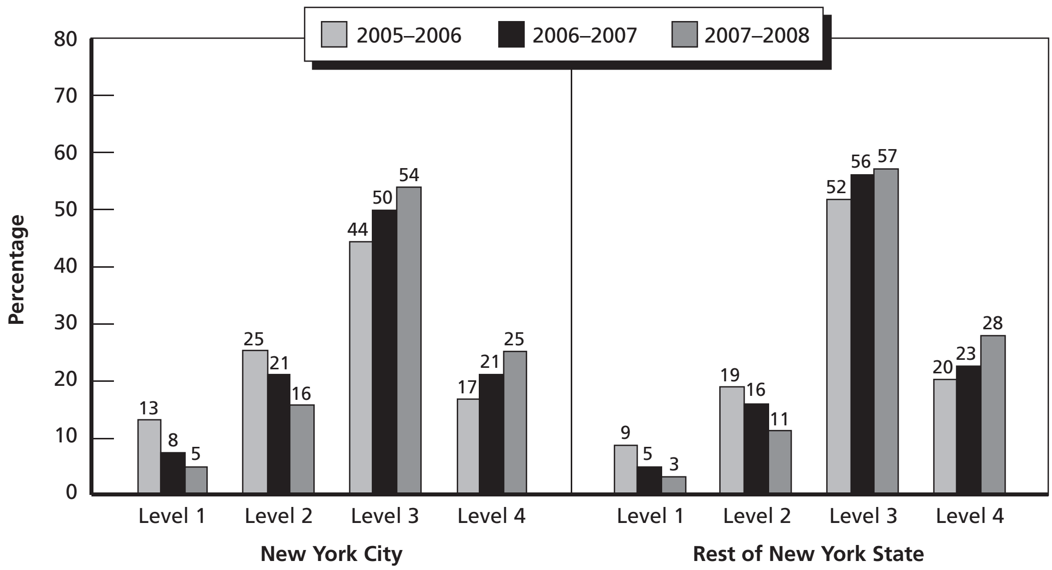
Traditionally, researchers have sought to validate rising test scores in a state against the results reported by the National Assessment of Educational Progress (NAEP). We examined the NAEP trends for both the state and the city to see whether NAEP

**Figure 7.15**
**Performance of 5th-Grade Students, by Level Achieved on the Spring ELA Assessment and Year, NYC and Rest of State**




RAND *MG894-7.15*

**Figure 7.16**
**Performance of 5th-Grade Students, by Level Achieved on the Spring Mathematics Assessment and Year, NYC and Rest of State**




RAND *MG894-7.16*

trends mirrored the observed improvements. Unfortunately, the period for which the NAEP data were available did not sufficiently coincide with our state data, as the last published NAEP data points were for 2005 and 2007, and NAEP scores are available only for grades 4 and 8. The most dramatic increases in state test scores occurred between 2007 and 2008. However, NAEP results show that in mathematics, average 4th-grade scale scores in 2007 posted an increase over scores in 2005 (243 versus 238), and the proficiency rate increased from 36 percent in 2005 to 43 percent in 2007. In grade 8, the scale scores showed no increase in mathematics since 2003. Average reading scores have remained essentially unchanged in grade 4 since 2002 and in grade 8 since 1998; neither have proficiency rates changed over the same period.

Results from NYC on the Trial Urban District Assessment NAEP tell a similar story. Fourth graders posted increases in mathematics between 2005 and 2007—moving from a 25-percent proficiency rate to a 34-percent proficiency rate. However, there were no significant differences in the mathematics performance of 8th-grade students in mathematics. There were no significant gains in reading in grade 4 or 8 between 2005 and 2007. However, the average scale score for 4th-grade students in 2007 was higher than in 2003 (270 versus 266), though this was not the case for grade 8.[7] For these two grades, we fail to see a concomitant increase in NAEP scores,

---

7 NAEP results can be found in National Center for Education Statistics (undated).

but it must be acknowledged that there are differences in what the NAEP and state tests are trying to measure, and the proficiency thresholds are defined differently.

## Summary

We used four 5th-grade cohorts (a comparison cohort and three policy cohorts) to track performance trends over time.

### Students Held to the Promotion Policy

Overall, approximately 60,000 5th-grade students were held to the policy in each of the three cohorts subject to the promotion policy, accounting for 75–78 percent of the cohort. Exemptions for special education and ELL students did not fully account for all students not held to the policy.

### Students Who Needed Services at the Beginning of the School Year

Students who scored Level 1 or low Level 2 on the 4th-grade assessments and retained students were categorized as needing services at the beginning of the school year—between 16 and 23 percent of 5th graders. The decrease in the percentage needing services between the P1 and P2 cohorts (from 23 percent to 16 percent) can be partly attributed to the additional services that low-performing students in the latter cohort may have received under the 3rd-grade promotion policy. This applied to the P3 cohort as well.

### Student Performance on the Proximal-Year Spring Assessments

The overall performance of the 5th-grade cohorts improved markedly over time. This mirrors trends in the rest of the state in grades 3–8 in both subjects from spring 2006 to spring 2008.

Students who were low-performing at the beginning of the school year mirrored this general improvement. For example, the percentage of students needing services in ELA who scored Level 1 on the ELA spring assessment declined from 36 percent in the C0 cohort to 9 percent in the P1 cohort. This sharp decline in the failure rate (both cohorts took the city assessment) was at least partly driven by the promotion policy and the supportive services provided to low-performing students during the school year.

### Fifth-Grade Promotion/Retention Outcomes for Students at Risk of Retention

About 30 percent of at-risk students passed the summer assessments and were promoted. There was a marked decrease in the percentage of students who failed the summer assessments but were promoted—from 17 percent in the C0 cohort to about 1–2 percent in the three policy cohorts.

About 2–3 percent of students in each cohort were retained in grade. By the P3 cohort, that percentage had fallen to 1 percent of the cohort (about 600 students out of approximately 58,000 students). Generally, retained students accounted for 30 percent of all students at risk of retention, as determined by the spring assessments.

### Retained Students

Hispanic and black students accounted for a disproportionate share of retained students, as did ELL students. Retained students tended to be absent from school more frequently and were more likely to be multiple holdovers.

A little less than 20 percent of retained students were exempt from the policy in the subsequent (retention) year. While many were special education referrals, some retained students were exempted from the policy using criteria other than those in the policy regulations.

The performance of retained students on the repeated 5th-grade assessment improved substantially from spring 2006, in terms of both pass rates and proficiency levels (particularly in mathematics).

### Performance of At-Risk Students on Higher-Grade Spring Assessments

Over time, proficiency rates in grades 6 and 7 for all at-risk student groups increased over time, while the percentage of students scoring Level 1 in these grades declined. Students at risk in mathematics were far more likely to achieve proficiency in that subject in future grades than were students at risk in ELA. In general, retained students performed better than promoted students in their cohort on the same-grade assessment (though they took these assessments in different years). Students promoted via appeals tended to do more poorly than their cohort peers. These trends were consistent in ELA and mathematics.

The improvements in 7th grade may have been at least partly due to the supportive services that low-performing students received under the 7th-grade promotion policy implemented in the spring of 2006. We see similar improvements in the rest of the state, however.

### Comparing Performance Trends in NYC and the Rest of the State

The state took over testing of all grades in spring 2006 and instituted a whole set of reforms aimed at improving student performance. Between 2006 and 2008, both NYC and the rest of the state saw marked improvements in the percentage of students scoring Level 1 on the spring assessments in almost every grade and in both subjects as well as in the percentage scoring Level 3 or higher. While, generally, NYC students lagged behind the rest of the state with respect to proficiency levels, they posted equal or larger increases in proficiency rates over this three-year period and equal or larger declines in rates of failing to pass the spring assessment.

CHAPTER EIGHT

# Measuring the Effect of Supportive Interventions on Proximal-Year Student Achievement

*Louis T. Mariano, Sheila Nataraj Kirby, Al Crego, and Claude Messan Setodji*

In this chapter, we examine the proximal-year academic outcomes of 5th graders treated under the promotion policy, i.e., the spring assessment outcomes of students identified as in need of services at the beginning of the year and the summer assessment and retention outcomes of students identified as at risk based on spring performance. The next chapter examines future outcomes of treated 5th-grade students. Appendix E presents the results of similar analyses of 3rd graders. Table 8.1 presents the research questions addressed in this chapter and maps them to the methods we use to answer them.

Two natural comparison groups are available for use in quasi-experimental designs to examine the academic outcomes of students treated under the promotion policy. In this chapter, as well as in Chapter Nine, we use students in the 5th-grade cohort in the year prior to the inception of the promotion policy as a control cohort to predict how students treated under the policy may have performed in the absence of the policy. To predict the performance of treated students under the control cohort and estimate

**Table 8.1**
**Proximal-Year Outcomes: Research Questions and Methods**

| Research Question | Method |
|---|---|
| Students needing services at the beginning of the year based on prior year's scores, relative to a comparison group | |
| How did these students perform on the spring assessments? | Propensity score weighting and doubly robust regression |
| What was the relationship between SPA attendance and performance on the spring assessments? | GAMM |
| Students at risk of retention based on performance on proximal-year spring assessments, relative to a comparison group | |
| How did these students perform on the summer assessments? | Propensity score weighting and doubly robust regression |
| What was the relationship between SSA attendance and performance on the summer assessments? | GAMM |

the treatment effect on the treated students, we use propensity score–weighted doubly robust regression models. When we examine future outcomes in Chapter Nine, we also use students in the policy year who were just above the treatment thresholds (i.e., students who were not at risk but very close to the cutoff) as a natural control group for students who were at risk and just below treatment thresholds, and we estimate the treatment effect on the treated students near the threshold using RDD. As discussed in Chapter Four, the RDD approach allows for stronger causal inference than the propensity-weighted regression model because it can account for unobserved confounders.

In both this chapter and the next, for each method, we display the estimated treatment effect on various groups of treated students, along with the 95-percent confidence interval for the estimate, in the tables and figures. Effect sizes are presented on the scale of the standard deviation of the outcome measure. Popular guidelines for interpreting effect sizes are those set forth by Cohen (1988), in which effect sizes of about 0.2 are considered "small," effect sizes of about 0.5 are considered "medium," and effect sizes of about 0.8 are considered "large." In addition, where possible, we provide additional context for the effect sizes by explaining what they imply for the relative position of the student in Performance Level 2. For example, suppose that Level 2 covers a range of 40 points on the assessment scale, and the scale has a standard deviation of 40 points; Level 2 spans one standard deviation on the scale. Then, a treatment effect size of 0.5 would imply that a student who, in the absence of treatment, would have been on the border of Levels 1 and 2 on the assessment scale would instead score in the middle of the Level 2 range after receiving the treatment.

We also examine the relationship between the level of attendance among students participating in SSA programs and 5th-grade summer assessment outcomes.

Chapter Four provided an overview of the methods used in this chapter. Next, we present a more detailed introduction to these analytic methods; a technical discussion may be found in Appendix A.

## Analytic Methods

### Propensity Score Weighting and Doubly Robust Regression

In this design approach, we use linear multiple regression models to isolate the difference in proximal-year outcomes among students identified for treatment under the promotion policy versus the expected outcomes of similar students in the control cohort. All available student and school covariates (listed in Chapter Four) are included in the regression models and, thus, controlled for in estimating the impact of being a member of a treated cohort. To ensure that we are comparing equivalent groups, students in the control cohort are weighted in the regression models so that the distributions of all the available covariates among the control cohort students match, as closely as possible, the distributions of these variables across the treated students. These weights are derived using propensity scoring techniques. The estimated regression coefficient for an indicator variable denoting membership in the treated cohort provides an estimate of the average treatment effect for students treated under the policy versus the expected outcomes of those students if they were members of the control cohort. In other words, it provides an estimate of the average treatment effect for treated students. Estimates from this approach are said to be "doubly robust" in that if either the propensity score model is correct or the regression model is correct, the treatment effect estimator will be consistent (Bang and Robins, 2005).

**Available Comparisons.** As we described briefly in the section on study limitations, the change in the administration of the assessments from the city to the state affected the comparisons that were possible between the control cohort and the cohorts treated under the promotion policy. Under the new state assessment program, as well as under the old city and state programs, assessment scores from an individual grade are equated across years. However, results from the old and new assessment programs are not equated. Table 8.2 shows the year in which students in the control and policy cohorts took the 4th- and 5th-grade assessments; the shaded cells indicate the tests administered by the city (5th grade) or by the state (4th grade) under the prior testing program.[1] As a result, for the 5th grade, we found the following pattern:

- The C0 and P1 cohorts took comparable assessments in both the 4th and 5th grades. Thus, we can use students in the C0 cohort to define control groups for students in the P1 cohort who (1) needed services at the beginning of the year, (2) were at risk of retention because of failure to score Level 2 or higher on the spring assessments, or (3) were retained in grade.
- For the remaining two cohorts, proximal-year comparisons using the C0 cohort as a control are unavailable because these later cohorts took spring state-administered tests in the 5th grade; the 5th-grade spring assessment serves as the outcome for students who needed services at the beginning of the year and the baseline score for students at risk of retention.

**Limitations.** As discussed earlier, equated assessment scores for proximal-year comparisons under the propensity scoring design are available only for the first treatment cohort, using the control cohort for comparison. This restriction may be lifted for later grade outcomes (see Chapter Nine for a full discussion of the available comparisons for later academic outcomes). We also needed to pay attention to the limitations of this particular quasi-experimental design, as applied in the context of this reform effort.

---

[1] All summer assessments were administered by the city and continue to be aligned with previous city spring assessments, which were discontinued after 2005.

**Table 8.2**
**Timing of Assessments, by Promotion Status, City- or State-Administered Test, and 5th-Grade Cohort, Proximal and Prior Years**

| Cohort | 4th-Grade Assessment | 5th-Grade Spring Assessment | 5th-Grade Summer Assessment |
|---|---|---|---|
| C0 | 2002–2003 | 2003–2004 | 2003–2004 |
| P1 | 2003–2004 | 2004–2005 | 2004–2005 |
| P2 | 2004–2005 | 2005–2006 | 2005–2006 |
| P3 | 2005–2006 | 2006–2007 | 2006–2007 |

NOTE: Shading represents tests administered by the city (5th grade) or by the state (4th grade) under the prior assessment programs. Cells not shaded represent tests administered by the state under the new assessment program.

In both the propensity scoring and regression phases of the method, we controlled for all available student and school covariates. This list is extensive and includes both variables that are known to be associated with outcomes (e.g., race/ethnicity) and variables that are particular to this reform, such as measures of the level of demand for the services available to at-risk students in the school. It does not, however, necessarily account for all factors that may have affected the success of 5th graders from cohort to cohort at the time. In particular, as discussed in Chapter Three, the promotion policy was just one element of the Children First initiative that was implemented in the city, nested within the wider set of reforms implemented by the state. Estimated treatment effects of the promotion policy under this design might also reflect the effects of these other initiatives; thus, they represent the net results of any changes that occurred between 2003–2004 and 2004–2005 that might have affected students qualifying for treatment under the promotion policy in the latter cohort.

Thus, to complement this design, we present additional information to help disentangle the effect of the promotion policy from that of other city or state efforts simultaneously in place. Along with estimates of effect sizes for students treated under the policy, we also display, where applicable, estimates of the effect of being a member of the policy cohort over expected results under the control cohort for students who were not in need of services or at risk of retention. Efforts that were directed at all students in the policy year but not in the control year would be reflected in the outcomes of all students, not just those treated under the policy. We use the nontreated students closest to the students on the appropriate baseline measure to generate an estimate of the possible effect size of such efforts, using the same doubly robust regression design. We then use a difference-in-difference approach to determine the net effect on the treated students, after adjusting for effects on all students in the policy year. This estimate should be viewed as exploratory at best. There is no guarantee that the mechanisms responsible for the estimated differences in the outcomes for the nontreated students in the policy cohort over those expected had they been in the control cohort would have an equal impact on the treated students.

### Modeling the Relationship Between Intensity of Participation and Outcomes

For the SPA and SSA programs, natural variation exists in the level of treatment that students receive because of differences in attendance. To explore the relationship between program attendance and 5th-grade outcomes (the same approach is used for both SPAs and SSAs), we use GAMM (Lin and Zhang, 1999). These models are similar to a standard linear mixed model approach (Raudenbush and Bryk, 2002), with the exception that we relax the assumption of a linear relationship between attendance and outcome by using a nonparametric form that is able to identify localized changes in the slope that indicate change points in the relationship. Such flexibility allows us to identify minimum thresholds of attendance before an effect is realized, as well as ranges of attendance over which the relationship is more intense. It is important to note that, while the RDD (used in Chapter Nine) and doubly robust methods discussed here are quasi-experimental approaches to understanding the treatment effect on the treated students, the GAMM approach provides only an exploratory examination of the relationships between program attendance and 5th-grade outcomes (see Appendix A for further details).

## Outcomes of Students Needing Services at the Beginning of the Year Based on Prior-Year Scores

### Performance of 5th Graders on the Proximal-Year Spring Assessments

In this section, we use a propensity-weighted doubly robust regression approach to estimate how students in the P1 cohort eligible to receive services in the beginning of the year under the promotion policy would have performed on the 5th-grade assessments had they been members of the C0 comparison cohort. This gives us an estimate of the treatment effect of early intervention services under the promotion policy for treated students in the P1 cohort. Table 8.3 shows the estimated treatment effects on 5th-grade ELA outcomes for different groups of students in the P1 cohort who were eligible to receive AIS for ELA in school and through the Saturday sessions as compared to their expected performance had they been members of the C0 cohort. As mentioned earlier, while the NYC regulations stated that students who scored Level 1 or Level 2 in the prior year were eligible to receive such services, in reality, most schools focused on students scoring Level 1 or low Level 2. Thus, we differentiate among students by whether they scored low or high Level 1 or low or high Level 2 to determine whether the estimated treatment effects differed across these various groups. As discussed earlier, to help disentangle the impact of other reforms on student achievement, we also include

**Table 8.3**
**Estimates of the Differences Between 5th-Grade Spring ELA Outcomes of the 2004–2005 (P1) Cohort and Expected Outcomes Under the Comparison 2003–2004 (C0) Cohort**

| 4th-Grade ELA Level | Estimate | 95% Confidence Interval |
|---|---|---|
| Low Level 1 | 0.506* | (0.248, 0.764) |
| High Level 1 | 0.621* | (0.567, 0.675) |
| Low Level 2 | 0.547* | (0.530, 0.565) |
| High Level 2 | 0.513* | (0.498, 0.528) |
| Low Level 3 | 0.415* | (0.401, 0.429) |
| High Level 3 | 0.347* | (0.329, 0.364) |
| Level 4 | 0.406* | (0.376, 0.436) |

NOTE: The estimates are reported as effect sizes. We standardized using the empirical standard deviation for the city assessments (39 for ELA).

* = statistically significant at the 0.05 level.

estimated treatment effects of being a member of the P1 cohort versus the C0 cohort for students who were not in need of services for ELA at the beginning of 5th grade.

Looking first at ELA, we find that students in this cohort who needed services in the policy year did much better on the 5th-grade assessment than they would have had they not been in the treated group. These differences hold regardless of level of performance on the prior year's assessment. We predict that, on average, the mean scores of the students who were in need of services in ELA were about 0.51–0.62 standard deviations higher than they would have been had these students taken the 5th-grade assessment in the earlier year, when they were not subject to the promotion policy. As a reference point to help put these effect sizes into context, the range of Level 2 scale scores on this assessment spans approximately one standard deviation. The estimate for low Level 1 students has a wide confidence interval, but it is still well above 0. This imprecision reflects the relatively small sample size of this group (fewer than 200 students in the P1 cohort). In contrast, there were well over 1,000 high Level 1 treated students, and the sample sizes for low and high Level 2 students were in the neighborhood of 10,000 each.

Turning attention to the lower half of Table 8.3, we find that students in the P1 cohort who were not in need of AIS for ELA when entering 5th grade also performed significantly better—approximately 0.39 standard deviations on average—on the spring 5th-grade ELA assessment than would have been expected had they been in the control cohort. The significant improvement among the students in the P1 cohort who were not treated under the promotion policy is indicative of additional programmatic differences between the C0 and P1 cohorts. Thus, it is unlikely that the estimated improvements demonstrated by the students eligible for treatment under the policy at the beginning of the P1 year are attributable to the promotion policy alone. They are more likely the result of the cumulative effects of the policy and additional simultaneous efforts to improve student outcomes at that time. To disaggregate these other systematic effects from those of the promotion policy, we use the experience of the nearest half performance level (in this case, low Level 3)[2] as a proxy for the level of improvements by the students treated under the policy that might be attributable to nonpolicy influences. Low Level 3 students in the P1 cohort scored an estimated 0.42 standard deviations better than would have been expected had they been members of the C0 cohort. We subtract this improvement from the estimated cohort effects on the treated students listed in Table 8.3 to produce difference-in-difference estimates of the net effect of the promotion policy on spring 5th-grade outcomes for students eligible for treatment under the promotion policy at the beginning of the P1 cohort year.

The estimated policy effects for ELA are displayed on the left side of Figure 8.1, along with 95-percent confidence intervals for the estimated effects. After accounting for influences other than the promotion policy, the net effects on 5th-grade spring assessment outcomes attributable to the policy (including receipt of AIS and, potentially, Saturday school) range from 0.21 standard deviations for high Level 1 4th graders to 0.10 standard deviations for high Level 2 4th graders. Note that, after accounting for nonpolicy effects, the effect of the promotion policy on low Level 1 4th graders is statistically indistinguishable from zero.

In mathematics, we see in Table 8.4 that effect sizes for being a member of the P1 cohort are smaller than for ELA and—with one exception—we estimate that the scores for these Level 1 and 2 students were about 0.36–0.40 standard deviations higher than they would have been in the nonpolicy control year. For context, the range of Level 2 scores for this mathematics assessment is also narrower than for ELA, spanning approximately 0.78 standard deviations. Thus, relative to the range of Level 2 scores, the estimated effect sizes for being in the P1 cohort over expected performance under the C0 cohort are roughly comparable in the two subjects, excluding low Level 1 students. The estimated mathematics effect was statistically insignificant at the traditional 0.05 significance level for the low Level 1 students; however, the sample size of these students was small, numbering less than 200.

Turning our attention to those students not in need of AIS in mathematics at the beginning of the school year, we see (lower half of Table 8.4) that, as in the ELA results, the nontreated members of the P1 cohort performed significantly higher on the 5th-grade mathematics assessment than predicted if they were part of the control cohort. The average effect size for the low Level 3 students was about 0.33 standard deviations. Taking a difference-in-difference approach would imply, as shown on

[2] Low Level 3 students are defined for our purposes as Level 3 students with a scale score within two standard errors of the Level 3 threshold score.

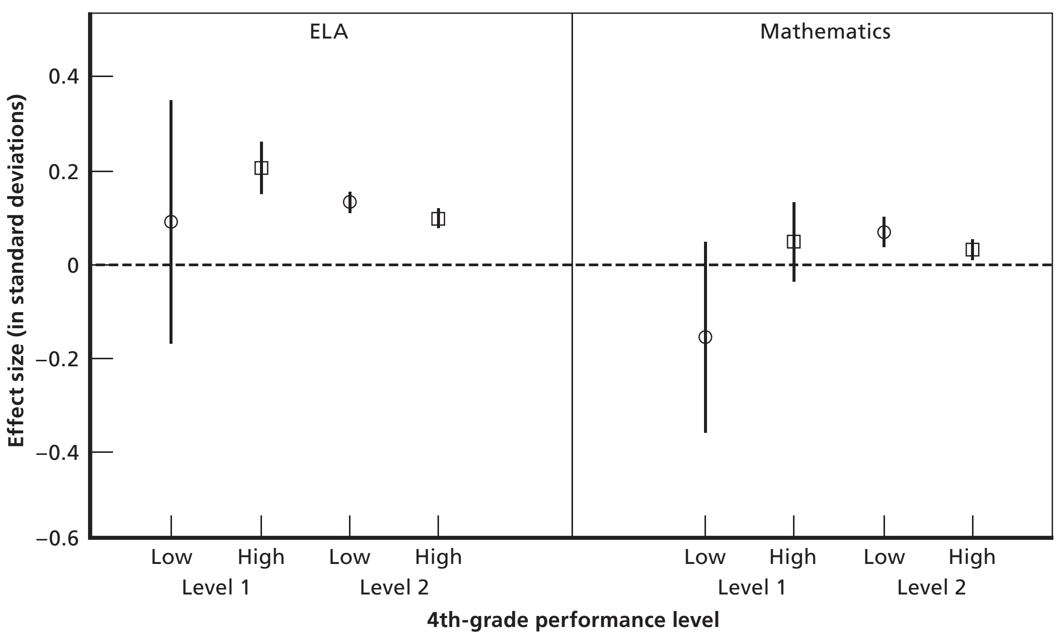
**Table 8.4**
**Estimates of the Differences Between 5th-Grade Spring Mathematics Outcomes of the 2004–2005 (P1) Cohort and Expected Outcomes Under the Comparison 2003–2004 (C0) Cohort**

| 4th-Grade Mathematics Level | Estimate | 95% Confidence Interval |
|---|---|---|
| Low Level 1 | 0.173 | (–0.029, 0.375) |
| High Level 1 | 0.377* | (0.293, 0.460) |
| Low Level 2 | 0.397* | (0.367, 0.426) |
| High Level 2 | 0.359* | (0.342, 0.376) |
| Low Level 3 | 0.327* | (0.314, 0.341) |
| High Level 3 | 0.332* | (0.321, 0.343) |
| Level 4 | 0.288* | (0.271, 0.305) |

NOTE: The estimates are reported as effect sizes. We standardized using the empirical standard deviation for the city assessments (45 for mathematics).

* = statistically significant at the 0.05 level.

**Figure 8.1**
**Difference-in-Difference Estimates of the Impact of the Promotion Policy on the 5th-Grade Spring Outcomes of 2004–2005 (P1) Cohort Students Needing Services at the Beginning of 5th Grade**




RAND *MG894-8.1*

the right side of Figure 8.1, that the estimated net effect of the promotion policy on students eligible for treatment is 0.03–0.07 standard deviations for Level 2 4th graders, while the effect attributable to the promotion policy alone is not statistically significant for P1 5th graders scoring Level 1 on the 4th-grade assessment.

Even relative to the range of Level 2 scores, the estimated difference-in-difference effects attributable to the promotion policy are much smaller for mathematics than for ELA for students who were eligible for services at the beginning of 5th grade.

As noted, the statistical precision of the effect estimates for low Level 1 students was low for both subjects due to the small sample size. Even if we were to assume effects similar to those for high Level 1 and low Level 2 students, the magnitude of those effects would not be great enough to improve the expected outcomes for incoming low Level 1 students sufficiently to reach the minimum Level 2 score. In other words, we estimate that the potential impact of AIS and other early intervention supports under the policy would not be large enough to push the low Level 1 students up to Level 2.

### The Relationship Between Frequency of SPA Attendance and Performance on the 5th-Grade Spring Assessments

As discussed in earlier chapters, students scoring Level 1 or Level 2 on either the ELA or mathematics 4th-grade spring assessments may be invited to attend SPA on Saturdays in the fall and winter months, where they receive additional instruction in both subjects. In this section, we explore the marginal relationship between the level of SPA attendance and performance on the 5th-grade ELA and mathematics assessments using GAMM. We estimated separate models for students in need of services for ELA and those in need of services for mathematics in the P2 and P3 cohorts. SPA attendance data were not available for the P1 cohort.

In our GAMM approach, we included district and region as random effects,[3] and we conditioned on available student covariates typical of such analyses: prior 4th-grade scores in both ELA and mathematics, gender, race/ethnicity, eligibility for free or reduced-price lunch, ELL status, and number of days absent in the school year. We used a subject-specific number of SPA sessions: For students at risk in ELA, the maximum number of sessions was capped at the number held before the January ELA assessment; for students at risk in mathematics, the maximum was set at the total number of sessions held because the SPA sessions ended before the mathematics assessment was administered. Thus, the maximum number of sessions for ELA was 10 for the P2 cohort and eight for the P3 cohort; for mathematics, the maximum number of sessions was 19 and 17, respectively. Students who did not attend any sessions were excluded from the models.

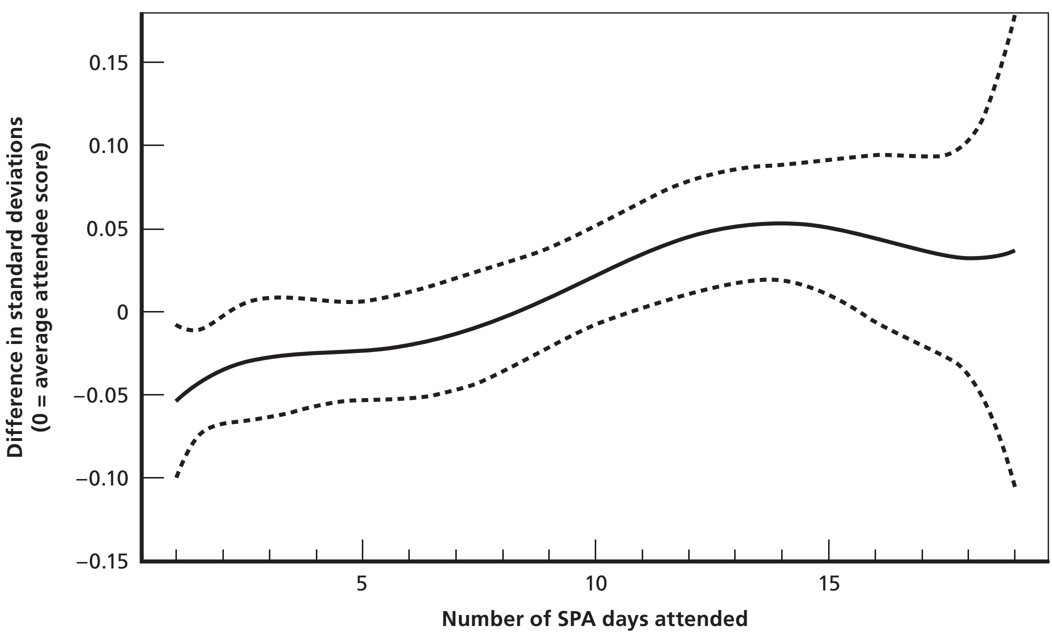
---

[3] Although the restructuring eliminated the regional structure in June 2007, regions were still in place during the P1 and P2 cohorts. Districts were retained under the new structure.

For both cohorts P2 and P3, the GAMM results for ELA were statistically insignificant; i.e., under the GAMM model in both years, the marginal effect of an additional day of SPA attendance on the spring ELA outcome was not distinguishable from zero.[4] The lack of a relationship between increased SPA attendance and spring ELA performance would be consistent with the idea that, in the presence of the other early intervention services, there is limited additional value in the SPA for ELA, or that there were too few days offered and too poor attendance for any potential benefit to manifest.

For mathematics, the GAMM results do imply a significant relationship between SPA attendance and the spring assessment score in both years. One advantage to the GAMM approach is the ability to visualize the relationship between dosage and outcome, along with any nonlinearity or change points that may exist (Figures 8.2 and 8.3). The x-axis on each figure indicates the number of days attended. The y-axis indicates the impact on the spring mathematics assessment scale score (in standard deviations so as to mirror our approach throughout this chapter) as the number of days attended increases. The y-axis is centered at the average spring score of the

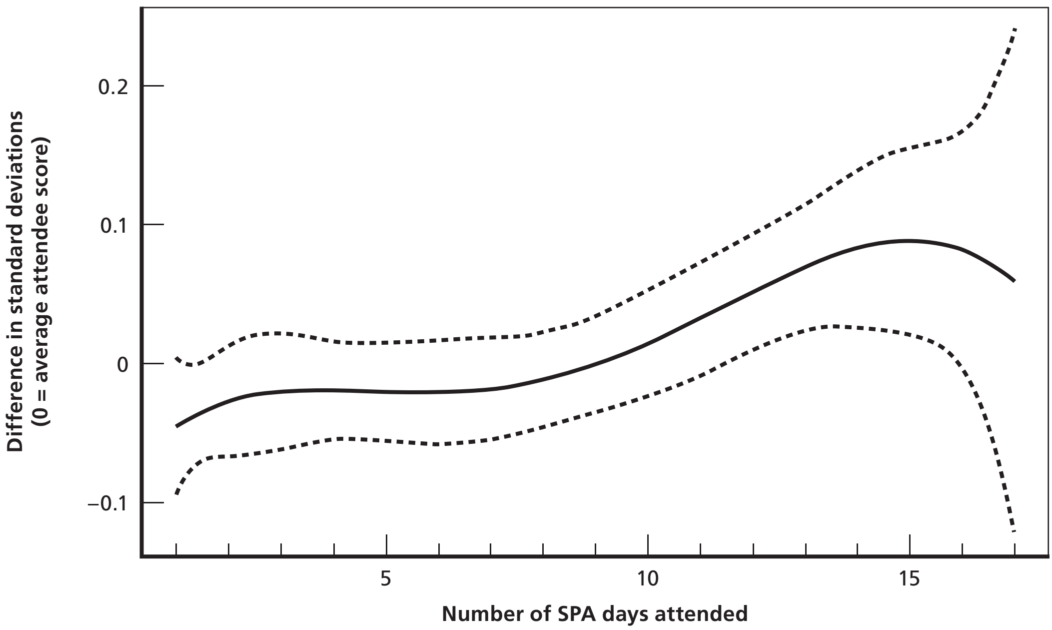
**Figure 8.2**
**Relationship Between Number of SPA Sessions Attended and 5th-Grade Spring Mathematics Assessment Scale Score, 2005–2006 (P2) Cohort**




RAND *MG894-8.2*

[4] This result is also supported by a more traditional fully parametric hierarchical linear model of the relationship between SPA attendance and spring ELA outcomes.

**Figure 8.3**
**Relationship Between Number of SPA Sessions Attended and 5th-Grade Spring Mathematics Assessment Scale Score, 2006–2007 (P3) Cohort**




RAND *MG894-8.3*

students attending the SPA. Figure 8.2 suggests that the effect of SPA attendance in the P2 cohort slopes upward until about 14 sessions. Between six and 14 days, we see an increase of approximately 0.075 standard deviations. As indicated by zero on the y-axis, the typical SPA student falls into this range. Part of the explanation for the discontinuance of an upward trend after 14 sessions is that not all the 19 SPA sessions focused on mathematics, and students may have failed to get much benefit with respect to their mathematics outcome from sessions that focused largely on ELA. The decreased precision in the estimated effect of attendance beyond 15 days reflects the fact that very few students attended all the sessions. A similar pattern emerges for SPA attendance in the P3 cohort, with an upward trend extending between seven and 15 days of attendance, with a student attending 15 sessions expected to score roughly 0.10 standard deviations above a student attending seven sessions.

These exploratory results suggest that there is little benefit to attending fewer than six or seven days, and, similarly, there appears to be no benefit in mathematics in attending more than 14 or 15 sessions. The peak of 14–15 days of additional mathematics attendance does not necessarily imply that additional days with a mathematics focus may not be beneficial, since the marginal additional days most likely were focused on ELA (and students might find have found these sessions useful for ELA).

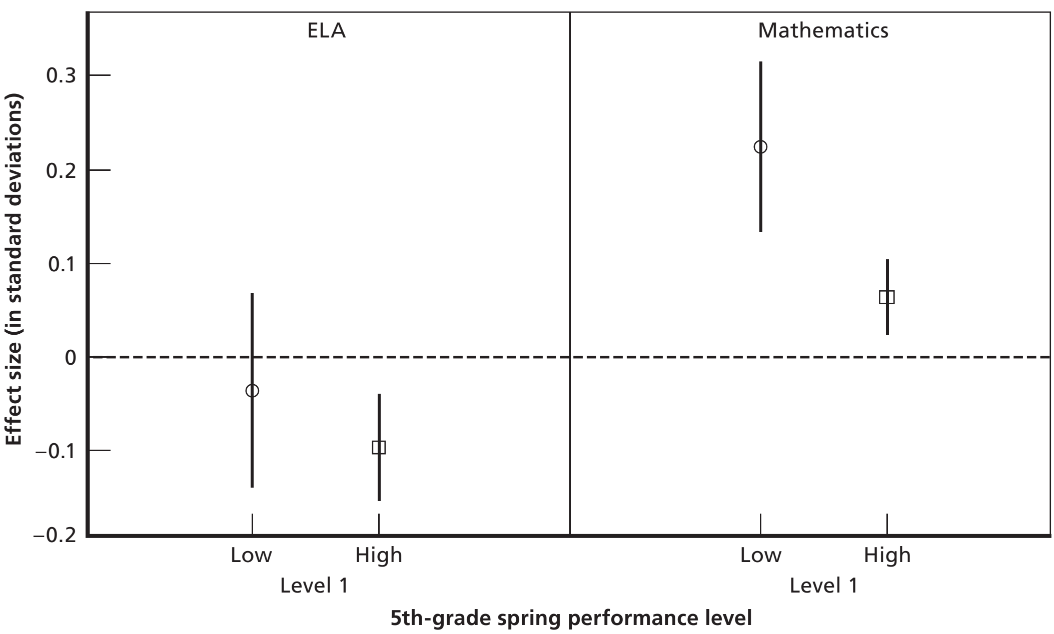
We now turn to the summer outcomes and retention status of students who were at risk of retention because they scored Level 1 on the spring assessments.

## Summer Outcomes of Students at Risk of Retention Based on Performance on the Proximal-Year Spring Assessments

### Performance of At-Risk 5th Graders on the Proximal Summer Assessments

We next turn our attention to the impact of the SSA on summer performance of 5th graders at risk of retention, relative to their expected performance under the prior summer school system. Here, we focus on a doubly robust regression approach.[5] Figure 8.4 presents the estimated SSA treatment effect on 5th-grade summer outcomes for P1 students who scored Level 1 on the spring assessments and thus were at risk of being retained in the 5th grade.[6] Again, these effect sizes represent the estimated standardized difference in performance of at-risk students in the treatment year versus their expected performance in the control year. For the P1 cohort, the 5th-grade summer

**Figure 8.4**
**Estimates of the Differences Between 5th-Grade Summer Outcomes of Students in the 2004–2005 (P1) Cohort and Their Expected Outcomes in the Comparison 2003–2004 (C0) Cohort, Students at Risk of Retention**




RAND *MG894-8.4*

[5] Regression discontinuity analyses are not applicable to summer performance outcomes because students above the treatment threshold do not have summer outcomes.

[6] SSA attendance data were not available for the summer of 2005. For spring Level 1 students in the P1 cohort, we used the presence of at least one summer test score for students not promoted via the spring portfolio as a proxy for SSA attendance.

assessments are on the same scale as the 5th-grade spring assessments. Looking first at ELA, we find that spring high Level 1 students performed slightly worse, by 0.10 standard deviations, under the promotion policy than how these students would have performed in the comparison year. We find no significant difference in the performance of those who scored low Level 1 on the spring ELA assessment and their expected performance had they been in the C0 cohort and attending a typical summer school.

We estimated that P1 students who scored Level 1 on the spring mathematics assessments did better in the treatment year (in an SSA) than their expected performance in the control year (in a typical summer school), and the effect was larger for low Level 1 students (0.22) than for high Level 1 students (0.06).

#### The Relationship Between Frequency of SSA Attendance and Performance on the Summer Assessments

We used GAMM to explore the relationship between SSA attendance and performance on the summer assessments for two 5th-grade cohorts (P2 and P3).[7] SSA attendance data were not available for the P1 cohort. Unlike SPA attendance, which was generally low, SSA attendance was high, with a large majority of attendees attending at least 14 SSA instructional days.

As was the case with SPA attendance, neither of the two 5th-grade ELA models detected a significant marginal relationship between SSA attendance and summer ELA assessment outcomes. We found a flat relationship beyond 14 days; because most students attended 14 days or more, little power exists to detect a relationship below 14 days.[8]

GAMM results for mathematics indicate a statistically significant relationship between days attended and summer mathematics outcomes in 2006 but an insignificant result in 2007. Because the number of at-risk students declined dramatically in the P3 cohort, the number of students attending SSA for mathematics declined by 70 percent in 2007 from the prior summer, greatly reducing the power to detect significant effects.

We illustrate the summer 2006 mathematics results in Figure 8.5.[9] Similarly to the SPA graph, the y-axis indicates the summer mathematics assessment scale score in standard deviations of the scale and is centered at the average summer score of the students attending the SSA. Between seven and 14 days of SSA attendance, the expected

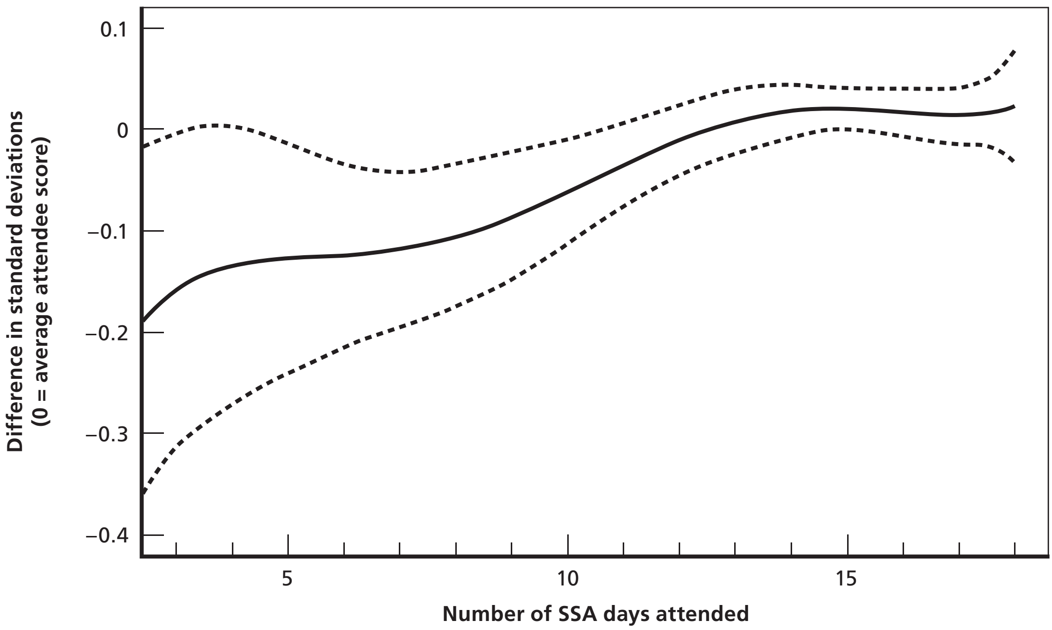
---

[7] SSA attendance data provided by NYCDOE included testing days. We estimated the actual number of instructional days attended, net of testing days, and use this "instructional days" variable in the models discussed in this section. Throughout this section, whenever we mention the number of SSA days attended, we are only including SSA instructional days.

[8] The GAMM graphs also revealed that linearity assumptions of a more traditional hierarchical linear model would be violated in these two cases.

[9] Fewer than three days of attendance is not shown in the figure to improve readability.

**Figure 8.5**
**Relationship Between Number of SSA Sessions Attended and 5th-Grade Summer Mathematics Assessment Scale Score, 2005–2006 (P2) Cohort**




RAND *MG894-8.5*

mathematics summer outcome increases by roughly 0.14 standard deviations. Beyond 14 days attended, the effect stays essentially flat; more than two-thirds of attendees in 2006 were present for at least 14 SSA instructional days and thus received the full benefit implied in this relationship. This heavy skew toward a large number of days attended, along with the nil marginal gain beyond 14 days, explains why the GAMM graph in Figure 8.5 peaks just above zero (the mean standardized outcome).

## Summary

### Effect of Being Identified as Needing Services at the Beginning of the School Year: P1 Cohort

Using the C0 cohort as a control to estimate the effects of being identified for services in a model that combined propensity score matching with a doubly robust regression, we found that students identified for services at the beginning of the school year scored about 0.50 standard deviations higher on the spring ELA assessment and about 0.40 standard deviations higher on the spring mathematics assessment than expected if they were in the comparison cohort. After considering the potential impact of other reforms simultaneously in place, we estimate that the effects directly attributable to the promotion policy are 0.10–0.21 standard deviations in ELA for Level 2 and high Level 1 students, with no significant effect for low Level 1 students, and 0.03–0.07 standard deviations in mathematics for Level 2 students, with no significant effect for Level 1 students.

#### Relationship Between SPA Attendance and Spring Assessment Outcomes: P1 and P2 Cohorts

Our results suggest that students who attended 14–15 Saturday sessions typically scored between 0.08 and 0.10 standard deviations higher on the spring mathematics assessment than students who attended six to seven sessions. We did not find a relationship between spring assessment outcomes and more frequent attendance in ELA, for which only eight to 10 sessions were offered prior to the assessment date.

#### Effect of Being Mandated to Attend SSA: P1 Cohort

Students in the policy cohorts who scored Level 1 on the spring assessments were mandated to attend an SSA, which, under NYCDOE guidelines, had small class sizes and a set instructional schedule. Students in the comparison cohort attended summer school. Relative to how they would have been expected to perform on the summer assessments in the control year, high Level 1 at-risk students in the in the P1 cohort performed slightly worse in ELA under the promotion policy by about 0.10 standard deviations and somewhat better in mathematics. Low Level 1 students gained about 0.20 standard deviations in mathematics over expected performance in the control year.

#### Relationship Between SSA Attendance and Summer Assessment Outcomes: P2 and P3 Cohorts

Using data on all attending students from both the cohorts, we found no relationship between more frequent SSA attendance and summer ELA outcomes, but we did find a small but statistically significant relationship between increased SSA attendance and 5th-grade summer assessments in mathematics for the P2 cohort in the summer of 2006. In 2007, the number of students who were mandated to attend SSA fell dramatically. Our P2 results imply that students attending 14 sessions typically scored about 0.14 standard deviations higher on the summer mathematics assessment than students attending seven sessions.

The next chapter examines future outcomes of both 5th-grade students needing services at the beginning of the school year and those at risk of retention based on their performance on the spring assessment.

CHAPTER NINE

# Future Outcomes of Students at Risk of Retention

*Louis T. Mariano, Sheila Nataraj Kirby, and Al Crego*

Chapter Eight examined the proximal-year outcomes for students who needed services at the beginning of the school year and those who were at risk of retention based on their performance on the spring assessments. In this chapter, we are interested in examining the future outcomes of students in need of services at the beginning of the school year, those who were at risk of retention, and those who were retained in grade once,[1] relative to their predicted performance in the absence of treatment under the policy. Table 9.1 presents the research questions addressed in this chapter along with the methods used to address them. This chapter presents results for the 5th-grade cohorts; see Appendix E for the 3rd-grade results.

**Table 9.1**
**Future Outcomes: Research Questions and Methods**

| Research Question | Method |
|---|---|
| Students needing services at the beginning of the year based on prior year's scores, relative to a comparison group | |
| How did these students perform on the spring assessments in future grades? | Propensity score weighting and doubly robust regression |
| Students at risk of retention based on performance on proximal-year spring assessments, relative to a comparison group | |
| How did these students perform on the spring assessments in future grades? | Propensity score weighting and doubly robust regression<br>RDD models |
| Students retained in grade (5th and 3rd grades), relative to a comparison group | |
| How did these students perform on the spring assessments in future grades? | Propensity score weighting and doubly robust regression<br>RDD models |

[1] Note that, due to data constraints, we were unable to analyze future outcomes for one small but important subgroup: students who had been retained in 5th grade under the promotion policy more than once. These students in the P1 cohort had not yet progressed to 7th grade by the 2007–2008 school year.

## Analytic Methods

As discussed previously, two natural comparison groups were available for use in quasi-experimental designs to examine the proximal-year outcomes. As in Chapter Eight, we use propensity score–weighted doubly robust regression models to predict how students treated under the policy may have performed in the absence of the policy as members of the C0 cohort. In that chapter, we also discussed some of the limitations of doubly robust regression models. However, as the students progress to later grades, additional concerns arise; these are discussed next.

We used students in the policy year who were just above the treatment thresholds (i.e., students who were not at risk but very close to the cutoff) as a natural control group for students who were at risk and just below the treatment thresholds, and we estimated the treatment effect on the treated students near the threshold using RDD.

Additional technical details for both methods can be found in Appendix A.

### Propensity Score Weighting and Doubly Robust Regression

**Available Comparisons.** Table 9.2 shows the year in which promoted and retained students took higher-grade assessments and the change from city- to state-administered assessments, with the shaded cells showing the tests administered by the city or by the state under the previous testing program.[2] For the 5th-grade cohorts, the table shows the following pattern of available comparisons:

- The C0 and P1 cohorts both took city-administered assessments in the 5th grade.
  - Thus, we can use the former to predict performance under the C0 cohort for students in the P1 cohort who (1) needed services at the beginning of the year, (2) were at risk of retention because of failure to score Level 2 or higher on the spring assessments, and (3) were retained in grade.
  - The 7th-grade outcomes of both promoted and retained students in these two cohorts are comparable, because all four subgroups took the state-administered assessment (which was not true of 6th grade), and the state assessments are aligned within grade across years.
- For the P2 and P3 cohorts, we were unable to carry out any comparisons using the control cohort with available data through the 2007–2008 school year.

**Limitations.** The major concern regarding the use of propensity score–weighted regression models for examining later grade outcomes is that students in the cohort treated under the policy may be retained. Unlike in Chapter Eight, where the proximal outcomes for the P1 cohort and the C0 cohort were measured one year apart,

2 All summer assessments were administered by the city and continue to be aligned with previous city spring assessments, which were discontinued after 2005.

**Table 9.2**
**Timing of Spring Assessments, by Promotion Status and City- Versus State-Administered Test**

| Cohort | 4th-Grade Assessment | 5th-Grade Assessment | Promotion Status | 6th-Grade Assessment | 7th-Grade Assessment | 8th-Grade Assessment |
|---|---|---|---|---|---|---|
| C0 | Spring 2003 | Spring 2004 | Promoted | Spring 2005 | Spring 2006 | Spring 2007 |
| | | | Retained | Spring 2006 | Spring 2007 | Spring 2008 |
| P1 | Spring 2004 | Spring 2005 | Promoted | Spring 2006 | Spring 2007 | Spring 2008 |
| | | | Retained | Spring 2007 | Spring 2008 | |
| P2 | Spring 2005 | Spring 2006 | Promoted | Spring 2007 | Spring 2008 | |
| | | | Retained | Spring 2008 | | |
| P3 | Spring 2006 | Spring 2007 | Promoted | Spring 2008 | | |
| | | | Retained | | | |

NOTE: Shaded cells represent tests administered by the city (5th grade) or by the state (4th grade) under the prior assessment programs. Cells not shaded represent tests administered by the state under the new assessment program.

comparisons of the treated and control cohorts in later grades were two years apart for retained students in the policy cohort and promoted control students. For retained students, this meant that we have to account for the cumulative effects of any additional programmatic changes that occurred between 2003–2004 and 2005–2006 that may have directly or indirectly affected these students.

As in Chapter Eight, we present additional information to help disentangle the effect of the promotion policy from that of other city or state efforts simultaneously in place. For promoted students, we use estimates of the effect of being a member of the policy cohort for students who were not in need of services or at risk of retention and a difference-in-difference approach to calculate the net effect of the policy treatment on at-risk students. For retained students, in addition to estimating the effect of being a nontreated member of the P1 cohort (relative to the control), we also estimate the effect of being a member of the P2 cohort relative to expected results under the P1 cohort for students who were not in need of services at the beginning of 5th grade to account for other programmatic differences that may have arisen in the retained year. The sum of these two nontreated effect estimates is used in a difference-in-difference approach to obtain the net policy effect on the retained students. As noted earlier, this approach may not adjust fully for the other influences, but it represents a reasonable approximation and is the best we can do with our data.

### Regression Discontinuity Design Models

Propensity score–weighted regressions provide a means of estimating the average treatment effect for all treated subjects. The methods should provide consistent estimates of the treatment effects. However, while we account for all available covariates in fitting the models, unobserved differences between treated and control cohorts could affect the outcome analyses. To complement the doubly robust regressions and to extend the analyses, we estimate a number of RDD models to compare students within a policy cohort whose assessment scores were just below or just above a threshold on a treatment assignment variable. These models control for unobserved covariates as long as their distribution is continuous at the treatment threshold; thus, they are superior to doubly robust regressions with respect to internal validity. However, estimated treatment effects are limited to treated students near the assignment variable threshold.[3]

The identification of an assignment variable with a threshold value (i.e., cutoff) that triggers treatment is crucial to the RDD framework. A treatment assignment variable is a measure that is used, at least in part, to assign a student to a treatment, and a treatment threshold is a value of the treatment assignment variable at which there exists a marked increase in the probability of a student being assigned to treatment. For example, students with the highest spring ELA assessment scale score in the range of scores that map to performance Level 1 will have a much higher average probability of attending SSA than students with the lowest ELA spring scale score in the Level 2 range.

There are two promotion policy treatments that lend themselves to RDD models: SSA enrollment and retention in 5th grade. The relevant assignment variables for these treatments are the spring 5th-grade and summer 5th-grade assessments, respectively. Under the promotion policy, treatment assignment is typically defined by performance-level scores, while the level scores represent a four-category ordinal variable. For each assessment, we can identify a scale score on a continuous range that marks the threshold between adjacent performance levels.[4] The relevant RDD outcomes for the SSA and retention treatments are assessment scores in future grades.[5]

**Methods.** To estimate the regression discontinuity models, we fit local linear regression models on either side of the treatment assignment threshold, including polynomial terms of the assignment variable, where appropriate, to estimate the effect of treatment eligibility at the threshold value on both the outcome measure and the probability of treatment. The ratio of these effects identifies the treatment effect on the treated students at the threshold. The design is implemented in a two-stage least-squares framework, generally following the procedure described by Imbens and

---

[3] More precisely, the RDD models provide an estimate of the treatment effect at the point of the assignment threshold. As a matter of generalizability, it would be appropriate to attribute these estimates only to those students with scores nearest the treatment threshold.

[4] All the assessments in both the state test program that began in 2006 and the prior city and state programs used a raw-score-to-scale-score conversion method; thus, there is always a one-to-one mapping between a performance-level score and a range of scale scores.

[5] Students above the SSA treatment threshold, i.e., spring Level 2 students, do not take the summer assessment, so the first available outcomes for comparison are the 6th-grade spring assessments.

Lemieux (2008), and standard errors are adjusted for discrete support in the assignment variable (Lee and Card, 2008) because only a finite set of scale scores may actually be observed for any given assessment.

Exposure to the NYC 5th-grade promotion policy SSA and retention treatments is affected by a student's standardized assessment scores (5th-grade spring or 5th-grade summer). Students scoring Level 1 on a particular ELA or mathematics assessment had a significantly increased probability of receiving the corresponding treatment. While the level scores represent a four-category ordinal variable, for each assessment, we can identify a scale score on a continuous range that marks the threshold between Level 1 and Level 2, with scale scores below the threshold mapping to Level 1 and scores above the threshold mapping to Level 2.

In each of the cases we analyze, not all students below the threshold will receive the treatment. For example, a student may avoid being retained by demonstrating Level 2 proficiency via a portfolio. Thus, having an assignment variable score on the treatment side of the threshold does not imply that a student will receive the treatment with absolute certainty. To account for this, we used a fuzzy RDD to model the effect of the three treatments. The fuzzy RDD approach is detailed in Appendix A.

Scoring Level 1 on either the ELA or the mathematics portion of one of the assessments used for assignment will trigger exposure to the treatment. However, there are not necessarily separate treatments for ELA and mathematics. All students attending an SSA will receive instruction in both ELA and mathematics, even if they scored Level 1 in only one of these subjects, and retained students repeat the entire grade curriculum. As such, the assignment mechanism is actually two-dimensional, and either of the subject scale scores being below the subject threshold scale causes an increase in the probability of exposure to a more likely common treatment.

Previous regression discontinuity analyses examining the effects of summer school and retention (Roderick and Nagaoka, 2005; Jacob and Lefgren, 2002; Matsudaira, 2008) addressed the problem of dual-exposure threshold scores by restricting analyses to a set of students who were above the threshold in one subject and below the threshold in the other, so only the threshold for the latter subject was of consequence for treatment exposure. For the purposes of this monograph, we adopt the approach taken by these other studies—in particular, that of Matsudaira (2008).[6] This approach focuses the RDD modeling effort to determine the treatment effect on a particular set of treated students, those closest to meeting the promotion standards.

---

6 We acknowledge that our current approach may be overly restrictive in that it excludes a subset of the treated students—those who are at risk in both subjects. These excluded students may have different experiences with respect to the treatment because they are attempting to meet standards in both subjects. The treatment effects for these excluded students may not be the same as for the included students, and excluding such differences could bias the estimated treatment effect. Further analyses are currently under way to model a multidimensional treatment assignment variable.

Two particular challenges to fitting the RDD models are choosing the bandwidth of the assignment variable (i.e., the range of values around the assignment variable used to estimate the effect at the treatment threshold) and choosing what order of polynomial terms of the assignment variable may be appropriate to include in the model. We adapt the cross-validation approach described by Ludwig and Miller (2005) and Imbens and Lemieux (2008) to choose the bandwidth and polynomial terms (details may be found in Appendix A). The number of observations included in the model both above and below the threshold is included along with the RDD estimates.

**Limitations.** The one-year delay in the observation of the later-grade outcomes for retained students creates two complications for the RDD approach. First, we must have comparable outcomes for both the promoted and retained students in later grades. As seen in Table 9.2, starting with 6th grade, later-grade assessment outcomes of all treated 5th-grade cohorts fall under the new state assessment program and so are aligned across years. While we can use RDD to examine 6th-grade outcomes for both the P1 and P2 cohorts, 7th-grade outcomes were available only for the first cohort, as our data extend only through the spring of 2008.

The second concern is that retained students joining the subsequent 5th-grade cohort may be exposed to further treatment, and the observed treatment effects may include the effects of these programmatic differences as well. The RDD models, in this case, are actually identifying the effect of being assigned to the subsequent 5th-grade cohort; that effect includes both a retention component and a programmatic differences component. To address this concern, we made the same adjustment as for the propensity score–weighted regressions when possible, subtracting out the estimated effect of being a nontreated student in the subsequent 5th-grade cohort versus the original cohort. As is the case with the weighted regression adjustments, there is no guarantee that the mechanisms responsible for the estimated differences in the outcomes for the nontreated students in the subsequent policy cohort had an equal impact on the treated students or that treated students received equal benefits from other reform efforts.

## Future Outcomes of Students Needing Services at the Beginning of 5th Grade Based on Prior-Year Scores

In this section, we examine the impact of the promotion policy on the future 7th-grade outcomes of students who were eligible for intervention services at the beginning of the 5th-grade year. We used propensity-weighted doubly robust linear regression to estimate the effect of being in the P1 5th-grade cohort versus expected performance under the C0 comparison cohort. These estimates are displayed in Tables 9.3 and 9.4 for ELA and mathematics, respectively. To help disentangle the impact of other reforms on student achievement, we also include estimated treatment effects of being a member of the

**Table 9.3**
**Estimates of the Differences Between 7th-Grade Spring ELA Outcomes of the 2004–2005 (P1) Cohort and Expected Outcomes Under the Comparison (C0) Cohort, 4th-Grade Spring Baseline**

| 4th-Grade ELA Performance | Estimate | 95% Confidence Interval |
|---|---|---|
| Low Level 1 | 0.299* | (0.005, 0.593) |
| High Level 1 | 0.410* | (0.343, 0.477) |
| Low Level 2 | 0.286* | (0.267, 0.305) |
| High Level 2 | 0.269* | (0.252, 0.285) |
| Low Level 3 | 0.189* | (0.173, 0.205) |
| High Level 3 | 0.102* | (0.083, 0.120) |
| Level 4 | 0.061* | (0.034, 0.087) |

NOTE: The estimates are reported as effect sizes. We standardized using the theoretical standard deviation for the state assessments (40 for ELA).

* = statistically significant at the 0.05 level.

**Table 9.4**
**Estimates of the Differences Between 7th-Grade Spring Mathematics Outcomes of the 2004–2005 (P1) Cohort and Expected Outcomes Under the Comparison (C0) Cohort, 4th-Grade Spring Baseline**

| 4th-Grade Mathematics Performance | Estimate | 95% Confidence Interval |
|---|---|---|
| Low Level 1 | 0.296* | (0.021, 0.571) |
| High Level 1 | 0.496* | (0.411, 0.581) |
| Low Level 2 | 0.422* | (0.384, 0.460) |
| High Level 2 | 0.318* | (0.296, 0.340) |
| Low Level 3 | 0.333* | (0.315, 0.351) |
| High Level 3 | 0.333* | (0.320, 0.346) |
| Level 4 | 0.374* | (0.356, 0.391) |

NOTE: The estimates are reported as effect sizes. We standardized using the theoretical standard deviation for the state assessments (40 for mathematics).

* = statistically significant at the 0.05 level.

P1 cohort versus the control cohort on 7th-grade outcomes for students who were not in need of services in ELA at the beginning of 5th grade.

Tables 9.3 and 9.4 underscore the need to disaggregate the students needing services at the beginning of the year by whether they scored at the lower or higher end of the level range. Students across the full range who had been eligible to receive services during the 5th grade (shown in the upper portion of each table) did significantly better on the 7th-grade ELA and mathematics spring assessments than they would have if they had not been part of the treated cohort. For example, compared to expected performance under the C0 cohort, students scoring high Level 1 in ELA, on average, scored about 0.41 standard deviations higher, and students who scored Level 2 on the 4th-grade assessment scored, on average, 0.27–0.29 standard deviations higher. In mathematics, the effect sizes were larger—0.50 standard deviations for high Level 1 students and 0.32–0.42 standard deviations for Level 2 students.

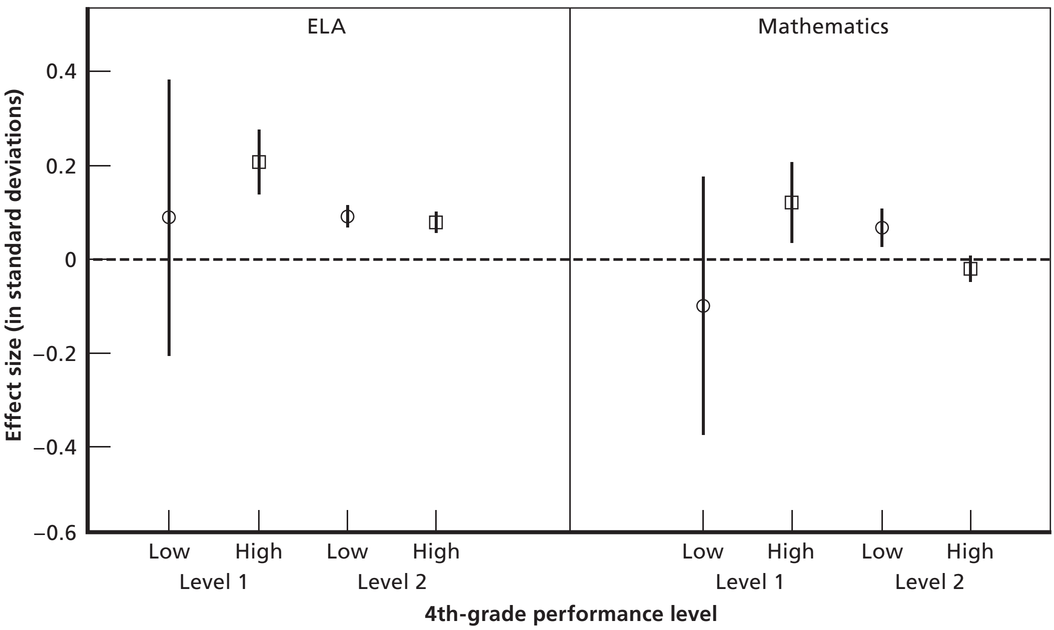
Turning our attention to the lower portions of Tables 9.3 and 9.4, we find that students in the P1 cohort who were not in need of AIS entering 5th grade also performed significantly better—the effect sizes were approximately 0.19 standard deviations for low Level 3 students in ELA and approximately 0.33 standard deviations for low Level 3 students in mathematics on the spring 7th-grade assessments. Recall that students retained in the 2004–2005 policy year become members of the P2 cohort; as such, to produce a policy effect estimate net of any other concurrent efforts to raise student performance, we also needed to adjust for the effect of systematic nonpolicy differences between the P1 and P2 cohorts for the fraction of students ultimately retained into the P2 5th-grade cohort. Low Level 3 students in the P2 cohort scored about 0.09 standard deviations higher than they would have if they were part of the P1 cohort on the 7th-grade ELA assessment and performed, on average, about 0.15 standard deviations higher on the 7th-grade mathematics assessment (not shown in the tables.

As we did in the previous chapter, we use a difference-in-difference approach to "net out" the effect of the promotion policy on 7th-grade ELA outcomes for students in need of services at the beginning of the 5th-grade year. The adjusted effect sizes attributable to the promotion policy for ELA ranged from 0.21 for students scoring high Level 1 on the 4th-grade assessment to 0.08 for those scoring high Level 2. For mathematics, the adjusted effect sizes are, on average, 0.16 standard deviations for high Level 1students and 0.09 for low Level 2 students. All of these estimated effects are significant and capture the impact of AIS, Saturday and summer school (for some students), retention (for a small percentage of students), and any net difference in additional instructional services received by treated and comparison students through the 7th-grade assessments that was not already captured by our difference-in-difference adjustment (Figure 9.1).[7]

---

7 Both P1 and C0 students may have received additional instructional services in later grades. Thus, the estimated effects from the doubly robust regression model capture the impact of the services that a P1 student may have received in later grades relative to what that student would have received had he or she been in the C0 cohort. The difference-in-difference adjustment corrects for any cohort-over-cohort differences in instructional services received in later grades that are common to the treated P1 students and the nontreated P1 students used for the adjustment, relative to their respective expected experiences under the C0 cohort. Services that were not common to both groups, such as the net difference in later promotion policy services received by treated P1 students in 7th grade relative to what they would have received had they been in the C0 cohort, remain in the adjusted net estimate of the effect of the policy presented here.

**Figure 9.1**
**Difference-in-Difference Estimates of the Impact of the Promotion Policy on the 7th-Grade Spring Outcomes of 2004–2005 (P1) Cohort Students Needing Services at the Beginning of 5th Grade**




RAND *MG894-9.1*

After the adjustment, low Level 1 students did not score significantly higher or lower than their expected outcomes had they been members of the control cohort. However, due to small sample sizes, these estimates have low levels of precision.

## Future Outcomes of Students at Risk of Retention in the 5th Grade

### Performance of All 5th Graders on the 7th-Grade Spring Assessments

We also examined 7th-grade outcomes for students in the P1 cohort who were at risk of retention in the 5th grade because they scored Level 1 on the 5th-grade ELA or mathematics assessment. The effect sizes—which measure the estimated difference between how at-risk students performed in the 7th grade and their expected performance had they been part of the control cohort—are shown in Tables 9.5 and 9.6. The results are disaggregated for at-risk promoted and at-risk retained students because effects differ substantially between the two groups. To help disentangle the impact of other reforms on student achievement, we also include estimated effects for students who were not at risk of retention in 5th grade.

Students at risk in ELA who were retained scored, on average, between 0.61 and 0.70 standard deviations higher on the 7th-grade ELA assessment, while high Level 1

**Table 9.5**
**Estimates of the Differences Between 7th-Grade Spring ELA Outcomes of the 2004–2005 (P1) Cohort and Expected Outcomes Under the Comparison (C0) Cohort, 5th-Grade Baseline**

| 5th-Grade ELA Performance and Outcome | Estimate | 95% Confidence Interval |
|---|---|---|
| Low Level 1—retained | 0.609* | (0.409, 0.809) |
| High Level 1—retained | 0.702* | (0.590, 0.813) |
| Low Level 1—promoted | –0.907* | (–1.474, –0.340) |
| High Level 1—promoted | –0.018 | (–0.150, 0.113) |
| Low Level 2 | –0.037* | (–0.066, –0.009) |
| High Level 2 | –0.116* | (–0.135, –0.097) |
| Low Level 3 | –0.146* | (–0.164, –0.129) |
| High Level 3 | –0.187* | (–0.202, –0.172) |
| Level 4 | –0.183* | (–0.212, –0.155) |

NOTE: The estimates are reported as effect sizes. We standardized using the theoretical standard deviation for the state assessments (40 for ELA).

* = statistically significant at the 0.05 level.

**Table 9.6**
**Estimates of the Differences Between 7th-Grade Spring Mathematics Outcomes of the 2004–2005 (P1) Cohort and Expected Outcomes Under the Comparison (C0) Cohort, 5th-Grade Baseline**

| 5th-Grade Mathematics Performance and Outcome | Estimate | 95% Confidence Interval |
|---|---|---|
| Low Level 1—retained | 0.602* | (0.468, 0.736) |
| High Level 1—retained | 0.648* | (0.584, 0.712) |
| Low Level 1—promoted | –0.163* | (–0.325, –0.002) |
| High Level 1—promoted | 0.027 | (–0.056, 0.110) |
| Low Level 2 | 0.058* | (0.034, 0.081) |
| High Level 2 | 0.005 | (–0.012, 0.022) |
| Low Level 3 | 0.004 | (–0.012, 0.020) |
| High Level 3 | 0.052* | (0.027, 0.078) |
| Level 4 | 0.160* | (0.140, 0.180) |

NOTE: The estimates are reported as effect sizes. We standardized using the theoretical standard deviation for the state assessments (40 for mathematics).

* = statistically significant at the 0.05 level.

ELA promoted students showed no significant difference in performance on the 7th-grade ELA assessment over their expected outcome had they been in the control cohort. Low Level 1 students who were promoted posted a large and significant decline in ELA of about 0.91 standard deviations, although this group was very small (32 students).

In mathematics, students who were at risk in 5th grade and retained also had large effects, scoring, on average, between 0.60 and 0.65 standard deviations higher on the 7th-grade mathematics assessment than they would have as part of the control cohort. Similar to the ELA results, there was no difference in the performance of high Level 1 promoted students relative to how they would have performed as part of the control cohort. Low Level 1 students had a small but significant negative effect of –0.16 standard deviations, but the sample size for the promoted low Level 1 mathematics students was fairly small (118 students).

It is not surprising that we see larger positive effects for retained students than for the at-risk promoted group. In the comparison cohort, at-risk students also attended summer school similar to the SSAs that treated students attended. Thus, the real difference was in the treatments and additional services that retained students received under the policy in the retention year, which retained students under the comparison cohort would not have received.

To adjust for the effect of other reforms that might affect all students, we examined the effects on students in the P1 cohort who were not at risk. The closest set of students not at risk of retention after the 5th-grade spring assessments were those scoring in the low Level 2 range.[8] They scored 0.04 standard deviations lower than expected under the C0 cohort on the 7th-grade ELA assessment. In mathematics, low Level 2 students scored approximately 0.06 standard deviations higher than expected had they been in the control cohort. These small adjustments in the effect sizes for the promoted at-risk students had no practical impact. In either subject, there was no significant promotion policy effect on 7th-grade outcomes of high Level 1 5th-grade students who were promoted, while low Level 1 5th-grade students showed a significant decline.

It is possible that the estimated negative effects for the low Level 1 students may be attributable to unobserved covariates not accounted for in the doubly robust regressions, in addition to the small sample sizes. The predicted counterfactual for these students is obtained by weighting the C0 cohort using available data; however, the fact that these students were granted an appeal for promotion in spite of being well below the Level 2 threshold indicates the presence of individual circumstances likely not captured by our data.

To account for other programmatic influences on retained students in the P1 cohort, we examined both the differential outcomes of the not-at-risk students in their original cohort and the differential outcomes of students in the P2 cohort not in need of

---

8 For these purposes, low Level 2 is defined as within two standard errors of the Level 2 threshold score on the assessment scale.

services at the beginning of the year, relative to their original P1 cohort, since retained students joined the P2 cohort after being retained and were designated for AIS at the beginning of that year. That is, retained students are taking the 7th-grade assessment two years removed from the control cohort, and so we account for the appropriate two years' worth of other programmatic differences. In ELA, low Level 3 students[9] in the P2 cohort scored approximately 0.07 standard deviations higher (not shown) on the 7th-grade assessment than expected had they been in the prior cohort. Together with the estimated difference in the 7th-grade outcomes of not-at-risk students in the P1 cohort versus the control year shown earlier, this implies a net adjustment of 0.04 standard deviations to account for the influences of other programs concurrently in place. Thus, we estimate the retention of students at risk in ELA under the promotion policy to have an adjusted effect size of 0.57–0.63 standard deviations on 7th-grade ELA outcomes.[10]

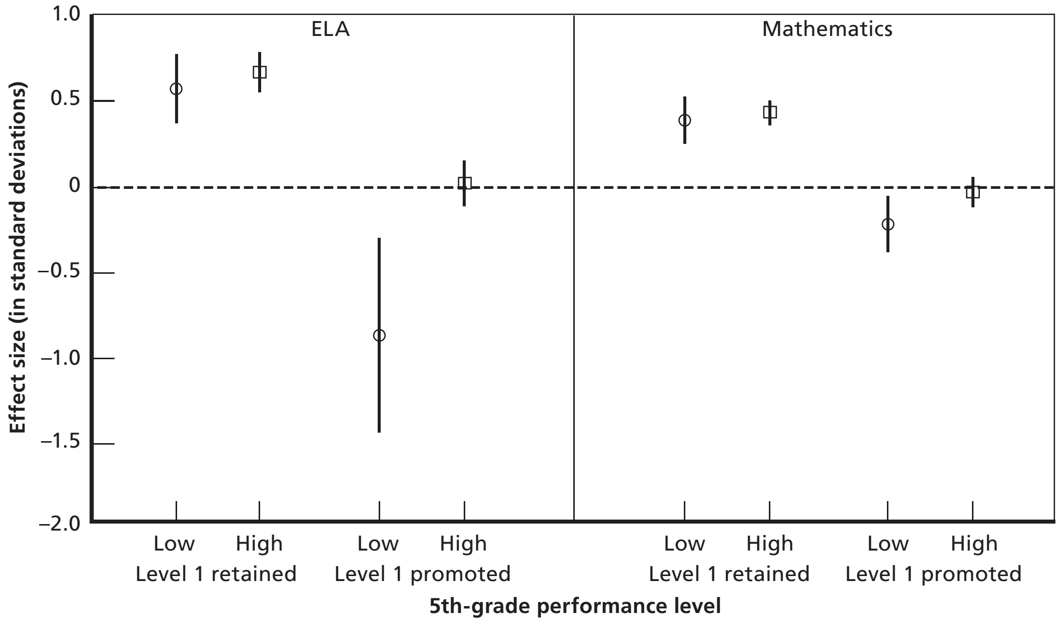
In mathematics, low Level 3 students in the P2 cohort scored approximately 0.16 standard deviations higher than expected under the prior cohort. Adding this difference to the increase of 0.06 standard deviations that students in the P1 cohort scored versus their expected results under the control year implies a net adjustment of 0.22. Our adjusted estimates of the effect of the retention of students at risk in mathematics under the promotion policy on 7th-grade mathematics outcomes are 0.38–0.42 standard deviations.

As was the case in the prior section, the estimated effects include the impact of the services received by treated students relative to what they would have received had they been in the comparison cohort through the 7th-grade assessments not already captured by our difference-in-difference adjustment (Figure 9.2). For additional context, on the 7th-grade state assessments, Level 2 scores span 1.25 standard deviations on the scale for ELA, while Level 2 mathematics scores span just under one standard deviation. For both subjects, the effect sizes imply that, based on the effects of retention alone, a 5th-grade high Level 1 student who was retained would, after an additional year of 5th-grade instruction and intervention, be expected to score well within the Level 2 range on the 7th-grade assessments, while the promoted high Level 1 student would be expected to remain a high Level 1 student not meeting the promotion standards in 7th grade.

[9] Since retained students repeat the full year of 5th grade, and Level 2 students are treated under the policy early in the year, low Level 3 students are the nearest nontreated students available for estimating the adjustment for the second year of programmatic differences between cohorts.

[10] Throughout this chapter, estimates of the effect of retention on future outcomes in both ELA and mathematics pertain to the effects of a single additional year of 5th-grade instruction and policy intervention before students proceed to later grades. The benefit of the single additional year of instruction is included in these same-grade estimates. The estimates provide no additional information about the potential effects of additional years of retention.

**Figure 9.2**
**Difference-in-Difference Estimates of the Impact of the Promotion Policy on the 7th-Grade Spring Outcomes of 2004–2005 (P1) Cohort Students at Risk of Retention in 5th Grade**




RAND *MG894-9.2*

## Disentangling the Effects of Summer Interventions and Retention on Future Outcomes: Comparing Students Near the Treatment Threshold

In Chapter Eight, we examined the effectiveness of the SSA program relative to the summer program that was in place prior to the promotion policy, using summer outcomes as the measures of interest. In this section, we use an RDD model to examine the impact of the SSA program relative to the absence of any summer intervention, by comparing the future outcomes of students just above and just below the spring Level 2 SSA assignment threshold. Concurrently, we also consider the effects of retention on the future outcomes of students just above and just below the summer Level 2 threshold. The value of considering the effects of these two treatments jointly may be seen by considering the paths to 6th grade taken by students scoring near the 5th-grade spring Level 2 threshold. Figure 3.1 in Chapter Three illustrated these paths. Here, we discuss how these influence the RDD analyses of the SSA and retention.

Consider two groups of students: group A, whose students score just above the Level 2 threshold on the 5th-grade spring ELA assessment, and group B, whose students score just below. Further, suppose that both groups score Level 2 or better on the 5th-grade spring mathematics assessment. Group A meets the promotion criteria in both subjects and is not required to attend the SSA, although some students may still attend for enrichment purposes. Group B students have not met the promotion criteria in ELA and would be enrolled in the SSA unless they scored a high Level 2 on the spring portfolio. The probability of the group B students attending the SSA is high relative to that of the group A students, indicating that a fuzzy RDD would be appropriate for evaluating the SSA treatment effect at the spring level 2 threshold. Because group A students do not take the summer assessment, the next standardized outcome measure available for both groups is the 6th-grade spring assessment.

Group B students take the summer assessment at the end of the summer. Some of the group B students will score Level 2 on the summer ELA assessment, thus meeting the promotion criteria. Of the group B students scoring Level 1 on the summer assessment, some will be promoted via the August portfolio or other appeal, while the remaining students will be retained.

The subset of group B students that is retained receives an additional treatment under the promotion policy—an additional year of 5th-grade instruction—before the 6th-grade assessment. Thus, the effect of retention on this subset of students estimated by the RDD model will be the cumulative treatment effect of the SSA on all group B students and the effect of retention on the proportion of the group B students ultimately retained in 5th grade. To isolate the effect of SSA alone, we need to "back out" a reasonable estimate for the effect of retention. For this reason, we proceed in investigating the effect of retention first, and then make use of the information learned in returning to the question of estimating the effect of the SSA.

#### The Effect of Retention on Future Outcomes: Comparing Students Near the Treatment Threshold

Students who remain at risk at the beginning of summer—those scoring below Level 2 on the 5th-grade ELA or mathematics assessments and not achieving a high Level 2 spring portfolio score—take the 5th-grade summer assessment. Students scoring Level 2 on the summer assessment meet the 5th-grade promotion standard; students scoring Level 1 on the summer assessment are retained unless they are promoted via the August portfolio or other appeal. Thus, there is a large drop in the probability of retention at the summer Level 2 threshold, and the summer assessment may be used as an RDD assignment variable in estimating the effect of retention on future outcomes for students at the Level 2 threshold.

Here, we need to consider two important points introduced earlier in the "Analytic Methods" section of this chapter. First, we consider the retention effects on students who scored Level 1 in one subject but not both. We examine the effects of retention on ELA outcomes for students at the threshold of retention in ELA and, similarly, the effects of retention on mathematics outcomes for students at the threshold of retention in that subject. Second, the treatment effect estimate produced by the RDD model is the effect of being retained into the subsequent cohort. That effect includes both a retention component and a component attributable to any programmatic differences between the original and subsequent cohort. As with the doubly robust estimates discussed earlier, the treatment effects produced by the RDD model need to be adjusted to account for programmatic differences to provide a reasonable net estimate of retention. Similarly to the doubly robust case, we use between-cohort differences for students not treated under the promotion policy as an estimate of these other programmatic differences.

Table 9.7 presents estimates from our RDD models of the effects of retention into the subsequent 5th-grade cohort on future-grade outcomes.[11,12] Sixth-grade spring assessment outcomes were available for retained students originally in the P1 and P2 5th-grade cohorts, and 7th-grade spring assessment outcomes were also available for the P1 cohort. Retained students in the P1 cohort who were just below the Level 2 cutoff on the summer assessments scored 0.43 standard deviations higher in ELA and 0.55 standard deviations higher in mathematics on the 6th-grade assessments, compared to students who were just above the Level 2 cutoff. The effects were smaller in the next cohort but still positive and significant (0.23 standard deviations in ELA and 0.43 standard deviations in mathematics). In the 7th grade, retained students from the P1 cohort continued to outperform their peers who were near the Level 2 cutoff on the summer assessments by 0.44 standard deviations in ELA and 0.60 standard deviations in mathematics.

Next, we adjust the RDD estimated retention effects in Table 9.7 to account for the effects of any programmatic differences between cohorts. We first consider the students retained from cohort P1; these results are presented in Table 9.8. To adjust the P1 retention estimates, we needed to identify year-to-year differences among nontreated students from the P1 and P2 cohorts. Mirroring the adjustments made to the doubly robust estimates of retention in the previous section, we adjusted using students scoring low Level 3 on the 4th-grade spring assessments as a baseline for comparison; these students scored, on average, about the same on the 6th-grade ELA assessment and 0.07 standard deviations higher on the 7th-grade ELA assessment than predicted if they were in the P1 cohort. This implies adjusted net estimates of the effects of retention for

---

[11] As noted earlier, to address the problem of dual-exposure threshold scores, we restricted the RDD analyses to the set of students who were above the threshold on one subject and below the threshold on the other. For the retention analyses, approximately 5–10 percent of students at the mathematics threshold were excluded, while approximately 37–51 percent of students at the ELA threshold were excluded. This disparity is due to the fact that 5th graders were much more likely to be at risk in mathematics.

[12] Consistent with other applications of RDD investigating the effects of summer instruction and retention, we do not adjust the standard errors for clustering at the school or district levels; we do not expect the treatment to vary at the level of the clusters. However, from the standpoint of independence, observations within a school or district are likely to be correlated, which could diminish the effective sample size of the observations and, consequently, the precision of the estimates. As a sensitivity check, we investigated including both district- and school-level clusters in the model. Even with their inclusion, all the estimated effects displayed in Table 9.7 remain significant; however, the precision of these estimates is affected, strongly in some cases, with standard errors increasing between 0 and 0.15 points on the standardized scale.

**Table 9.7**
**Effect of Being Retained into the Subsequent 5th-Grade Cohort on Future-Grade Spring Assessments, Students with Scale Scores Just Below the Summer Level 2 Threshold**

| | Effect of Being Retained, Including Cohort Differences[a] | | | |
|---|---|---|---|---|
| Grade and Cohort | Estimate | 95% Confidence Interval | Number of Observations Below Cutoff | Number of Observations Above Cutoff |
| 6th grade | | | | |
| ELA | | | | |
| P1 | 0.43* | (0.29, 0.57) | 124 | 328 |
| P2 | 0.23* | (0.03, 0.43) | 374 | 1,649 |
| Mathematics | | | | |
| P1 | 0.55* | (0.45, 0.64) | 1,015 | 977 |
| P2 | 0.43* | (0.35, 0.50) | 1,412 | 1,719 |
| 7th grade | | | | |
| ELA | | | | |
| P1 | 0.44* | (0.32, 0.57) | 124 | 324 |
| Mathematics | | | | |
| P1 | 0.60* | (0.50, 0.70) | 1,000 | 957 |

NOTE: The estimates are reported as effect sizes. We standardized using the theoretical standard deviation for the state assessments (40 for both subjects).

* = statistically significant at the 0.05 level.

[a] Outcomes of retained students are observed one year later.

retained students near the Level 2 threshold in ELA on the 5th-grade summer assessment of 0.43 and 0.39 standard deviations on the 6th- and 7th-grade ELA assessments, respectively.

In mathematics, there were greater year-to-year improvements between the P1 and P2 cohorts. Low Level 3 students scored, on average, 0.15 standard deviations higher on the 6th-grade mathematics assessment than predicted if they were in the P1 cohort and 0.16 standard deviations higher on the 7th-grade mathematics assessment. Thus, we calculate effect sizes of 0.37 and 0.44 standard deviations as reasonable estimates of the net effects of retention for students near the Level 2 threshold in mathematics on the 5th-grade summer assessment. While the 7th-grade estimated effect size is larger for mathematics, we note the strong overlap of the confidence intervals for the retention effect on the 6th- and 7th-grade outcomes. We find this to be evidence that the retention effect is sustained two grades later but is not necessarily increasing.

**Table 9.8**
**Effect of Retention in 5th Grade on Future-Grade Spring Assessments, Students with Scale Scores Just Below the Summer Level 2 Threshold**

| | Effect of Being Retained, Excluding Cohort Differences[a] | |
|---|---|---|
| **Grade and Cohort** | **Estimate** | **95% Confidence Interval** |
| 6th grade | | |
| P1 ELA | 0.43* | (0.30, 0.56) |
| P1 mathematics | 0.39* | (0.30, 0.49) |
| 7th grade | | |
| P1 ELA | 0.37* | (0.25, 0.49) |
| P1 mathematics | 0.44* | (0.35, 0.54) |

NOTE: The estimates are reported as effect sizes. We standardized using the theoretical standard deviation for the state assessments (40 for both subjects).

* = statistically significant at the 0.05 level.

[a] Outcome measures for retained students are observed one year later.

For the estimates of the effects of retention for the P2 cohort, we are not able to control for the potential impact of other programmatic effects in the same way as outlined earlier. Retained students from this cohort joined the P3 cohort and received AIS at the beginning of that school year. The analogous comparison to those previously made in this chapter would be the outcomes of P3 5th graders not in need of services at the beginning of the year versus similar students in the prior cohort. However, the 4th-grade assessments for these two cohorts are not equated due to the change to the new state assessment program and so are not available as a baseline for examining the differences. An exploratory examination of citywide scoring trends from 3rd through 6th grades in the appropriate years for these two 5th-grade cohorts (not shown) indicates that there is likely not much change in the relative difference in mathematics performance between these two cohorts, with the later cohort averaging roughly 0.15 standard deviations higher than the earlier cohort in both 3rd and 6th grades. ELA performance over that span shows a downward relative trend, with the later cohort again scoring about 0.15 standard deviations higher than the earlier cohort in 3rd grade, but with both cohorts averaging about the same in 6th grade. However, as seen in Figure 7.15 in Chapter Seven, this downward trend may be attributable in part to performance of students at the upper end of the scale, where there was a decrease in the percentage of Level 4 students. Thus, we expect the effect directly attributable to retention for the P2 cohort to be comparable to those reported in Table 9.3, and possibly slightly higher than reported in that table for ELA.

### The Effect of SSA on Future Outcomes: Comparing Students Near the Treatment Threshold

We now return to the effects of the SSA on students at the 5th-grade spring assessment Level 2 threshold in order to disaggregate SSA effects from the cumulative effects of SSA and retention provided by the RDD model. Note that, theoretically, we cannot simply use the retention effects identified from the RDD models to isolate the effects of SSA from the cumulative effects of SSA and retention. This is because the RDD model for SSA assignment estimates treatment effects for students just below the Level 2 threshold on the *spring* 5th-grade assessment, while the RDD model for retention estimates treatment effects for students just below the Level 2 threshold on the *summer* assessment. As Jacob and Lefgren (2002) point out, it would be reasonable to believe that the students around the summer Level 1 cut-point are of lower proficiency than those around the spring Level 1 cut-point. Thus, a simple subtraction of the appropriate terms might not apply. However, we can use what is learned about the affected students to suggest a reasonable estimate for the marginal effect of the SSA.

An anomaly due to the shift from city to state testing in spring 2006 also allows for a more direct isolation of the effect of the SSA on future outcomes. Level score cut-points for the new state assessment were not available for the P2 cohort at the end of the 2005–2006 school year, so the city assigned interim level scores in order to carry out SSA assignment under the policy. The interim Level 2 thresholds in both subjects turned out to be just slightly above the true Level 2 threshold, which became known at the end of the summer of that year. Thus, the SSA treatment effects estimated from the RDD model apply to the lowest Level 2 students—more than 98 percent of whom were promoted—instead of the highest Level 1 students, as would have been the case if the true level scores had been available in June. Thus, while RDD estimates for the P1 cohort need to be adjusted for retention, we may essentially view the estimated SSA effect under the RDD model for the P2 cohort as a true effect for SSA, albeit for very low Level 2 students instead of very high Level 1 students.

In consideration of the needed adjustment to account for retention in the cumulative SSA assignment estimates for the P1 cohort, we explored the spring scores of the summer Level 1 students at the Level 2 threshold on the summer assessment. Of those students obtaining the highest observable ELA Level 1 scale score on the 2005 summer assessment, more than 50 percent had scored within 0.25 standard deviations of the Level 2 threshold on the spring assessment, and more than 95 percent had scores within 0.5 standard deviations of Level 2 in the spring. For mathematics, more than 50 percent of those students obtaining the highest observable Level 1 mathematics scale score on the 2005 summer assessment had scored within 0.3 standard deviations of Level 2 in the spring, and more than 85 percent had scored within 0.7 standard deviations of Level 2 in the spring. Thus, although most of the Level 1 students nearest the Level 2 threshold in the summer were not also at the threshold in the spring, the overwhelming majority of these students were in the upper half of the high

Level 1 range in the spring. Unless there is a drastic drop-off in the SSA effect for students just below the spring Level 2 threshold—exploratory investigations of the data suggest that this is not the case—the estimated retention effects from the RDD models for summer Level 2 threshold students should be a reasonable estimate for the retention effects for spring Level 2 threshold students, and we proceed accordingly. This would be true for both 6th- and 7th-grade outcomes, since the same assignment variable applies to both.

Table 9.9 presents the estimated effects of the SSA,[13,14] relative to no additional summer instruction, for the P1 and P2 cohorts.[15] We again consider 6th- and 7th-grade outcomes. The upper portion of the table shows the RDD estimates for the P1 cohort that represent the cumulative effects of the SSA for all students just below the Level 2 threshold on the spring assessment and retention for those students just below the Level 2 threshold on the spring assessment who were eventually retained. The lower portion of the table shows the estimated marginal effect of being mandated to SSA, including adjusted estimates for the P1 cohort and estimates directly from the RDD model for the P2 cohort.

The effect of being mandated to SSA has a significant impact on 6th-grade outcomes in both subjects. That is, compared to students who were just above the true or interim Level 2 cutoffs, respectively, students who scored just below the cutoff and were mandated to attend summer school experienced an average increase of 0.09–0.12 standard deviations on their 6th-grade ELA and mathematics assessments due to SSA alone.

For the P1 cohort, the estimated effect is slightly larger in the 7th grade than in the 6th grade: 0.15 versus 0.10 in ELA, and 0.14 versus 0.11 in mathematics. However, as evidenced by the substantial overlap in the confidence intervals, this is evidence that the effects are sustaining for two to three years beyond the SSA, but not necessarily increasing.

---

[13] Restricting the RDD analyses to the set of students who were above the threshold on one subject and below the threshold on the other caused approximately 9–12 percent of students at the mathematics threshold and approximately 23–45 percent of students at the ELA threshold to be excluded. This disparity is due to the fact that 5th graders were much more likely to be at risk in mathematics.

[14] As with the RDD retention effects presented in Table 9.7, for sensitivity, we investigated the implications of including district- and school-level clustering in the RDD models. With their inclusion, those effects that are significant in Table 9.9 remain so; precision of the estimated effects is affected, however, with standard errors increasing from 0 to 0.03 points on the standardized scale.

[15] SSA attendance data were not available for the summer of 2005. For spring Level 1 students in the P1 cohort, we used the presence of at least one summer test score among students not promoted via the spring portfolio as a proxy for SSA attendees.

**Table 9.9**
**Effect of Being Mandated to SSA on Future-Grade Spring Assessments, Students with Scale Scores Just Below the Spring Level 2 Threshold**

| Grade and Cohort | Estimate | 95% Confidence Interval | Number of Observations Below Cutoff | Number of Observations Above Cutoff |
|---|---|---|---|---|
| 6th grade | | | | |
| ELA | | | | |
| P1 | 0.10 | (–0.01, 0.21) | 579 | 41,719 |
| P2 | 0.09* | (0.05, 0.13) | 2,308 | 27,681 |
| Mathematics | | | | |
| P1 | 0.11* | (0.05, 0.16) | 3,146 | 36,536 |
| P2 | 0.12* | (0.06, 0.18) | 3,221 | 26,147 |
| 7th grade | | | | |
| ELA | | | | |
| P1 | 0.15* | (0.06, 0.24) | 579 | 41,917 |
| Mathematics | | | | |
| P1 | 0.14* | (0.09, 0.20) | 3,111 | 36,660 |

NOTE: The estimates are reported as effect sizes. We standardized using the theoretical standard deviation for the state assessments (40 for both subjects). SSA effect estimates from the RDD model for the P1 cohort have been adjusted to account for the retention effects of marginal Level 1 students who were ultimately retained. SSA effect estimates from the RDD model for the P2 cohort did not require adjustment for retention effects and are based on an assignment variable just above the minimum Level 2 score.

* = statistically significant at the 0.05 level.

## Summary

In this chapter, we used two quasi-experimental methods to investigate the effects of AIS, Saturday and summer schools (for some students), and retention (for some students) on the future outcomes of students who were promoted or retained in grade once.

### Seventh-Grade Outcomes for Students Identified for Services in the 5th Grade, P1 Cohort

Using the comparison cohort as a control, we examined the 7th-grade outcomes of students needing services in the beginning of 5th grade in the first policy cohort. After adjusting for the potential impact of other reforms simultaneously in place, we estimated that, for students needing services in 5th grade, the effects on 7th-grade outcomes directly attributable to the promotion policy are 0.08–0.21 standard deviations for high Level 1 and low Level 2 students in ELA and 0.09–0.16 standard deviations for high Level 1 and low Level 2 students in mathematics.

### Seventh-Grade Outcomes for Students at Risk of Retention, P1 Cohort

Using the comparison cohort as a control, we also examined the 7th-grade outcomes of students who were identified as at risk of retention on the 5th-grade spring assessments, as well as those who were retained. After adjusting for the influences of concurrent reforms, we estimated that the effects of retention on 7th-grade outcomes directly attributable to the promotion policy are 0.57–0.63 standard deviations in ELA and 0.38–0.42 standard deviations in mathematics for all Level 1 retained students.

### Sixth- and 7th-Grade Outcomes for Students Attending SSA and Retained, P1 and P2 Cohorts

The effect of being mandated to SSA has a small but significant impact on 6th-grade outcomes in both subjects. That is, compared to students who were just above the Level 2 cutoff, students in the P1 cohort who scored just below the cutoff on the 5th-grade spring assessment and were mandated to attend summer school had scores that were about 0.10 standard deviations higher in ELA and mathematics in the 6th grade and slightly higher in the 7th grade (0.15 standard deviations).

The effects for retained students are larger than for the set of students mandated to summer school and then promoted. Retained students in the P1 cohort who were just below the Level 2 cutoff on the summer assessments scored roughly 0.40 standard deviations higher in ELA and mathematics on the 6th- and 7th-grade assessments, compared to students who were just above the Level 2 cutoff (though these students took the assessments in different years). The 6th-grade results for the P2 cohort were smaller in ELA (0.23 standard deviations).

We now turn to results from our student surveys on the socioemotional status of at-risk and not-at-risk students.

CHAPTER TEN

# The Impact of New York City's Promotion Policy on Students' Socioemotional Status

*Vi-Nhuan Le, Louis T. Mariano, and Al Crego*

One of the key concerns regarding the use of grade retention as an intervention is the potential impact on students' socioemotional outcomes. This chapter examines the social and emotional outcomes of three groups of students held to the promotion policy: (1) students who were at risk of retention and retained, (2) students who were at risk of retention but promoted (hereafter referred to as "at-risk promoted" students), and (3) students who were not at risk of retention and promoted (hereafter referred to as "not-at-risk" students). We surveyed students regarding three measures of socioemotional functioning—namely, school belonging, mathematics confidence, and reading confidence—over a three-year period, following cohorts of students into subsequent grades. We used the survey responses to address the following research questions:

1. Are there differences among retained, at-risk promoted, and not-at-risk students with respect to school belonging or self-confidence in mathematics and reading? Do these differences change over time?
2. After controlling for student-level characteristics, is retention status related to students' later socioemotional outcomes?
3. Among at-risk students, are socioemotional responses related to final promotion outcomes?
4. Are students' socioemotional responses predictive of future retention status?

Chapter Four described the survey methodology, including the sampling procedures, response rates, and survey weighting procedures. We focus here on the findings from the three student surveys we administered—in particular, the differences in the affective outcomes of retained students relative to nonretained students. This chapter provides details of our results for the 5th-grade cohorts. We also conducted these analyses for the two cohorts of 3rd-grade students (most of whom are also members of the surveyed 5th-grade cohorts). The details of those analyses can be found in Appendix E.

## Survey Measures

As mentioned earlier, in developing the survey, we chose socioemotional measures with a strong theoretical basis that had been previously used in large-scale research and did not overburden respondents. The three measures included in the survey are discussed next.

### School Belonging Scale

We administered an adapted version of Goodenow's Psychological Sense of School Membership to assess students' perception of school belonging. With the input of NYCDOE, we modified the wording of some items to make them more appropriate for the context of this study. The 18-item scale was intended to capture a student's sense of school belonging and included items about students' perceptions in the following areas:

- Personal acceptance:
  - I feel very different from most other students here.
  - I can really be myself at school.
  - I wish I were in a different school.
- Inclusion:
  - I feel like a real part of this school.
  - It is hard for me to be accepted here.
  - Sometimes I feel as if I don't belong here.
  - I am included in lots of activities at this school.
  - I feel proud of belonging in this school.
- Their relationships with their peers and teachers:
  - People here notice when I'm good at something.
  - Other students in this school take my opinions seriously.
  - Most teachers here are interested in me.
  - There is at least one teacher or adult in this school I can talk to if I have a problem.
  - People at this school are friendly to me.
  - Sometimes I feel that people here are not interested me.
  - I am treated with as much respect as other students.
  - The teachers and principals here respect me.
  - People here know I can do good work.
  - Other students here like me the way I am.

To keep the response categories consistent with the other measures included on the survey, we also modified the response scale from its original five-category scale to a four-category scale: "disagree a lot," "disagree a little," "agree a little," and "agree a lot."

### Mathematics Confidence Scale

We administered a four-item scale assessing students' self-confidence in mathematics. The scale had previously been used in the 2003 TIMSS (Martin, Mullis, and Chrostowski, 2004), and included items about students' perceptions of how quickly they learned mathematics and whether they performed well in that subject:

1. I usually do well in mathematics.
2. Mathematics is harder for me than for many of my classmates.
3. I am just not good at mathematics.
4. I learn things quickly in mathematics.

Students were asked to respond using a four-category response scale ranging from "disagree a lot" to "agree a lot."

### Reading Confidence Scale

To assess reading confidence, we used a measure that was analogous to the mathematics confidence scale and had also been used in the 2003 TIMSS:

1. I usually do well in reading.
2. Reading is harder for me than for many of my classmates.
3. I am just not good at reading.
4. I learn things quickly in reading.

## Timing of Surveys and Total Number of Responses

The analysis reported in this chapter focused on comparing the affective outcomes of retained, at-risk promoted, and not-at-risk students in two 5th-grade cohorts: the 2005–2006 and 2006–2007 5th graders.[1] We administered surveys to these cohorts during the 2006–2007, 2007–2008, and 2008–2009 academic years. Table 10.1 shows the timing of the survey administrations relative to the retention decision.

**Table 10.1**
**Student Surveys, by Number of Academic Years After the Retention Decision**

| Cohort | 2006–2007 | 2007–2008 | 2008–2009 |
|---|---|---|---|
| 2006–2007 5th graders | 0 | 1 | 2 |
| 2005–2006 5th graders | 1 | 2 | 3 |

[1] We also surveyed two 3rd-grade cohorts. These results are presented in Appendix E.

For the 2005–2006 5th-grade cohort, we obtained survey responses only after the retention decision was made. We used this cohort to examine relationships between retention and socioemotional outcomes over the short term (i.e., one to three years after the retention decision). For the 2006–2007 5th-grade cohort, we obtained survey responses prior to students being identified as at risk of retention, allowing us to examine socioemotional responses before and after the retention decision was made. We presented overall and adjusted response rates in Chapter Four, but we briefly repeat the discussion here. Because we administered the survey to complete classrooms in many cases to avoid burdens on the schools, our survey administration included some ineligible respondents. In addition, some of the returned surveys were only partially complete. We discarded surveys that did not have answers on at least 75 percent of the items in each of the three scales. The number of eligible usable surveys and the adjusted response rates across the three surveys were as follows:

- 2006–2007 student survey: 7,982 (55 percent)
- 2007–2008 student survey: 7,260 (41 percent)
- 2008–2009 student survey: 8,769 (41 percent).

## Analytic Methods

### Survey Weighting

We used post-stratification weighting of the survey responses to ensure that the actual distribution of at-risk students across schools was properly represented in the survey analyses. For each of the four survey cohorts, a set of weights was developed for each subgroup: at-risk retained, at-risk promoted, and not at risk. Because all eligible students in a sampled school were included in the sample, all responding students in a school within the same subgroup were assigned the same weight.

Table 10.2 shows the final sample sizes for each subgroup, by cohort and survey year. The table shows the number of students who were both subjected to the retention policy and members of the original cohort being studied (i.e., they enrolled in the NYC school system before the retention decision was made). With the exception of the 2006–2007 cohort during the second survey year,[2] the sample sizes were sufficiently large that we could reasonably make inferences to the full subgroup population.

---

[2] In the first two years, because of a delay in the receipt of data and the planned fall survey administrations, complete school population data were not available in time for sampling. Therefore, each school's population with respect to each subcohort to be sampled was estimated using the data from the prior year. Because of the unanticipated sharp decline in the number of retained students in the 2006–2007 5th-grade cohort, the distribution of students retained from the 2005–2006 5th-grade cohort was not a good estimate for the distribution of students retained in the 2006–2007 5th-grade cohort. This caused a serious shortfall in the number of retained students from that cohort represented in the sample in the second survey year. Additional detail regarding survey design and sampling may be found in Chapter Four.

**Table 10.2**
**Sample Sizes for Each Subgroup, by Cohort and Survey Year**

| Cohort | 2006–2007 | 2007–2008 | 2008–2009 |
|---|---|---|---|
| 2006–2007 5th graders | | | |
| Not at risk | 3,104 | 4,199 | 4,452 |
| At-risk promoted | 179 | 163 | 150 |
| Retained | 67 | 10 | 122 |
| 2005–2006 5th graders | | | |
| Not at risk | 3,328 | 1,695 | 2,696 |
| At-risk promoted | 249 | 183 | 160 |
| Retained | 189 | 169 | 152 |

## Imputation

Given that the survey was voluntary, students were informed in the survey instructions that they could skip any question that they did not wish to answer. About 15 percent of respondents skipped one or more items. To maximize the number of responses, we imputed the values of skipped items for students who answered at least 75 percent of the questions in each of the three scales.

To impute the missing values, we first fit a separate polytomous item response theory (IRT) model to the item responses from each survey scale. For each missing response, we then used the fitted IRT model to estimate the probabilities that the student would have responded in each of the four available response categories. The response category with the largest probability of response was then chosen as the imputed value. For consistency across years, the IRT item parameters were estimated using the year 1 survey responses and then held fixed at those values for the year 2 and year 3 imputations.

## Estimating Survey Scale Scores

For the analyses shown here, survey responses were coded by mapping the response scale ("disagree a lot," "disagree a little," "agree a little," and "agree a lot") to integer values from 1 to 4.[3] This coding is in accordance with previous uses of the three scales included in the survey instrument.

For each eligible respondent, we produced scale scores for each of the three survey scales by averaging the integer values corresponding to their responses.[4] Using the aver-

---

3 For consistency, responses to negatively phrased items were reverse-coded before mapping to the integer scale.

4 The mapping of the response categories to integer values 1–4 carries an implied assumption of equal distance between adjacent response categories. An alternative to using average scale scores that would avoid this assump-

age, as opposed to a sum, allows interpretation of the scale score in terms of the original response scale. For example, a value of 3.0 on the original response scale corresponds to mild agreement (i.e., "agree a little"). Thus, an average scale score of 3.0 would correspond to mild agreement with the qualities of the construct represented in the scale.

All three survey scales demonstrated sufficient reliability across the three waves of the survey. Cronbach's alpha coefficients for the school belonging, reading confidence, and mathematics confidence scales were 0.83, 0.74, and 0.82, respectively, in the first wave of the survey. They were comparable in the second and third waves as well (0.84, 0.74, and 0.83, respectively, in the second wave; 0.86, 0.75, and 0.83, respectively, in the third wave).

### Multiple Statistical Tests

We conducted pairwise comparisons for each subgroup and survey year, resulting in 200 statistical tests to address certain research questions. As mentioned earlier, we adjusted for a 10-percent false discovery rate by using a p-value of less than 0.04 as the criterion for statistical significance. This standard was applied to all of the results reported here.

To aid in the interpretation of the results, for cases in which two subgroups had significantly different mean scores on a survey scale,[5] we also include an estimated effect size. Effect sizes were calculated by dividing the magnitude of the difference in means between the two subgroups by the standard deviation of the pooled standard deviation of all three subgroups. As in previous chapters, following the guidelines set forth by Cohen (1988), effect sizes, *d*, of about 0.2 are considered "small," effect sizes of about 0.5 are considered "medium," and effect sizes of about 0.8 are considered "large."

### Regression Models

We conducted weighted least-squares linear regression or logistic regression, where appropriate, and adjusted the standard errors to account for grade-range stratification and clustering at the school level. We also standardized the survey scales so that the reported coefficient is the expected differences between two groups for an increase of one standard deviation unit in scale scores.

We explored models that included both student- and school-level covariates. At the individual student level, we controlled for gender, eligibility for free or reduced-price

---

tion would be to use the student latent scale parameters from the IRT models used in the imputation of missing responses. We compared the IRT estimates to the survey scale averages and found that they had a correlation of over 97 percent on all three scales.

[5] All tests for differences in means were conducted as two-sided t-tests.

lunche, special education status, number of days absent from school, and ethnicity.[6] Because we were most interested in understanding how socioemotional responses are related to retention status, much of our discussion focuses on differences in the affective outcomes among the retained, at-risk promoted, and not-at-risk students. Readers interested in the relationship between student characteristics and retention status are referred to other studies (e.g., Dauber, Alexander, and Entwisle, 1993; Jimerson, Carlson, et al., 1997; Meisels and Liaw, 1993; Owings and Magliaro, 1998).

At the school level, we controlled for school configuration, NCLB status, school size, and student composition variables, including the percentage of students in each racial/ethnic category, the percentage eligible for free or reduced-price lunch, the percentage of males, and the percentage of immigrants. We included the school-level variables because previous studies have shown that school context factors, such as school size and grade configuration, can be related to students' socioemotional responses, particularly their sense of school belonging (Klonsky, 2002; Lee, Ready, and Welner, 2002). However, in our analysis, few of the school-level variables were significant, and none showed consistent patterns across different cohorts or years. In the interest of parsimony, we dropped the school-level variables from our models, so the regression results presented here do not control for school-level covariates.

## Student Survey Results

### Distributional Differences in Socioemotional Responses Among Retained, At-Risk Promoted, and Not-At-Risk Students

To understand how retention is related to students' school attitudes, we compared the retained, at-risk promoted, and not-at-risk students' perceptions with respect to school belonging and self-confidence in mathematics and reading. If being retained negatively affects students' socioemotional well-being, we would expect retained students to show lower perceptions of school belonging and express less confidence in their mathematics or reading abilities than nonretained students.

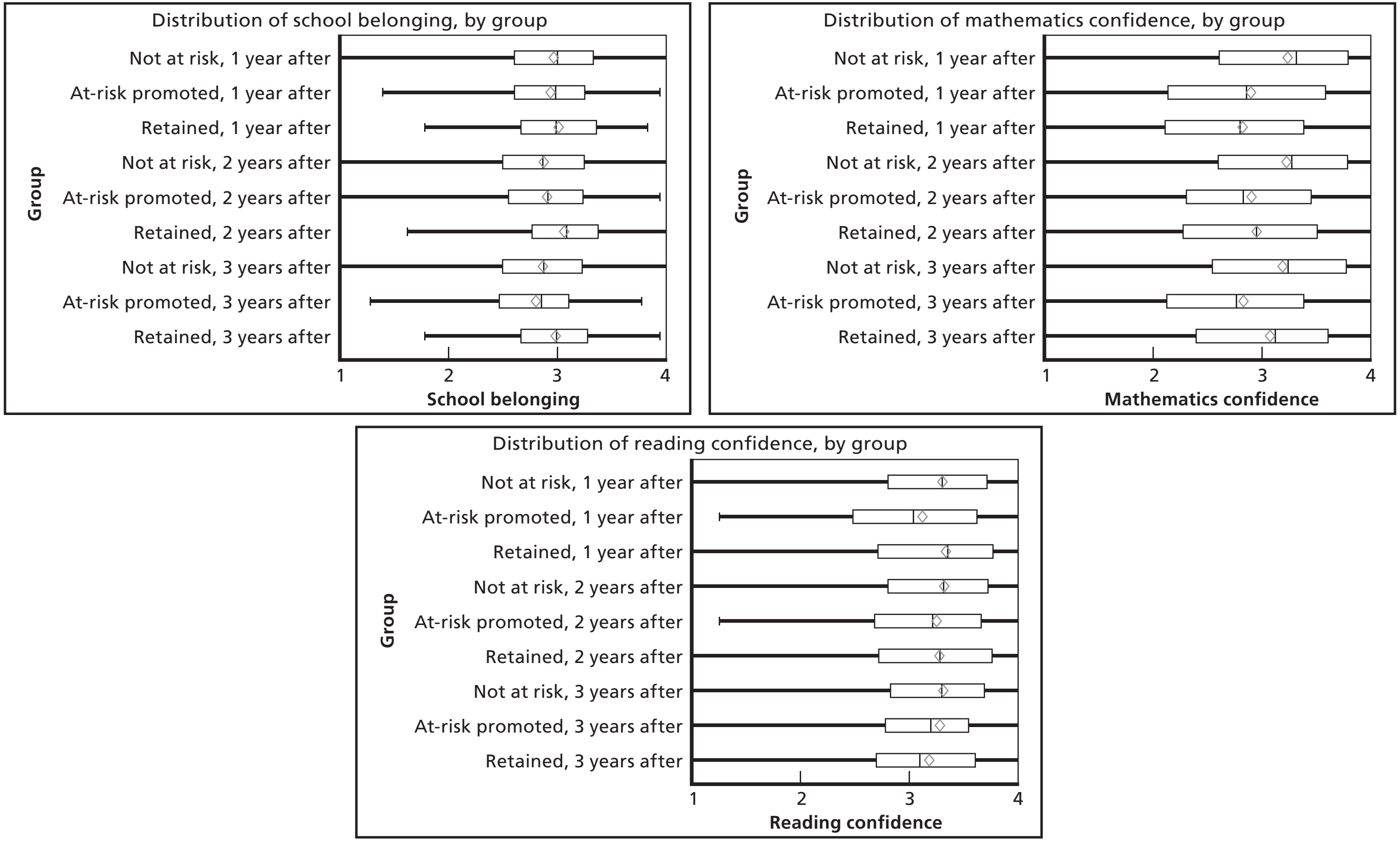
Figure 10.1 shows the distributions of socioemotional outcomes for students who were one, two, or three years after their 5th-grade retention decision.[7] (The associated descriptive statistics are provided in Appendix D). There were no differences between

---

[6] Because we received data early during the 2008–2009 academic year, the number of days absent from school was not yet available. Thus, we averaged the number of days absent from the two previous school years and used that figure as a proxy for absences during the 2008–2009 year.

[7] To explore the feasibility of combining different cohorts, we examined differences in their distributions. For example, we explored whether the responses of the 2005–2006 cohort in the 2006–2007 survey year could be combined with the responses of the 2006–2007 cohort in the 2007–2008 survey year (i.e., when both groups were one year removed from the retention decision). There were few significant differences between any of the cohorts across the years, and for all but one comparison, the effect sizes did not exceed 0.30 in magnitude. Given the small differences, we combined the cohorts across years for our analyses.

**Figure 10.1**
**Students' Socioemotional Responses One, Two, and Three Years After the Retention Decision**




**RAND** *MG894-10.1*

the three groups one year after the retention decision, but retained students expressed higher levels of school belonging than both the at-risk promoted students and not-at-risk students two and three years after the retention decision. However, the mean differences in favor of the retained students were small, with effect sizes ranging from 0.08 to 0.13.

A more complex picture emerged when we considered mathematics attitudes. Both retained and at-risk promoted students were significantly less confident in their mathematics skills than the not-at-risk students for the first two years after the retention decision, although the mean differences were small (i.e., the effect sizes ranged from 0.15 to 0.32). By the third year, there were no differences in the mathematics attitudes of the retained and not-at-risk groups. In contrast, three years after the retention decision, the at-risk promoted students continued to express lower levels of mathematics confidence than the not-at-risk students ($d$ = 0.18) and demonstrated significantly poorer attitudes toward mathematics than the retained students ($d$ = 0.12). An examination of mean differences in school belonging across the survey years suggested that these findings were due mostly to gains in the retained students' confidence over time ($d$ = 0.27), as opposed to declines in confidence among the not-at-risk or at-risk promoted students.

With respect to reading attitudes, retained students were comparable to the not-at-risk students, and both groups were significantly more confident than the at-risk promoted students one year after the retention decision ($d$ = 0.14 and 0.12 in favor of the retained and not-at-risk students, respectively). Two years after the retention decision, the at-risk promoted students had closed the gap with both groups such that all three groups were comparable in terms of reading attitudes. By the third year after the retention decision, the not-at-risk students expressed more reading confidence than the retained students ($d$ = 0.11), but there were no significant differences between the not-at-risk students and the at-risk promoted students, nor were there differences between the at-risk promoted and retained groups.

Overall, the not-at-risk students remained stable in their reading confidence across years, whereas the at-risk promoted students experienced a slight increase in reading confidence, and the retained students experienced a slight decrease.

### Relationship Between Grade Levels, Promotion Status, and Socioemotional Outcomes

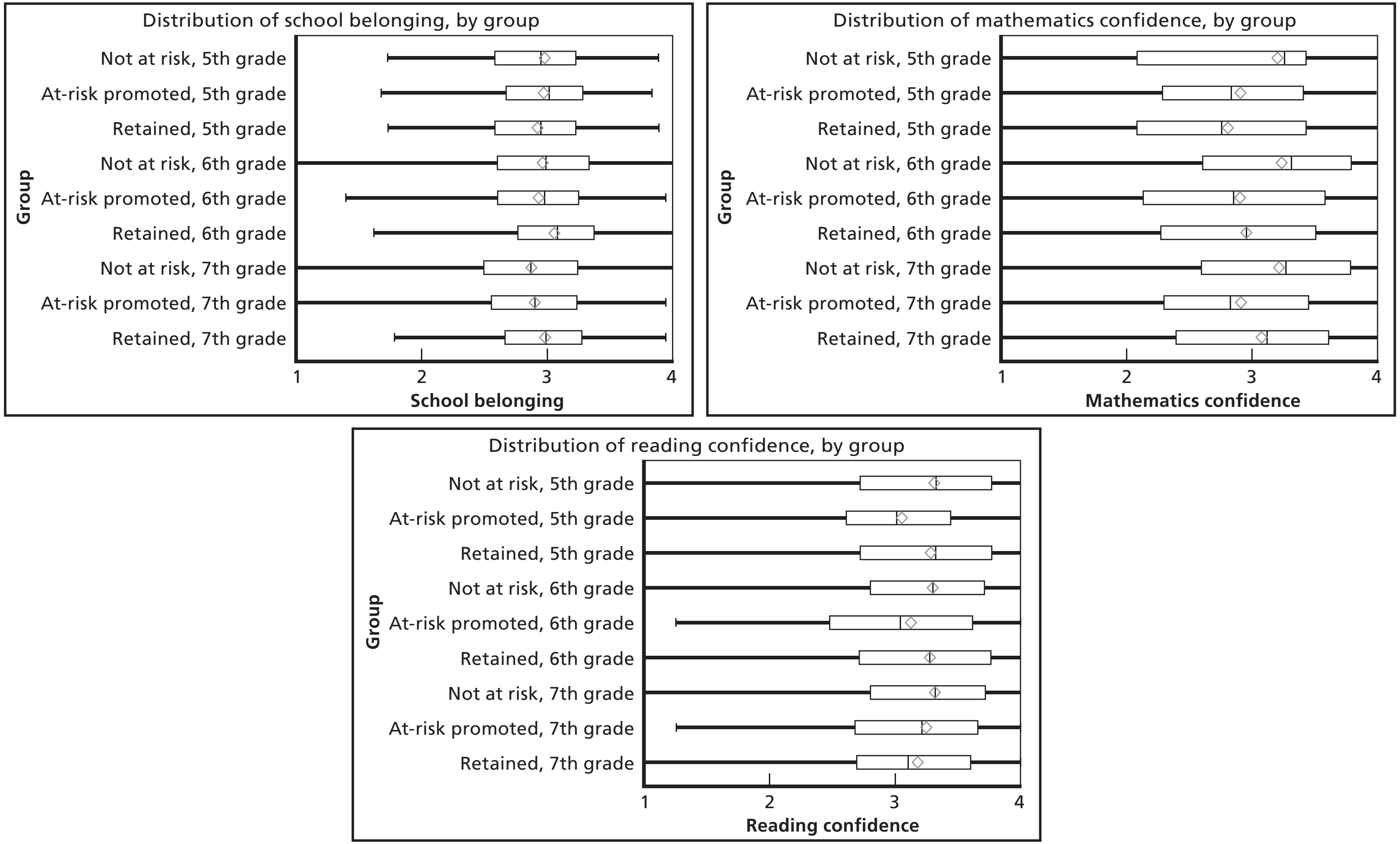
Of particular interest in our analysis were students' socioemotional outcomes during the middle-grade years. Many studies have found that problem behaviors (e.g., feelings of alienation, anxiety, and conduct disorders) tend to increase during the middle school years (Alexander, Entwisle, and Kabanni, 2001; Finn, 1993; Kazdin, 1993), partly because of a potential mismatch between students' developmental needs and the school environment (Carnegie Council on Adolescent Development, 1989; Eccles and Midgley, 1989). At the onset of puberty, students undergo a host of psychological, emotional, and cognitive changes. At the same time, many students are also experiencing a change in school environment as they transition from an elementary school to a junior high school (i.e., serving grades 7–8) or middle school (i.e., serving grades 6–8). It has been hypothesized that the junior high/middle school's structure and instructional practices do not support the needs of adolescents. For example, the junior high/middle schools' departmentalization of subjects is thought to be at odds with adolescents' desire for more personal relationships (Juvonen et al., 2004). Indeed, there is some support for a developmental mismatch; 7th-grade students enrolled in junior high schools reported having lower self-esteem, poorer school attitudes, and lower grades than 7th-grade students enrolled in schools with a K–8 configuration (Simmons and Blyth, 1987).

These findings have implications for the 5th-grade students in NYC, the majority of whom were enrolled in middle schools at the time the second and third surveys were fielded. Namely, a developmental mismatch between students' needs and learning environment would suggest an observed decrease in positive school attitudes over time, especially from 5th to 6th grade, when the transition to a middle school occurs. However, as shown in Figure 10.2, there was not a decrease in students' belonging or confidence, and the at-risk promoted and retained students even showed an increase in school attitudes over time (Appendix D provides the associated descriptive statistics). Retained students were more connected to their school in 6th grade than in 5th grade ($d = 0.26$) and expressed more confidence in their mathematics ability in 7th grade than in 5th grade ($d = 0.29$). Similarly, at-risk promoted students were more confident in their reading ability in 7th grade than in 5th grade ($d = 0.22$). In contrast, the not-at-risk students expressed similar levels of school attitudes across the three grade levels. It is important to note that, relative to the not-at-risk students, the retained students expressed significantly lower mathematics confidence in 5th grade ($d = 0.30$), while the at-risk promoted students expressed significantly lower confidence in both subjects ($d = 0.23$ and 0.27 in mathematics and reading, respectively). Thus, the gains in the at-risk students' mathematics and reading confidence over time led to their "catching up" to the not-at-risk students, rather than to a surpassing of the not-at-risk students with respect to perceived competence.

#### Relationship Between Retention and Students' Later Socioemotional Outcomes

While the descriptive statistics provide a useful preliminary indication of students' socioemotional responses in relation to their retention status over time, the analyses did not take into account student-level characteristics. It is important to control for potential changes in student demographics over time because we may observe trends in students' socioemotional outcomes that are unrelated to their retention status (Jacob, Stone, and Roderick, 2004). For example, suppose that, relative to other racial/ethnic

**Figure 10.2**
**Students' Socioemotional Responses, by Retention Decision and Grade Level**




RAND *MG894-10.2*

groups, students in one particular group were more likely to be retained, to feel connected to school, and to leave the NYC school after being retained. Regardless of whether retention negatively affects students' perceptions of school belonging, we would expect to observe a decrease in retained students' school connectedness over time due to the changing racial composition of the sample (i.e., fewer students from that particular group). By controlling for student demographics, we can better estimate the relationships between retention and socioemotional responses.

In this section, we analyze how prior retention status is related to later socioemotional outcomes, after controlling for student demographics. As indicated earlier, we use weighted least-squares linear regression models and adjusted standard errors to account for the survey design. For the outcomes discussed here, the response variables are the respective student survey scale scores, and the student subgroup (retained, at risk promoted, or not at risk), along with the available student-level characteristics included as explanatory variables. Table 10.3 provides the standardized regression coefficients for the not-at-risk and at-risk promoted subgroups for the 5th-grade cohort one to three years after the retention decision (see Appendix D for the regression estimates for the remaining variables). Because the student subgroup is categorical, we set "retained" as the reference category; the regression coefficients presented show how the average scale scores of the other two groups differ from retained students. The regression results corroborated most of the unadjusted mean-difference analyses presented earlier. For example, even after accounting for student demographics, retained students showed

**Table 10.3**
**Standardized Regression Coefficients One to Three Years After the Retention Decision**

| Variable | 1 Year | 2 Years | 3 Years |
|---|---|---|---|
| School belonging | | | |
| At-risk promoted | –0.09 (0.13) | –0.29 (0.09)** | –0.34 (0.11)** |
| Not at risk | –0.06 (0.09) | –0.36 (0.08)** | –0.23 (0.09)** |
| Mathematics confidence | | | |
| At-risk promoted | –0.04 (0.14) | –0.06 (0.10) | –0.33 (0.12)** |
| Not at risk | 0.39 (0.11)** | 0.30 (0.08)** | 0.09 (0.09) |
| Reading confidence | | | |
| At-risk promoted | –0.47 (0.13)** | –0.04 (0.11) | 0.10 (0.10) |
| Not at risk | –0.07 (0.49) | 0.02 (0.09) | 0.14 (0.09) |

NOTE: Standard errors are in parentheses.

* = statistically significant at the 0.04 level.

** = statistically significant at the 0.01 level.

advantages over both at-risk promoted and not-at-risk students with respect to school belonging. There may be several reasons that retained students felt a greater sense of school belonging than the nonretained students. Retained students may have received additional supports and services not offered to the nonretained students, which might have fostered greater attachment to the school. The extra year may have also offered retained students additional opportunities to participate in school-based activities. Indeed, Jacob and Stone (2005) reported a large increase in after-school participation by at-risk students after the implementation of a grade retention policy in Chicago. Finally, because retained students were learning alongside younger students, they may have enjoyed a higher social status among their peers (Gottfredson, Fink, and Graham, 1994).

The regression analyses also confirmed the results of the descriptive analyses with respect to mathematics confidence. Namely, the not-at-risk students held an advantage in mathematics confidence over the retained students for the first two years after the retention decision, but by the third year, both groups expressed similar levels of mathematics confidence.

There were, however, some differences in results after adjusting for student demographics. After controlling for student-level characteristics, the not-at-risk and retained students had comparable levels of reading confidence three years after the retention decision. This was a departure from the earlier descriptive analysis, in which the not-at-risk students had expressed slightly more positive attitudes about reading than did the retained students three years after the retention decision.

### Relationship Between Promotion Criteria and At-Risk Students' Socioemotional Outcomes

Arguably, the at-risk promoted students represent the best comparison group in terms of evaluating the effect of the retention policy on the retained students. Both groups were at risk of retention, but the at-risk promoted students had managed to successfully meet the promotion criteria (or successfully secure an appeal). It is important to note that, within the at-risk promoted group, there were different "types" of students, some of whom could be considered better matches to the retained students than others. For example, students who successfully passed the spring portfolio and were not required to attend summer school may have had different levels of socioemotional adjustment from students who were mandated to attend summer school and subsequently met the promotion criteria by passing the summer assessment or the August portfolio. In turn, both of these groups of students can be contrasted with the students who were promoted only after they appealed their nonsuccessful summer performance. In terms of a continuum, we might expect the students who appealed their retention decision to be most similar to the retained students with respect to socioemotional adjustment, whereas the students who passed the spring portfolio may be least similar.

We explored this possibility by further comparing the socioemotional responses of the retained students to those of the at-risk students who met the promotion criteria through the spring portfolio, their summer performance (i.e., passing either the summer assessments or the August portfolio), or an appeal to district officials. The regression coefficients relating socioemotional outcomes to promotion type are provided in Table 10.4 (see Appendix D for the regression coefficients for the covariates). The regression model was similar to those used earlier, except that the possible subgroups were altered to include retained students (still the reference category) and the three at-risk promotion types. The differences in school membership reported earlier between at-risk promoted and retained students were significant only for the subset of students who had been mandated to attend summer school. The regression coefficients indicated that students who were promoted via the spring portfolio had a lower sense of school belonging than retained students two years out, but the differences were not statistically significant.[8]

**Table 10.4**
**Standardized Regression Coefficients, by Promotion Type, One to Two Years After the Retention Decision**

| Variable | 1 Year | 2 Years |
|---|---|---|
| School belonging | | |
| Spring portfolio | −0.40 (0.25) | −0.29 (0.28) |
| Summer performance | −0.03 (0.14) | −0.47 (0.20)* |
| Appeal | 0.02 (0.18) | −0.58 (0.22)* |
| Mathematics confidence | | |
| Spring portfolio | −0.06 (0.16) | 0.67 (0.23)** |
| Summer performance | −0.06 (0.14) | 0.14 (0.17) |
| Appeal | −0.11 (0.19) | −0.11 (0.24) |
| Reading confidence | | |
| Spring portfolio | −0.44 (0.17)* | 0.24 (0.20) |
| Summer performance | −0.49 (0.15)** | 0.06 (0.17) |
| Appeal | −0.18 (0.13) | −0.02 (0.23) |

NOTE: Standard errors are in parentheses.

* = statistically significant at the 0.04 level.

** = statistically significant at the 0.01 level.

[8] Because only 15 percent of responding at-risk promoted students were promoted via the spring portfolio, we have a small sample size and, thus, less power to identify significant differences for this group.

For the most part, the retained students and the at-risk promoted students expressed comparable levels of mathematics confidence, regardless of promotion type. Students who had passed the spring portfolio reported being more confident in mathematics than did the retained students two years after the retention decision. However, no other group differences were observed for any other year.

Similarly, there were few differences in the reading confidence levels of at-risk promoted and retained students. Students who were promoted through the spring portfolio or their summer performance were less confident in their reading abilities than were retained students one year after the retention decision. However, this advantage dissipated by the second year.

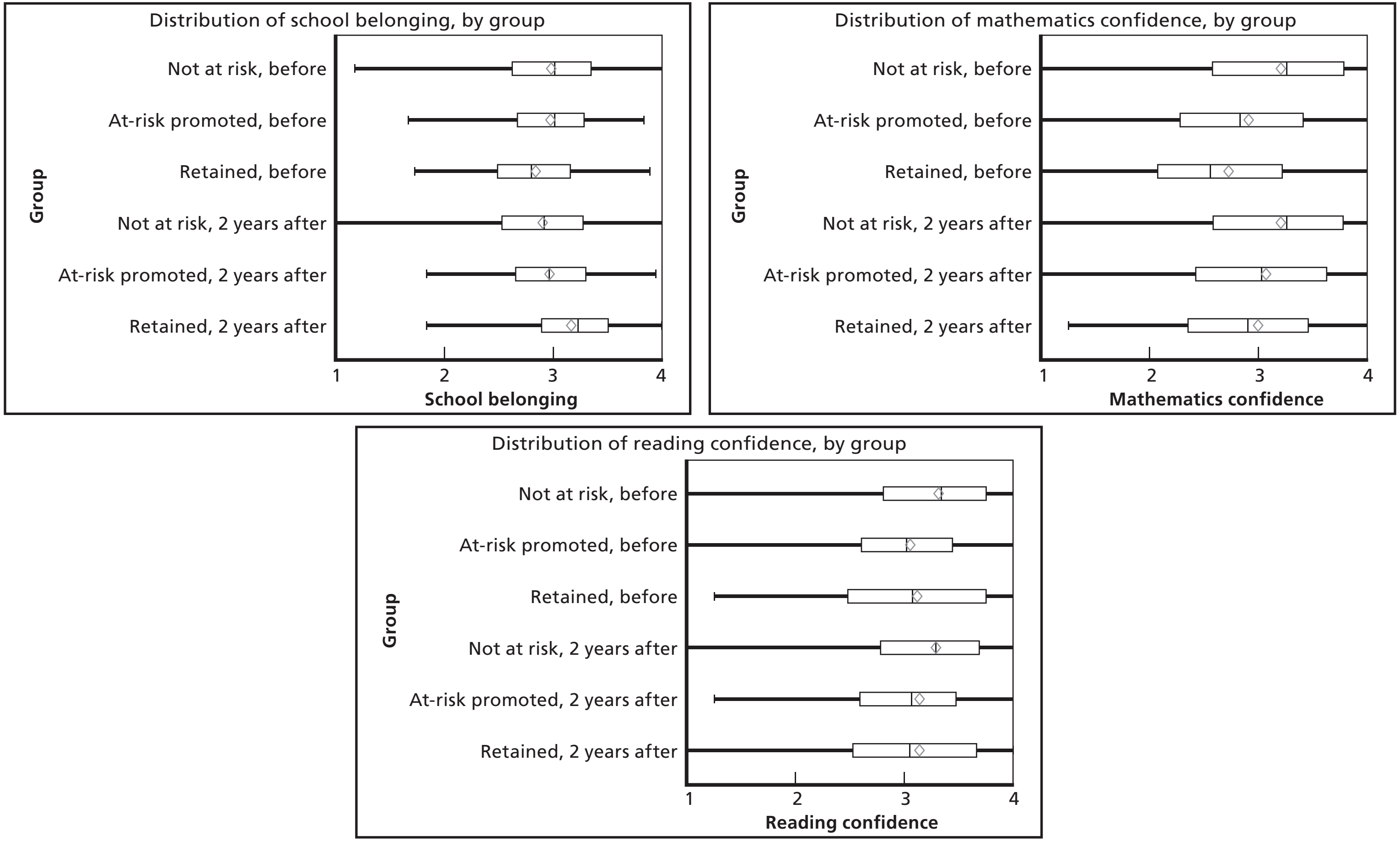
Taken as a whole, there were few differences among at-risk students with respect to subject-specific confidence. Regardless of promotion type, at-risk promoted students were generally comparable to retained students in their reading and mathematics attitudes. Retained students reported a higher level of school belonging than did at-risk promoted students, although these differences were most pronounced among the promoted students who attended summer school.

### Relationship Between Students' Socioemotional Responses and Future Retention Status

Up to this point, we have examined students' socioemotional responses only after the retention decision had been made. However, for the 2006–2007 5th-grade cohort, we were able to examine students' socioemotional responses prior to their being identified as at risk of retention. Figure 10.3 presents the mean differences for each group before and after the retention outcome. (Appendix D provides the descriptive statistics.) At the outset, the not-at-risk and retained students reported comparable levels of school belonging, but moderate growth over time ($d$ = 0.56) meant that the retained students surpassed the not-at-risk students in school belonging two years later ($d$ = 0.18). The retained students also surpassed the at-risk promoted students in school belonging two years after the retention decision ($d$ = 0.14). It is important to emphasize, however, that mean differences favoring the retained group were small.

Prior to the retention decision, the not-at-risk students had more positive mathematics attitudes than both the at-risk promoted and retained groups, although the effect sizes, which were in the range of 0.22 to 0.36, are considered small. Two years after the retention decision, the not-at-risk students continued to express more confidence in their mathematics and reading skills than both the at-risk promoted and retained students. However, the mean differences became markedly smaller, particularly for the retained students. Although retained students continued to lag behind their not-at-risk peers in mathematics confidence, they showed significant gains in mathematics confidence two years after the retention outcome ($d$ = 0.36). In contrast, neither the at-risk promoted nor the not-at-risk students showed increases in mathematics confidence over the two-year time frame.

**Figure 10.3**
**2006–2007 5th-Grade Cohort's Socioemotional Responses Before and After the Retention Decision**




**RAND** *MG894-10.3*

While a number of reasons underlie this finding, it is possible that the increase in confidence reflects a regression to the mean effect. As pointed out by Roderick and Nagaoka (2005), students are not retained at random but tend to be retained after a particularly low-performing school year. Because we surveyed retained students during the academic year in which they were to be held back, their confidence may have been uncharacteristically low. If regression to the mean is occurring, we would expect the retained students to follow the subpar year with a return to their more typical responses. This may explain why we observed a significant gain in their mathematics confidence two years after the retention decision. Indeed, other studies have confirmed the existence of regression to the mean in retention contexts (Alexander, Entwisle, and Dauber, 1994; Shepard and Smith, 1989).

Because we measured socioemotional attitudes prior to the retention decision, we were able to examine whether school belonging or self-confidence in mathematics and reading were predictive of students' future retention status. Table 10.5 provides the standardized regression coefficients for the weighted logistic regression model that predicts the probability of being retained at the end of the academic year. (See Appendix D for the full regression results.) In the model, retention is the outcome variable, and the scale scores are included as explanatory variables, along with the student-level demographics. Thus, a significant negative coefficient for a scale score indicates that a higher scale score is related to a reduction in the probability of retention. Neither school belonging nor reading confidence was related to retention status, but mathematics confidence was a significant predictor of retention. Students who reported being less confident in their mathematics abilities early in the school year were more likely to be retained at the end of that academic year.[9] While there are a variety of mechanisms by which confidence can be related to academic performance, this finding

**Table 10.5**
**Standardized Regression Coefficients Predicting Retention Status at the End of the Academic Year**

| Variable | Coefficient |
|---|---|
| School belonging | –0.04 (0.16) |
| Mathematics confidence | –0.46 (0.15)** |
| Reading confidence | –0.20 (0.18) |

NOTE: Standard errors are in parentheses.
* = statistically significant at the 0.04 level.
** = statistically significant at the 0.01 level.

[9] We explored the feasibility of additional analyses examining the relationships between socioemotional responses and reason for retention (i.e., failing the ELA assessment, failing the mathematics assessment, or failing both), but small sample sizes precluded these analyses.

supports the notion that retained students were likely to have already been struggling with their confidence prior to the retention decision, and any association between retention and subsequent confidence problems is unlikely to be solely attributable to the experience of being retained (Alexander, Entwisle, and Dauber, 1994).

## Summary

The effects of grade retention policies on students' affective well-being have been the subject of much debate. Taken as a whole, there were few differences among at-risk students with respect to subject-specific confidence. Regardless of promotion type, at-risk promoted students were generally comparable to retained students in their reading and mathematics attitudes. One consistent finding concerned school belonging: Retained students reported a greater sense of school connectedness than at-risk promoted students and not-at-risk students, even four years after the retention decision.

These results suggest that additional research is needed to shed more light on how retention is related to students' socioemotional well-being. Future research that explores a broad range of socioemotional outcomes, examines these affective responses over an extended time frame, and incorporates students' perspectives on being held back would provide policymakers with a better basis on which to understand the effects of retention policies on students' socioemotional well-being.

CHAPTER ELEVEN

# Conclusions and Policy Implications

*Sheila Nataraj Kirby, Jennifer Sloan McCombs, and Louis T. Mariano*

The NYC public school system is the largest in the country, with about 1.1 million students, 79,000 teachers, and more than 1,400 schools. As part of an ambitious citywide initiative to improve student performance, NYCDOE implemented a test-based promotion policy for students in grades 3, 5, 7, and, most recently, 8. General education students in these grades are required to meet promotion criteria on the ELA and mathematics assessments in order to be promoted. The grade promotion policy is not based on a sole criterion: Students may demonstrate basic proficiency in ELA and mathematics either through their performance on standardized tests administered during the school year or in August or through a review of a portfolio of their work. In addition, students may be promoted via an appeal. The policy places considerable emphasis on identifying struggling students early, providing them with additional instructional time, both in school and outside of school in Saturday and summer school programs, and continuously monitoring their progress.

Our evaluation examined several aspects of the 5th-grade promotion policy: (1) policy implementation, factors affecting implementation, and implementation progress over time; (2) the impact of the policy on student academic and socioemotional outcomes; and (3) the links between implementation, components of the policy, and desired outcomes. This monograph reports the results of the overall study and is one of three publications documenting the overall evaluation of NYC's 5th-grade promotion policy. The two other reports—Marsh et al. (2009) and Xia and Kirby (2009)—help situate the NYC policy in the broader national context and in the context of prior literature.

We collected data through a variety of methods over a three-year period, including case studies; surveys of administrators of schools, Saturday programs, and summer programs; surveys of students; interviews with leaders of NYC agencies designated to support schools; and interviews with officials of other states and districts implementing promotion policies similar to NYC's. We also obtained data on four cohorts of 5th-grade students: The first (the 2003–2004 cohort) was not subject to the promotion policy; the remaining three (the 2004–2005, 2005–2006, and 2006–2007 cohorts) were subject to the policy. We refer to these cohorts as the C0, P1, P2, and P3 cohorts, respectively. Additional data were obtained on three 3rd-grade cohorts, but the focus of the study was the 5th-grade promotion policy. The study used a combination of quantitative and qualitative methods to examine the implementation and impact of NYC's 5th-grade promotion policy on the immediate and future academic and socioemotional outcomes of students.

Our study faced two major limitations. The first was the change in test administration from the city to the state in the spring of 2006. The establishment of a new assessment, which was not equated to prior assessments, restricted our ability to compare longitudinal outcomes over time for some cohorts in certain years. The second limitation arose from the difficulty of sorting out the relative influences of the city, state, and national contexts on student outcomes. The study was intended to evaluate NYC's 5th-grade promotion policy and to estimate the effects of retention as well as the supportive interventions offered to at-risk students on current and future student outcomes. The fact that the promotion policy was only one piece of a larger set of reforms adopted by the city under the Children First initiative, enacted against the larger backdrop of a whole series of reforms undertaken by New York State and the pressures engendered by the federal NCLB mandates, makes the question of attribution challenging.

## Findings

In this monograph, we organized our findings around two main topics:

- implementation of the policy and effects on students, including
  - supports provided to 5th-grade students to help them meet promotion standards
  - effectiveness of supports in helping students meet 5th-grade promotion standards
  - performance of low-performing 5th-grade students in later years
- socioemotional outcomes of students held to the promotion policy.

### Implementation of the Policy and Effects on Students' Academic Outcomes

**Supports Provided to 5th-Grade Students to Help Them Meet the Promotion Standards.** Overall, approximately 60,000 5th-grade students were held to policy in each of the three cohorts subject to the promotion policy, accounting for 75–78 percent of the cohorts. Special education students, those classified as in ELL status for three or fewer years, and those in charter schools were exempt from the policy.

Students who scored Level 1 or low Level 2 on the 4th-grade assessments and retained students were categorized as needing services at the beginning of the school year—between 19 and 24 percent of 5th graders. Students were more likely to enter 5th grade needing services in ELA than in mathematics. The decrease in the percentage needing services between the P1 and P2 cohorts (from 24 percent to 19 percent) can be partly attributed to the additional services that low-performing students in the latter cohort may have received under the 3rd-grade promotion policy. This applied to the P3 cohort as well.

Students needing services were not evenly distributed across schools, and their numbers ranged from 0 to 80 percent of a school's 5th-grade students in 2005. Schools seemed to provide AIS to as many students as they had the capacity to serve. High-need schools (those with larger percentages of in-need students) were less likely to serve all students needing services than were other schools. However, high-need schools were significantly more likely than low-need schools to have additional resources for students, such as mathematics specialists, reading coaches, and mathematics coaches.

Schools provided additional services to students both during the school day and outside the school day. Across the three survey years (2006–2008), almost all schools reported serving students in need of services through small-group tutoring during the school day. Fewer schools were able to offer one-on-one tutoring during the school day (64–80 percent). While in-school services were valuable to students, our case-study visits suggest that services may not have been offered consistently when AIS providers were pulled away for other duties. Teaching test-taking strategies using professionally compiled test-preparation materials, such as Kaplan, or teacher-generated test-preparation materials was a focus to a moderate or great extent both during the school day and outside the school day.

Across the three years, the majority of respondents reported receiving useful support from their region or SSO. While the majority of schools did not report many major hindrances to improving student achievement, respondents in high-need schools were much more likely to do so than respondents in low-need schools.

Principals and teachers tended to be positive about many aspects of NYC's promotion policy, and approximately three-quarters of AIS leaders and principals agreed that the policy focused their schools' instruction efforts in a positive way and that it had made parents and guardians more concerned about their children's progress. However, the majority of respondents thought that the promotion policy relied too heavily on state assessment scores and, interestingly, that the policy made it more difficult to retain students who should be retained but passed the test.

Many students were supported by SPAs and SSAs—including students not technically in need of services or at risk of being retained. Both SSAs and SPAs consistently focused instructional time on ELA and mathematics and typically maintained small class sizes of 15 or fewer students. Because principals were given more autonomy, both programs began to vary more in terms of schedule and the types of students invited to attend. Overall, issues hindering implementation seem to have improved over time, perhaps because the increased autonomy has allowed sites to address them more effectively or because the number of students requiring services has declined.

### What Were the Trends in Student Achievement for Students Subject to the Policy?

The overall performance of the 5th-grade cohorts improved markedly over time. By 2006–2007 (P3 cohort), only 4 percent failed the spring assessments, while 65 percent and 80 percent scored at the proficient level in ELA and mathematics, respectively.

The performance of students needing services improved under the promotion policy. For example, the percentage of students needing services in ELA who scored Level 1 on the ELA spring assessment declined from 36 percent in the C0 cohort to 9 percent in the P1 cohort. The sharp decline in failure rate of the P1 cohort compared to the C0 cohort (both of which took the city assessment) is at least partly driven by the promotion policy and the supportive services provided to low-performing students during the school year. While students were more likely to need services in ELA at the beginning of the year, they were more likely to be retained for failing to meet promotion criteria in mathematics.

### How Did 5th Graders Needing Services at the Beginning of the Year Perform on the Spring Assessments Relative to the Control Cohort?

Overall, we found small positive effects of AIS intervention for in-need students. Our results show that in-need students in the P1 cohort increased their 5th-grade assessment scores by between 0.10 and 0.20 standard deviations in ELA and by less than 0.10 standard deviations in mathematics over their expected performance in the control year, and this is likely because of the supports provided under the promotion policy. However, for the small group of students entering 5th grade as low Level 1 students, we found little effect of services on their performance.

### How Effective Were Specific Supports in Helping Low-Performing Students Meet Promotion Standards in the Spring?

**Instructional Strategies.** Our exploratory models suggest that, in terms of school strategies, being in a school that offered one-on-one tutoring in mathematics to some or all students needing services increased students' likelihood of passing the spring mathematics assessment. Results for ELA one-on-one tutoring were promising in some models as well.

**SPA Attendance.** There was no relationship between additional SPA attendance and ELA outcomes in the P2 and P3 cohorts. In mathematics, we found that there was little benefit to attending fewer than six or seven sessions, while students who attended 14–15 sessions gained about 0.10 standard deviations over those attending six to seven sessions.

### Is SSA More Effective Than Typical Summer School in Improving Performance on the Summer Assessments?

Students who scored Level 1 on the spring assessments were mandated to attend SSA unless they demonstrated high Level 2 performance on a spring portfolio review. SSAs had a set instructional schedule and smaller class sizes than the summer school that students in the C0 cohort attended. The SSA intervention also included an additional potential motivator for students—the very real threat of being retained in grade. Relative to how they would have performed on the summer assessments in the control year, at-risk students in the P1 cohort performed slightly worse in ELA under the promotion policy (by about 0.10 standard deviations) and somewhat better in mathematics (by 0.20 standard deviations). This suggests that, while SSA in the first policy year was somewhat more effective in mathematics than typical summer school, this was not the case for ELA.

#### What Is the Relationship Between SSA Attendance and Performance on the Summer Assessments?

Using data on all students in the P1 and P2 cohorts who attended SSA, we found no relationship between longer attendance and summer ELA outcomes. There was a small but statistically significant relationship between attendance and performance on the summer mathematics assessment for one of the two cohorts (approximately 0.10 standard deviations) for students attending 14 sessions versus seven sessions. Prior research has found that summer school appears to have more beneficial effects on mathematics than on reading.

#### Who Was Retained Under the Promotion Policy?

About 2–3 percent of students in each cohort were retained in grade. By the P3 cohort, the percentage had fallen to 1 percent of the cohort (about 600 students out of approximately 58,000). Generally, retained students accounted for 30 percent of all students at risk of retention following the spring assessments.

Of those at risk of retention in the spring, about 30 percent passed the summer assessments and were promoted. There was a marked decrease in the percentage of students who failed the summer assessments but were promoted under the promotion policy—from 17 percent of the C0 cohort to about 1–2 percent in the three policy cohorts.

A little less than 20 percent of retained students were exempt from the policy in the subsequent (retention) year. While many were special education referrals, a number of retained students were exempt based on criteria other than those in the policy regulations.

#### How Did Retained Students Perform on the Repeated 5th-Grade Assessments and on Higher-Grade Assessments?

The performance of retained students on the repeated 5th-grade assessment improved substantially from spring 2006, both in terms of pass rates and proficiency levels (particularly in mathematics). For example, only 7–9 percent of retained students in the P3 cohort failed to meet promotion standards on the repeated 5th-grade assessments, while 20 percent of students retained for ELA and close to 45 percent retained for mathematics increased their scores to Level 3 or higher.

### How Did At-Risk Students Perform on Higher-Grade Assessments?

Over time, proficiency rates in grades 6 and 7 for all at-risk student groups increased, while the percentage of students scoring Level 1 in these grades declined. Students at risk in mathematics were far more likely to attain proficiency in that subject in future grades than were students at risk in ELA. In general, retained students performed better than at-risk promoted students in their cohort on the same-grade assessment (though they took these assessments in different years). Students promoted via appeals tended to do more poorly than their cohort peers. These trends were consistent in ELA and mathematics.

The improvements in 7th grade may have been at least partly due to the supportive services that low-performing students received under the 7th-grade promotion policy implemented in 2005–2006. However, we see similar improvements in the rest of the state.

### What Were the Effects of the Promotion Policy on Student Achievement Over Time?

We examined how specific groups of low-performing students subject to the promotion policy performed on higher-grade assessments relative to comparable groups of students. We were unable to track future outcomes for the very small group of students who had been twice retained in 5th grade. Overall, our estimates point to small to moderate positive effects of components of the promotion policy:

- *Small, positive effects of early identification and intervention.* Students in the P1 cohort who needed services at the beginning of 5th grade tended to score somewhat higher on the 7th-grade assessments than expected in the absence of the promotion policy (by 0.10–0.20 standard deviations in ELA and mathematics). This estimated effect captures the impact of AIS, Saturday and summer school (for some students), retention (for a small percentage of students), and any net difference in instructional services exclusive to lower-achieving students received by treated and comparison students through the time of the 7th-grade assessments.
- *Small, positive effects of summer school.* Compared to students who scored just above the Level 2 cutoff on the spring assessment, students in the P1 cohort who scored just below the cutoff and were required to attend summer school scored somewhat higher in both subjects (0.10–0.15 standard deviations higher) in both the 6th and 7th grades.
- *Moderate, positive effects of an additional year of instruction due to retention.* Retained students in the P1 cohort performed considerably better on the 7th-grade assessments (0.60 standard deviations in ELA and 0.40 standard deviations

in mathematics) than expected in the absence of the promotion policy. The effect sizes imply that a 5th-grade high Level 1 student who was retained would, after an additional year of 5th-grade instruction and intervention, be expected to score well within the Level 2 range on the 7th-grade assessments, based on the effects of retention alone.

Analyses focused on high Level 1 retained students in the P1 and P2 cohorts (compared to their peers scoring right above the Level 2 cutoff) also point to moderate, positive effects of retention over time (0.40 standard deviations higher in ELA and mathematics in both the 6th and 7th grades). The 6th-grade results for the P2 cohort were smaller in ELA. These estimated effects capture the impact of an additional year of schooling during the retained year and any net difference in instructional services received by treated and comparison students through 7th-grade assessments.

### Socioemotional Outcomes of Students Held to the Promotion Policy

We used data from student surveys to compare three groups of students—not at risk, at risk but promoted, and at risk and retained—on three socioemotional measures: school belonging, mathematics confidence, and reading confidence. Our analyses resulted in the following observations:

- *Retention had no negative effects on students' sense of school belonging or confidence in mathematics and reading.* On all three measures, the mean responses of retained students were comparable to or higher than those of their at-risk promoted peers. In all instances in which significant group differences were observed, the mean differences were considered small. These findings are consistent across grades as well. There was no difference with respect to self-confidence between retained students and their at-risk peers who had been promoted.
- *Retained students reported a greater sense of school belonging than at-risk promoted and not-at-risk students, even four years after the retention decision.* In all instances, the mean differences were small but statistically significant. These results are consistent with prior literature.

### Lessons Learned About the Design and Implementation of Promotion Policies

Although not reported in this monograph, we also analyzed the experience of other states and districts in designing and implementing test-based promotion policies similar to that of NYC (see Marsh et al., 2009). Here, we draw on those findings to provide a brief overview of the design and implementation of promotion policies.

NYC's own experience, combined with a set of interviews we conducted with officials from states and districts with promotion policies similar to NYC's, point to a useful set of lessons and practical insights into implementation for those who have adopted or are considering changes to promotion and retention policies.

With respect to design, the National Research Council (Heubert and Hauser, 1999, p. 135) pointed out that the validity and fairness of test-based promotion decisions can be enhanced by adopting the following strategies:

> (1) identifying at-risk or struggling students early so they can be targeted for extra help; (2) providing students with multiple opportunities to demonstrate their knowledge through repeated testing with alternate forms or other appropriate means; (3) taking into account other relevant information about individual students [American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1985].

NYC's promotion policy follows these tenets: It emphasizes early identification of at-risk students and provides them with additional instructional services both in school and out of school and it offers several opportunities and means for students to meet the promotion standards, including the summer assessment, the spring portfolio review, the August portfolio review, and an appeals process that allows other information to be taken into account. Other states and districts considering adopting promotion policies would do well to consider the key components of the NYC policy in their own design.

Additional lessons from NYC and other states and districts regarding design and implementation include the following:

- Invest in building support and maintaining ongoing communication with parents and educators.
- Anticipate the need to handle requests for exceptions to the policy.
- Identify students early in their careers and expand support beyond the promotion gateway grades.
- Provide adequate professional development for teachers.
- Invest in monitoring implementation and effects.
- Link the policy to a broader set of supports for at-risk students.
- Provide adequate funding.[1]

NYC, for example, rolled out the policy in stages—starting with students in grade 3 and then expanding it to higher grades as it gained more experience with the policy. NYC emphasized the need for open communication with parents, including sending out letters to parents in nine different languages to overcome language barriers. In addition, the policy was linked to a broad set of supports for schools and

---

1 See Marsh et al. (2009) for a full description of these and related findings.

students, and considerable funding for both professional development and supportive services was provided.

## Interpreting the Findings

Using a variety of methods, we found evidence that the supportive interventions had positive impacts in both the proximal year and two to three years later for students needing services and those who were retained in grade. However, it is important to keep in mind the larger context in which the promotion policy was enacted: It was one piece of a large, citywide initiative aimed at improving the conditions of learning and teaching in schools. On top of the city's efforts, the state invested a large amount of resources ($3.4 billion over a two-year period) in supporting the schools and adopted several reforms, including a well-defined curriculum aligned to state standards and increased emphasis on teacher qualifications and professional development. Pursuant to NCLB requirements, the state also took over testing in spring 2006, instituting new assessments in grades 3–8. Schools also face pressures from the NCLB requirements to bring all their students to the proficiency level and showing improvements every year.

Across the board, since 2006 in almost every grade, the percentage of students scoring at the proficiency level (Level 3 or higher) in NYC schools has increased dramatically, while the percentage of students scoring Level 1 has declined equally dramatically—and this was true across the rest of New York State as well. For example, the percentage proficient in grade 7 mathematics increased by 25 percentage points in NYC and 22 percentage points in the rest of the state between 2005–2006 and 2007–2008. The percentage scoring Level 1 declined to 3 percent by 2007–2008 in NYC and to 1 percent in the rest of the state. There were similar improvements in ELA. In the 5th grade, for example, the percentage scoring Level 3 increased 17 percentage points over the three years in both the city and the state. It appears likely that these improvements are related to the set of reforms enacted by the city and state in the NCLB context. However, other mechanisms may be partly responsible for the observed improvement as well. For example, prior research has shown that large increases in test scores are associated with score inflation resulting from teachers becoming more familiar with the test content and structure and thus "teaching to the test" (Koretz, 2008). We found that test-taking strategies were being emphasized in both in-school and out-of-school programs. If the tests and the standards and curriculum to which they are aligned are rigorous, there may be nothing adverse about the strategy and the results. It was beyond the scope of this study to attribute the increasing trends in city and state results to any particular source or to evaluate the rigor of the state test.

However, we needed to account for the effects of these other factors to determine the "true" effects of the promotion policy, to the extent feasible. Thus, we applied a correction for the upward trajectory in test scores. While our adjusted results for the effects of supportive interventions fall within the range reported by other studies (Matsudaira, 2008; Jacob and Lefgren, 2002; Roderick, Engel, et al., 2003), our estimates of the effects of retention on higher-grade assessments are larger than those previously reported (Roderick and Nagaoka, 2005; Greene and Winters, 2006; Jacob and Lefgren, 2002). There are three possible reasons for the divergence in estimates. First, like other studies, ours may not have been able to fully control for unobservable differences between the comparison and treated students. Second, the procedure used to account for a wider set of influences may not have fully adjusted for these effects; as a result, our estimates may not be true "net" estimates of the effect of the promotion policy and may be biased. Third, the larger effects for retention may be partly due to differences in methodology. We used same-grade comparisons to capture the effect of an additional year of instruction on the performance of retained students in future grades, while other studies used same-age comparisons to estimate students' ability to catch up to same-aged peers at a fixed time in the future.

One larger question remains: Will the benefits of the policy outweigh its costs in the long term? This question has two components. First, from an individual student's point of view, the question is whether these short-term gains will persist into high school. Recent studies that have used quasi-experimental methods similar to those used in our study have shown mixed results—short-term gains that decline over time. Prior studies have also shown that retained students have an increased probability of dropping out. We do not have data to address either of these issues. For the 5th-grade students retained under the policy, we are limited to two to three years from the retention decision for all cohorts. Second, from a societal point of view, it is important to understand the overall relative benefits and costs associated with implementing a promotion policy and holding students back a year. Eide and Goldhaber (2005), for example, focused strictly on the economic consequences of retention and concluded that the overall longer-term benefits from retention fall far short of covering the costs to society in terms of financing an extra year of education and loss of future earnings. The NYC policy has simply not been in place long enough to address these larger systemic costs and benefits.

Nonetheless, the bottom line is clear. Since the implementation of the promotion policy, substantially fewer 5th graders are being identified as needing services at the beginning of the year, the percentage of students failing the spring assessments is quite low, proficiency rates are high, and very few students are being retained—and those who are retained show substantial improvement on the repeated 5th-grade assessments. In terms of outcomes on the 6th- and 7th-grade assessments, we found small to moderate positive effects of early identification and intervention, summer school, an additional year of instruction for retained students. We also found no negative effects of retention on students' sense of school belonging or confidence in mathematics and reading over time. The near-term benefits we found hold out the possibility of longer-term benefits as well.

## Policy Recommendations

Based on our findings, we offer some recommendations for policymakers and administrators in NYC at the city, district, and school levels. While targeted at NYC, these recommendations may be of use to other districts and states considering or implementing test-based promotion policies.

*Continue early identification of students and provision of AIS.* A key component of NYC's promotion policy is the identification of students who need services early in the year so that these students can receive additional support to meet the promotion standards. Our findings suggest that the process of identification and support helped students meet the promotion standards and had positive benefits for student achievement in future grades. Results from our exploratory models suggest that the provision of one-on-one tutoring may be a particularly helpful form of AIS and should be continued and expanded where possible.

*Enable AIS providers to work consistently with students who need services.* While identification and AIS appear to have benefited students, our qualitative findings suggest that the consistency of AIS provision is sometimes a weakness in schools. In many cases, we found that AIS providers were not able to meet consistently with students as planned because they were often asked to assume other responsibilities, such as substitute teaching, lunch duty, or proctoring assessments. Although the teachers we interviewed considered AIS important and helpful for their students' academic growth, they often noted that AIS could be more effective if students received services on a more consistent basis.

*Consider the expected duration and participation when constructing Saturday programs.* Principals now have the authority to choose whether to operate a Saturday program for their students. While many principals have chosen not to operate a Saturday school, others do offer this intervention program. When making this decision or constructing a program, principals may want to consider that the intensity of the treatment seems to matter in terms of mathematics achievement. Our results suggest that students attending at least six to seven sessions typically have higher achievement outcomes.

*Continue to encourage struggling students to attend summer school.* Summer school attendance appears to have a positive relationship to achievement on the summer assessment, particularly in mathematics, and summer school may also have a positive impact on future achievement in grades 6 and 7. Our results suggest that these benefits accrue, in particular, for students close to the Level 2 cutoff. Thus, principals may want to encourage high Level 1 and low Level 2 students to attend summer school programs.

*Collect data on interventions being provided to at-risk students.* Our study highlighted a few instructional strategies that appeared to hold promise for helping low-performing students. Under the current structure, principals have considerable autonomy over the supports they provide to these students, and there is no centralized data-collection effort to track the treatments that individual students are receiving. However, it is important to collect and analyze these data to determine what works. The school data system implemented by NYC—the Achievement Reporting and Innovation System—may enable the collection of such data.

*Continue to monitor the effects of retention on students.* Our study tracked students through the 7th grade and found that retention appeared to have positive effects on achievement and no negative effects in terms of socioemotional status. As we noted earlier, one of the most important questions regarding the effects of retention is whether short-term positive effects persist over the longer term and whether the policy is cost-effective compared to alternatives. We could not answer these questions with our data, and they remain important topics for NYC to address in the future.

APPENDIX A

# Technical Appendix for Achievement Models

This appendix provides the methodological details for the achievement analyses presented in Chapters Eight and Nine. It details the propensity score weighting and doubly robust regression and RDD methods that were used to estimate treatment effects for students in the policy cohorts. It also provides additional details regarding the hierarchical and GAMM models used to estimate the relationship between intensity of participation in SPA and SSA and outcomes.

## Propensity Score Weighting and Doubly Robust Regression

We estimated the propensity scores using a flexible, nonparametric generalized boosting method, including the method described in McCaffrey, Ridgeway, and Morral (2004) and implemented in the Toolkit for Weighting and Analysis of Nonequivalent Groups (twang) package for the R statistical environment. One unique aspect of our propensity score analysis was that we did not restrict the comparison group to participants in the same schools as students who needed services, were at risk of retention in the spring, or were retained in grade. Instead, we conditioned on all available school-level characteristics and school district, thereby allowing us to match both within and across schools in the comparison cohort. Some schools had very few students treated under the policy; restricting comparisons to within schools would necessitate poor comparisons in some of these cases. Instead, conditioning on school-level characteristics allowed us to compare similar students within similar schools in the district. Additionally, we conducted separate propensity weighting for each outcome measure in question, thus giving us more refined balance than would have been the case if we used a single control group as the basis of comparison for all outcomes. A propensity score was assigned to each student in the comparison group, with students who were most similar to the students in the policy cohort receiving greater weight in the analysis.

The goal of the propensity scoring process was to equalize, to the greatest extent possible, the distribution of all available covariates across the treated and control cohorts. In addition to baseline assessment scores, available covariates included student characteristics, school characteristics, and district.[1] For each available covariate, we calculated a Kolmogorov-Smirnov (K-S) statistic to examine how well the distributions of that covariate were matched between the treated and control cohorts under the calculated propensity scores, with the goal of choosing scores that minimized the set of K-S statistics for all covariates to the greatest extent possible.

Given propensity scores that equalized the distribution of the features of the treated and untreated subjects, we assigned each treated subject a weight of 1 and each control subject a weight of $w_i = P(\text{treated} \mid \mathbf{x}_i)/\left(1 - P(\text{treated} \mid \mathbf{x}_i)\right)$, so that the control weights corresponded to the odds that a subject with features $\mathbf{X}_i$ would be in the treated group. We then estimated a weighted regression model that also included the same set of all available covariates used in the propensity scoring process and an indicator variable ($Z_i$) denoting whether the student was from the treated cohort ($Z_i = 1$) or the comparison cohort ($Z_i = 0$). Because the comparison students had already been weighted according to their similarity to the treated students through the propensity scoring process, there was actually no need to include demographics or school-level variables in the model (if the K-S statistic indicated that the distributions between the two groups on these variables were statistically equivalent). However, the inclusion of covariates also makes the treatment effect estimate "doubly robust" in the sense that if either the propensity score model or the regression model is correct, then the treatment effect estimator will be consistent. The goal is not to make the most accurate inferences about predictors of treatment but to make the most accurate inferences about the effects of treatment. Note that the counterfactual we are using here is "How would the treated students have performed had they not been treated?" Thus we are interested in the treatment effect on the *treated*.

When the outcome variable ($Y_i$) is continuous (i.e., an assessment scale score), the fitted weighted linear regression takes the following form:

$$Y_i = \mu_o + Z_i\beta_w + \mathbf{X}_i\boldsymbol{\psi}_w + \varepsilon_i. \tag{A.1}$$

The estimate of the coefficient of the treatment indicator, $\hat{\beta}_w$, representing the mean outcome difference between the treated students in the policy cohort and weighted control students in the comparison cohort, is a doubly robust estimate of the treatment effect on the treated students (Bang and Robins, 2005).

When the outcome variable $Y_i$ is binary (i.e., the proportion of students scoring at least Level 2 or the proportion retained), we fit a logistic regression:

$$\text{logit}\left(p(Y_i = 1)\right) = \mu_o + Z_i\beta_w + \mathbf{X}_i\boldsymbol{\psi}_w + \varepsilon_i. \tag{A.2}$$

[1] See the Chapter Four for a discussion of the available student and school-level data.

In this case, the estimated coefficient of the treatment indicator is on the logit scale. The true proportion of students with positive outcomes is, of course, known for the treated students in the policy cohort. We use the fitted weighted logistic regression in equation A.2 to predict the counterfactual responses for these students by setting $Z_i = 0$ and keeping all other covariates constant at their original values, and we also record the standard error of those predictions. Similarly, the fitted version of equation A.2 may be used to predict the difference $(Y_i \mid Z_i = 1) - (Y_i \mid Z_i = 0)$, which is then a doubly robust estimate of the treatment effect of the treated students in the policy cohort.

## Regression Discontinuity Design Models

Our modeling approach for estimating the treatment effects of interest in all three research questions was essentially the same, so we describe the procedure in a more generic fashion that applies to all three. Treatment assignment under consideration in each of the research questions is not strictly a deterministic function of the assignment variable; students below the threshold have an increased probability of treatment, indicating a fuzzy RDD. In this case, the treatment effect on the treated students below the threshold is identified by dividing the effect of treatment eligibility on the outcome at the threshold value by the effect of treatment eligibility on the probability of receiving the treatment at the threshold value. We may model these two values separately via local linear regression techniques and then take the ratio of these two estimates as the estimate of the treatment effect on the treated, generally following the procedure described by Imbens and Lemieux (2008).

Let $T_i$ be an indicator variable of the assessment score in a given subject falling below the threshold value for student *i*:

$$T_i = 1\left\{S_i < 0\right\}, \tag{A.3}$$

where $S_i$ is the student's test score, centered on the pass-fail cutoff so that negative values of $S_i$ indicate treatment eligibility.

Let $W_i$ be an indicator of the receipt of treatment by student *i*:

$$W_i = 1\left\{\text{student } i \text{ treated}\right\}. \tag{A.4}$$

Finally, let $Y_i$ represent a future outcome of interest (e.g., a future ELA or mathematics assessment score). We can estimate the effect of treatment eligibility on the outcome at the threshold $\tau_y$ by fitting the regression equation

$$Y_i = \alpha_y + \beta_y S_i + \tau_y T_i + \psi_y T_i S_i + \varepsilon_i, \tag{A.5}$$

using observations where the values of $S_i$ are in a local neighborhood around the threshold of zero. Notice that equation A.5 allows for different intercepts and slopes on either side of the threshold:

$$\text{For } S \leq 0\text{: } Y_i = (\alpha_y + \tau_y) + (\beta_y + \psi_y)S_i + \varepsilon_i. \tag{A.6}$$

$$\text{For } S > 0\text{: } Y_i = \alpha_y + \beta_y S_i + \varepsilon_i. \tag{A.7}$$

At the threshold value $S_i = 0$, the difference in intercepts in equations A.6 and A.7 is then $(\alpha_y + \tau_y) - \alpha_y = \tau_y$. We denote the estimated value of $\tau_y$ from the fitted regression equation by $\hat{\tau}_y$. In the same way, we can estimate the effect of treatment eligibility on the probability of receiving the treatment $\tau_w$ by fitting the regression equation

$$W_i = \alpha_w + \beta_w S_i + \tau_w T_i + \psi_w T_i S_i + \varepsilon_i, \tag{A.8}$$

again using observations where the values of $S_i$ are in a local neighborhood around the threshold of zero.

An estimate of the treatment effect on the treated students near the treatment threshold is then calculated as

$$\hat{\theta} = \frac{\hat{\tau}_y}{\hat{\tau}_w}. \tag{A.9}$$

The model can be extended to include other covariates of interest, and, as we show below, the assignment variable can be modeled using a more flexible parametric representation. For example, Matsudaira (2008) allowed for inclusion of a *p*th-degree polynomial in $S_i$ to be added to the regression model and ultimately included up to a third-order polynomial for some outcomes based on the Bayesian information criterion (Schwarz, 1978; Kass and Raftery, 1995) for model selection.

The estimation of the RDD models required the following steps.

**Step 1.** The validity of the RDD rests on the assumption that the observed baseline covariates are locally balanced on either side of the threshold (i.e., the conditional expectations of all other characteristics affecting outcomes are continuous at the threshold). This implies that students on either side of the threshold have, on average, the same value on any observed covariate (analogous to a randomized controlled trial), and the only factor that is changing discontinuously at the threshold is the probability that the student receives the treatment (Matsudaira, 2008). We used canonical correlation procedures to find the linear combination of covariates most highly correlated with outcomes, estimated that vector as a function of the treatment variable in an RDD analysis, and checked that the treatment effect was insignificant, suggesting that there was no discontinuity in these covariates at the threshold. As another check, we also examined selected covariates using a graphical RDD analysis.

**Step 2.** There were two critical choices to be made—the selection of the bandwidth and the choice of the order of the polynomial regressions. As Lee and Lemieux (2009, p. 39) point out, choosing the bandwidth involves finding an optimal balance between precision and bias. Using a larger bandwidth means that more observations are available to estimate the regression, and this leads to greater precision. However, there is no guarantee that a larger bandwidth will reduce the bias in the estimate at the discontinuity point, which is essentially at the boundary of the range of data being fit to the model. We use the cross-validation approach outlined by Ludwig and Miller (2005) and Imbens and Lemieux (2008) to jointly determine both the bandwidth and the appropriate degree polynomial form for the baseline test score. As candidate bandwidths, we considered a sequence of increments of 0.25 standard deviations of *S*, from 0.25 to 3.5. For each potential bandwidth, for any observation of the assignment variable $s_k$ within one-half of the bandwidth of *0*, we ran four separate regressions with observation $S = s_k$ left out and used the estimates to predict the values of all responses, *Y*, at $S = s_k$ from each fitted regression. The four regression models included different polynomial term for *S*, from only a linear *S* through a quartic $S^4$ term. Because the RDD estimates were based on regression estimates at the boundary, this procedure uses only observations with values of *S* on the left of $s_k$ for observations on the left of the cutoff point ($S_i < 0$) and values of *S* on the right of $s_k$ for observations on the right of the cutoff point ($S_i \geq 0$). We then compared the set of predicted values of *Y* against the actual values of *Y*. The optimal bandwidth and appropriate degree polynomial were selected by choosing the values of both that minimized the mean square error (MSE) of the difference between the predicted and actual values of *Y*.

This procedure was then repeated for the treatment assignment, replacing *W* for *Y* analogous to equations A.5 and A.8. It was rare that the optimal choice for both the outcome and treatment assignment coincided. To reconcile the two, we sought to find choices for bandwidth and polynomial that were within 1 percent of the optimal MSE for both. Where multiple such choices existed, we chose the cell with the smallest average percentage MSE. In the minority of cases in which there was no overlap, we followed the recommendation of Imbens and Lemieux (2008) in taking the lower suggested bandwidth between the two and then choosing the appropriate polynomial within that bandwidth with the smallest average percentage MSE.

Note that we implemented this cross-validation procedure for each treatment and assignment variable, choosing the optimal bandwidth and degree of polynomial individually for each. This localized optimization produced a range of bandwidth choices, and it was not pragmatic to use a generalized overall choice. In the RDD results presented next, we note the number of observations within the bandwidth used to fit each model. Typically, a large majority of those falling within the bandwidth are above the treatment threshold; thus, we display the sample size separately for both above and below the threshold.

**Step 3.** For each model, we estimated the RDD model as a two-stage least-squares design in Stata using the *ivreg* procedure. We instrument for $W_i$ in the structural model using $T_i$ and the polynomial terms of $S_i$ and their interactions. Following Lee and Lemieux (2009), we allowed the parameters of each term of the polynomial to vary on either side by fully interacting with the treatment indicator. We report the heteroskedasticity-robust standard errors to correct for clustering at each discrete support point of the prior test score.

## Modeling the Relationship Between Intensity of Participation and Outcomes

We examine how variation in attendance might affect outcomes by considering the following hierarchical model, in which students are in programs (SPA or SSA) that are nested within districts and region, forming levels 2 and 3 of the model, and are modeled as random effects (not shown):

$$Y_{idr} = \mu_d + \beta I_i + \mathbf{X}_i \gamma + \varepsilon_{ij} \quad \text{(level 1)}, \tag{A.10}$$

where $Y_{idr}$ represents the outcome of student *i* in district *d* and region *r*; $I_i$ is the intensity of program participation for student *i* (i.e., number of days attended); $\mathbf{X}_i$ represents a vector of student variables; and $\gamma$ is the vectors of regression coefficients on the student-level variables, $\mu_d$ is the district random effect, and $\varepsilon_{ij}$ represents student-level variance.

Note that this implies a linear relationship between student achievement and number of days attended. We then relax this assumption and estimate

$$Y_{ij} = \mu_j + \varphi(I_i) + \mathbf{X}_i \gamma + \varepsilon_{ij} \quad \text{(level 1)}. \tag{A.11}$$

In equation A.11, we represent the marginal relationship between program attendance, $I_i$, and the outcome as a smooth arbitrary nonparametric function, $\varphi(I_i)$, rather than a linear coefficient $\varphi$. We used a GAMM framework (Lin and Zhang, 1999) to estimate equation A.11 and the other two levels of the model. Spline functions offer the best estimate of the nonparametric function $\varphi(I_i)$, with smoothing parameters estimated through cross-validation. The GAMM framework is implemented in the *mgcv* package for the R statistical environment.

GAMM provides a graphical representation of the relationship between attendance and outcomes after adjusting for covariates and clustering. A plot of the values from the smooth function $\varphi(I_i)$ against the attendance values allows us to inspect the trend lines for indications of the existence of change-points in the relationship (such as material changes in the slope).

APPENDIX B

# Supporting Data for Chapter Five

This appendix provides supporting data for the analyses presented in Chapter Five. The figures and tables are based on principal and AIS administrator survey and indicate the survey years from which the data were drawn. "NA" indicates that the particular question or item did not appear on that year's survey.

**Table B.1**
**Percentage of Schools with an AIS Team**

| Need Status of School | 2006 | 2007 |
|---|---|---|
| All schools | 99.3 | 97.2 |
| Low-need schools | 100.0 | 98.4 |
| Medium-need schools | 100.0 | 98.5 |
| High-need schools | 98.0 | 94.7 |

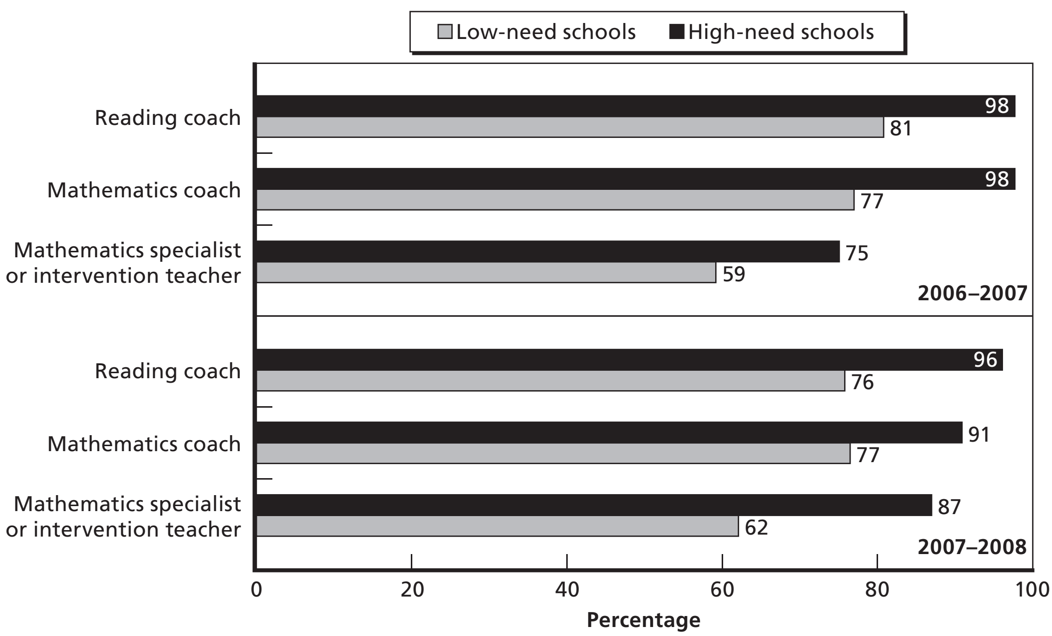
**Table B.2**
**Percentage of Respondents Reporting That Data Sources Were Used to a Moderate or Great Extent to Select Students for AIS**

| Data Source | 2007 | 2008 |
|---|---|---|
| Prior-year state assessment results | 97.9 | 95.3 |
| Teacher recommendations | 95.0 | 95.8 |
| Progress or interim assessment results | 81.7 | NA |
| Parent requests | 44.6 | 54.9 |

**Table B.3**
**Percentage of Respondents Reporting That Data Sources Were Helpful to a Moderate or Great Extent in Monitoring the Progress of In-Need 5th-Grade Students**

| Data Source | 2006 | 2007 |
|---|---|---|
| City interim assessments | 67.6 | 77.6 |
| Diagnostic or other non-city interim assessments (e.g., Degrees of Reading Power, Kaplan) | 89.9 | 79.6 |
| Teacher-created tests | 87.9 | 88.4 |
| Letter grades or GPAs | 46.3 | 48.1 |
| Student work | 98.5 | 96.0 |
| Progress reports from AIS teachers | NA | 89.6 |

**Figure B.1**
**AIS Staffing Resources Available, by Need Status of School, 2007 and 2008**




RAND *MG894-B.1*

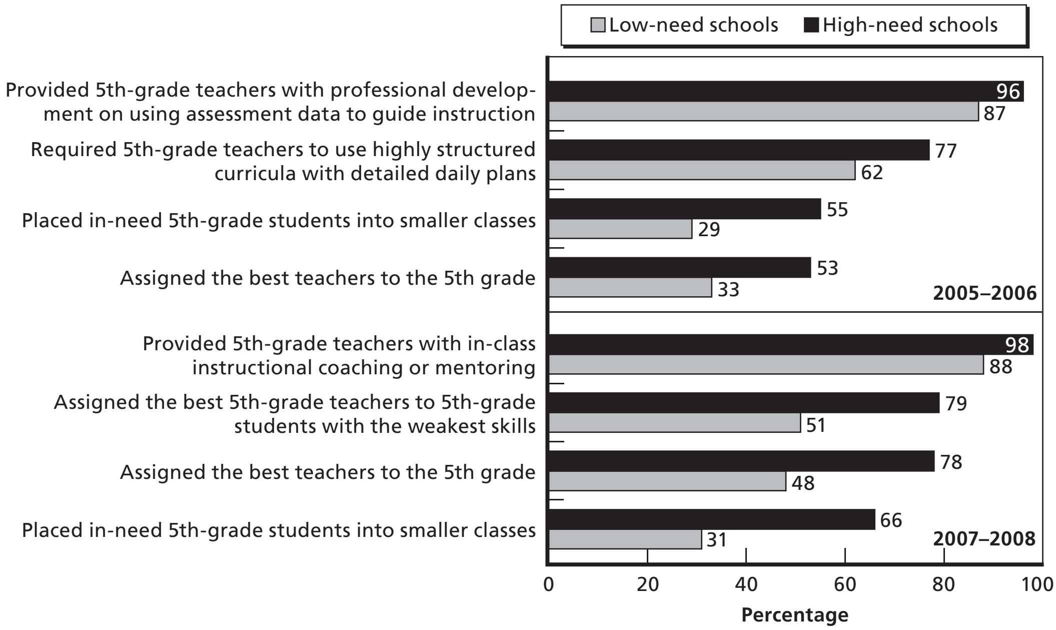
**Table B.4**
**Percentage of Administrators Who Perceived Various Programs to Be Moderately or Very Effective, Schools with Programs**

| Program | 2006 | 2007 | 2008 |
|---|---|---|---|
| Before- or after-school program | 98.0 | 89.7 | 90.2 |
| One-on-one tutoring with a specialist during the school day | 96.6 | 94.8 | 95.9 |
| Small-group tutoring with a specialist during the school day | 97.8 | 99.6 | 98.6 |
| 37.5-minute program | NA | 80.8 | 84.6 |
| Intersession | NA | 59.3 | 80.5 |

**Table B.5**
**Percentage of Administrators Using School Strategies to Improve the Academic Performance of In-Need 5th-Grade Students Who Perceive Them to Be Moderately or Very Effective**

| Strategy | 2006 | 2007 | 2008 |
|---|---|---|---|
| Assigned the best teachers to the 5th grade | 93.7 | 91.0 | 96.6 |
| Assigned the best 5th-grade teachers to 5th-grade students with the weakest skills | 94.1 | 92.6 | 99.1 |
| Placed in-need 5th-grade students into smaller classes | 92.4 | 86.3 | 92.3 |
| Grouped 5th-grade retained students together for instruction for the entire day | 75.9 | 60.2 | 69.3 |
| Grouped 5th-grade retained students together for instruction for part of the day | 76.8 | 77.8 | 90.1 |
| Provided 5th-grade teachers with in-class instructional coaching or mentoring | 93.5 | 82.3 | 96.0 |
| Provided 5th-grade teachers with professional development on using assessment data to guide instruction | 92.5 | 86.1 | 97.6 |
| Provided 5th-grade teachers with professional development on differentiated instruction | 90.7 | 84.9 | 96.6 |
| Required 5th-grade teachers to use highly structured curricula with detailed daily plans | 91.3 | 84.4 | 93.2 |
| Required 5th-grade teachers to use materials that taught test-taking strategies | 91.1 | 84.4 | 91.0 |

**Figure B.2**
**Principals' Reports About Adopting Certain School Strategies in the 5th Grade, by Need Status of School, 2006 and 2008**

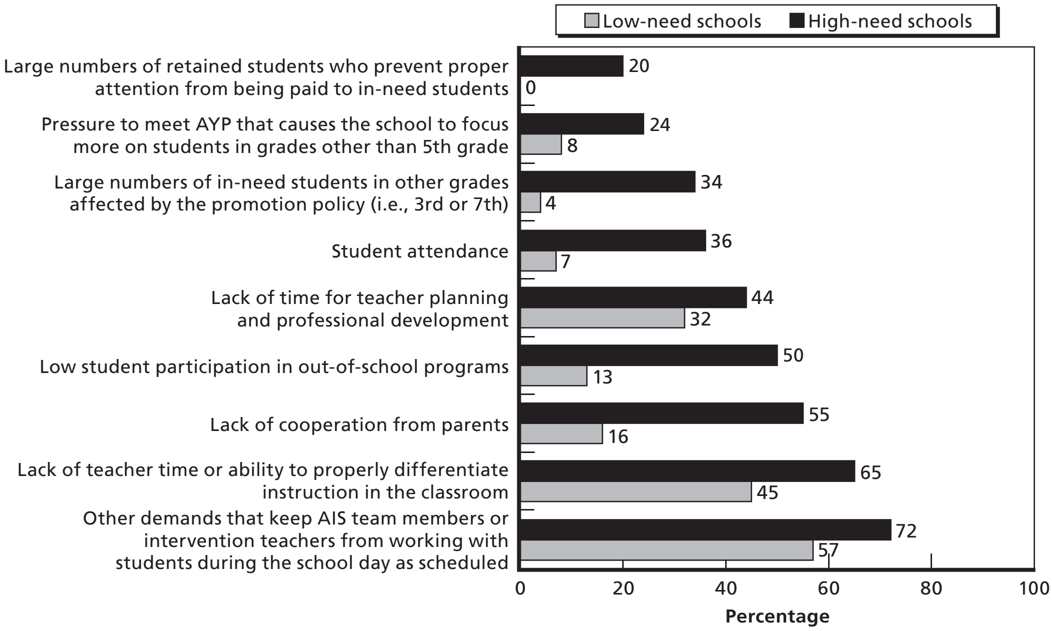



RAND *MG894-B.2*

**Table B.6**
**School Administrators' Reports About Student Retention**

| Perception | 2006 | 2007 |
|---|---|---|
| Holding students back a grade improves their long-term chances of success. | 52.0 | 49.0 |
| Retention negatively affects students' self-esteem. | 69.8 | 67.3 |
| The threat of retention motivates students to work harder in school. | 43.3 | 44.6 |
| The threat of retention upsets some students so badly that they learn less. | 46.2 | 37.3 |
| Holding students back makes them more likely to dislike school. | 56.0 | 48.8 |
| Retaining a student almost guarantees that he or she will drop out of school. | 17.8 | 11.9 |
| Most students end up with stronger skills after repeating a grade. | 60.9 | 64.7 |
| Retaining a student can have a negative or positive impact, depending on the needs of the individual student. | 91.0 | 88.4 |

**Figure B.3**
**Extent to Which Factors Were a Hindrance to Improving the Performance of In-Need 5th-Grade Students, 2007**




RAND *MG894-B.3*

**Table B.7**
**Definitions of Independent Variables Used in the Analyses**

| Variable | Definition |
|---|---|
| Percentage of in-need students in school | Percentage of 5th-grade students scoring Level 1 or low Level 2 on the prior year's assessment |
| Percentage of students not held to the policy | Percentage of 5th-grade students in the school who were not held to the promotion policy—ELL students and those needing special education services |
| NCLB status | = 0, if school is in good standing under the state NCLB requirements<br>= 1, otherwise |
| Prior year's achievement | Scale score from the prior year's assessment |
| Served all students needing services in reading | = 1, if school served all Level 1 and low Level 2 students in reading<br>= 0, otherwise |
| Served all students needing services in mathematics | = 1, if school served all Level 1 and low Level 2 students in mathematics<br>= 0, otherwise |

**Table B.7—Continued**

| Variable | Definition |
|---|---|
| One-on-one tutoring in reading | = 1, if school provided one-on-one tutoring to some or all in-need students in reading<br>= 0, otherwise |
| One-on-one tutoring in mathematics | = 1, if school provided one-on-one tutoring to some or all in-need students in mathematics<br>= 0, otherwise |
| Reduced class size | = 1, if school placed in-need 5th-grade students into smaller classes<br>= 0, otherwise |
| Special teacher assignment | = 1, if school assigned the best teachers to the 5th grade and/or assigned the best 5th-grade teachers to 5th-grade students with the weakest skills<br>= 0, otherwise |
| Perceived hindrances, four-point Likert scale ranging from "not a hindrance" to "a great hindrance"<br>(year 1 alpha = 0.84)<br>(year 2 alpha = 0.81)<br>(year 3 alpha = 0.74) | To what extent is each of the following a hindrance to the school's implementation of the 5th-grade promotion policy?<br>Limited time available for AIS leader to work with teachers (year 1)<br>Lack of teacher time or ability to properly differentiate instruction in the classroom (year 1)<br>Pressure to meet AYP causes the school to focus more on improving achievement of Level 2 students than Level 1 students (year 1)<br>Large numbers of retained students who prevent proper attention from being paid to in-need students (years 1 and 2)<br>Large numbers of previously retained students who prevent proper attention from being paid to other students at risk of being retained (year 3)<br>Large numbers of in-need students in other grades affected by the promotion policy (i.e., 3rd or 7th) (years 1 and 2)<br>Large numbers of in-need students retained in other grades affected by the promotion policy (i.e., 3rd or 7th) (year 3)<br>AIS leader's limited skills or experience working with teachers (year 1)<br>Pressure to meet AYP causes the school to focus more on students in grades other than 5th (year 1)<br>Lack of cooperation from parents (years 1, 2, and 3)<br>Other demands that keep AIS team members or intervention teachers from working with students during the school day as scheduled (years 2 and 3)<br>Low student participation in out-of-school (e.g., before- or after-school, Saturday, or intersession) programs (years 2 and 3)<br>Large class size (years 2 and 3) |

**Table B.7—Continued**

| Variable | Definition |
|---|---|
| Perceived hindrances (cont.) | Lack of time for teacher planning and professional development (years 2 and 3)<br>Student attendance (year 2)<br>Poor student attendance (year 3) |
| Level of support received, four-point Likert scale ranging from "strongly disagree" to "strongly agree"<br>(year 1 alpha = 0.87)<br>(year 2 alpha = 0.85)<br>(year 3 alpha = 0.92) | To what extent do you agree or disagree with the following statements about the support you receive from your region and/or district for implementing the 5th-grade promotion policy? (year 1)<br>To what extent do you agree or disagree with the following statements about the support you receive from your region and/or district as an AIS leader? (year 2)<br>To what extent do you agree or disagree with the following statements about the support you receive from your SSO? (year 3)<br>Provides the school with necessary guidance on selecting curriculum materials to support in-need students (years 1, 2, and 3)<br>Provides principals with high-quality professional development focused on supporting in-need students and their teachers (year 1)<br>Provides me (the AIS leader) with high-quality professional development focused on supporting in-need students and their teachers (year 2)<br>Provides me (the principal) with high-quality professional development focused on supporting in-need students and their teachers (year 3)<br>Provides teachers with high-quality professional development focused on teaching in-need students (years 1, 2, and 3)<br>Helps the school get the resources it needs to support in-need students (years 1, 2, and 3)<br>Provides useful assistance in analyzing student data (years 1, 2, and 3)<br>Provides useful support for AIS leaders (year 3) |
| Teacher quality, four-point Likert scale ranging from "strongly disagree" to "strongly agree"<br>(year 1 alpha = 0.92)<br>(year 2 alpha = 0.93) | To what extent do you agree or disagree with the following statements about the teachers at your school? (years 1 and 2)<br>Have a "can-do" attitude<br>Are continually learning and seeking new ideas<br>Work together to do what is "best for kids"<br>Work hard to help their students succeed<br>Really care about their students<br>Consistently enforce rules of the school, even for students who are not in their classes<br>Share beliefs and values about what the central mission of this school should be<br>Make an effort to reach out to the community |

**Table B.7—Continued**

| Variable | Definition |
|---|---|
| Principal leadership, four-point Likert scale ranging from "strongly disagree" to "strongly agree"<br><br>(year 2 alpha = 0.95) | To what extent do you agree or disagree with each of the following statements about your principal's leadership? (year 2)<br>Communicates a clear academic vision for the school<br>Sets high standards for teaching<br>Encourages teachers to review the New York State standards and incorporate them into their teaching<br>Helps teachers adapt their curriculum based on an analysis of state test results<br>Ensures that teachers have sufficient time for professional development and planning<br>Enforces school rules for student conduct and backs up teachers when needed<br>Makes the school run smoothly |

**Table B.8**
**Means and Standard Deviations of Variables Used in the Analysis for In-Need Students, by Year, ELA**

| | 2005–2006 | | 2006–2007 | | 2007–2008 | |
|---|---|---|---|---|---|---|
| **Variable** | **Mean** | **Standard Deviation** | **Mean** | **Standard Deviation** | **Mean** | **Standard Deviation** |
| Passed ELA assessment | 0.14 | 0.35 | 0.07 | 0.25 | 0.03 | 0.17 |
| Current year's achievement in ELA | 627.85 | 24.25 | 636.26 | 20.18 | 639.89 | 18.02 |
| Prior year's achievement in ELA | 614.84 | 14.07 | 613.35 | 17.19 | 614.62 | 16.17 |
| NCLB status | 0.29 | 0.45 | 0.24 | 0.43 | NA | NA |
| Percentage of students needing services | 7.35 | 6.43 | 12.51 | 8.01 | 9.17 | 6.56 |
| Percentage of students not held to policy | 21.37 | 8.08 | 21.22 | 8.41 | 21.50 | 8.16 |
| Serving all students needing services | 0.63 | 0.48 | 0.48 | 0.50 | 0.68 | 0.47 |
| One-on-one tutoring | 0.81 | 0.39 | 0.66 | 0.48 | 0.78 | 0.42 |
| Special teacher assignments | 0.62 | 0.48 | NA | NA | 0.76 | 0.43 |
| Before- or after-school programs | 0.60 | 0.49 | 0.93 | 0.26 | 0.84 | 0.37 |
| Reduced class size | 0.44 | 0.50 | 0.37 | 0.48 | 0.52 | 0.50 |
| Hindrances | 2.13 | 0.65 | 2.30 | 0.69 | 2.03 | 0.60 |
| Level of support | 2.74 | 0.56 | 2.78 | 0.58 | 2.95 | 0.73 |
| Teacher quality | 3.25 | 0.52 | 3.16 | 0.52 | NA | NA |
| Principal leadership | NA | NA | 3.23 | 0.69 | NA | NA |

**Table B.9**
**Means and Standard Deviations of Variables Used in the Analysis for In-Need Students, by Year, Mathematics**

| | 2005–2006 | | 2006–2007 | | 2007–2008 | |
|---|---|---|---|---|---|---|
| Variable | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| Passed mathematics assessment | 0.38 | 0.49 | 0.16 | 0.36 | 0.12 | 0.33 |
| Current year's achievement in mathematics | 623.59 | 25.10 | 637.32 | 22.48 | 638.32 | 22.45 |
| Prior year's achievement in mathematics | 606.63 | 20.01 | 613.29 | 16.95 | 616.05 | 16.32 |
| NCLB status | 0.35 | 0.48 | 0.29 | 0.46 | NA | NA |
| Percentage of students needing services | 9.28 | 6.57 | 14.36 | 7.87 | 10.47 | 6.65 |
| Percentage of students not held to policy | 21.93 | 7.98 | 21.80 | 8.32 | 22.10 | 8.48 |
| Serving all students needing services | 0.60 | 0.49 | 0.45 | 0.50 | 0.64 | 0.48 |
| One-on-one tutoring | 0.81 | 0.39 | 0.65 | 0.48 | 0.78 | 0.42 |
| Special teacher assignments | 0.62 | 0.49 | NA | NA | 0.78 | 0.41 |
| Before- or after-school programs | 0.59 | 0.49 | 0.93 | 0.25 | 0.84 | 0.36 |
| Reduced class size | 0.47 | 0.50 | 0.36 | 0.48 | 0.55 | 0.50 |
| Hindrances | 2.22 | 0.65 | 2.33 | 0.71 | 2.09 | 0.63 |
| Level of support | 2.71 | 0.59 | 2.77 | 0.58 | 2.99 | 0.72 |
| Teacher quality | 3.20 | 0.50 | 3.13 | 0.53 | NA | NA |
| Principal leadership | NA | NA | 3.19 | 0.70 | NA | NA |

**Table B.10**
**Means and Standard Deviations of Variables Used in the Analysis for All Held-to-Policy Students, by Year**

| | 2005–2006 | | 2006–2007 | | 2007–2008 | |
|---|---|---|---|---|---|---|
| **Variable** | **Mean** | **Standard Deviation** | **Mean** | **Standard Deviation** | **Mean** | **Standard Deviation** |
| Passed ELA | 0.04 | 0.19 | 0.02 | 0.13 | 0.01 | 0.07 |
| Current year's achievement in ELA | 662.22 | 35.54 | 665.51 | 31.93 | 666.03 | 26.52 |
| Prior year's achievement in ELA | 663.29 | 37.83 | 666.41 | 34.34 | 661.83 | 31.80 |
| Passed mathematics | 0.06 | 0.23 | 0.02 | 0.14 | 0.01 | 0.11 |
| Current year's achievement in mathematics | 670.97 | 37.07 | 682.07 | 35.90 | 683.78 | 34.18 |
| Prior year's achievement in mathematics | 670.58 | 35.91 | 680.25 | 37.84 | 682.57 | 36.69 |
| NCLB status | 0.19 | 0.39 | 0.17 | 0.38 | NA | NA |
| Percentage of students needing services | 4.41 | 5.15 | 8.08 | 7.37 | 5.77 | 5.56 |
| Percentage of students not held to policy | 19.99 | 7.98 | 18.87 | 8.23 | 19.47 | 8.12 |
| Serving all students needing services | 0.64 | 0.48 | NA | NA | NA | NA |
| Serving all students needing services in ELA | NA | NA | 0.55 | 0.50 | 0.70 | 0.46 |
| Serving all students needing services in mathematics | NA | NA | 0.54 | 0.50 | 0.64 | 0.48 |
| One-on-one tutoring | 0.80 | 0.40 | 0.66 | 0.47 | 0.74 | 0.44 |
| Special teacher assignments | 0.58 | 0.49 | NA | NA | 0.68 | 0.47 |
| Before- or after-school programs | 0.61 | 0.49 | 0.91 | 0.29 | 0.79 | 0.40 |
| Reduced class size | 0.38 | 0.49 | 0.37 | 0.48 | 0.44 | 0.50 |
| Hindrances | 1.98 | 0.63 | 2.15 | 0.67 | 1.97 | 0.56 |
| Level of support | 2.76 | 0.55 | 2.74 | 0.57 | 2.95 | 0.69 |
| Teacher quality | 3.35 | 0.51 | NA | NA | NA | NA |

**Table B.11**
**Influence of School Characteristics on the Odds of In-Need Students Scoring Level 2 or Higher on the Spring Assessment**

| | 2005–2006 | | 2006–2007 | | 2007–2008 | |
|---|---|---|---|---|---|---|
| Variable | ELA | Mathematics | ELA | Mathematics | ELA | Mathematics |
| Prior year's achievement | 1.04** (1.04, 1.05) | 1.02** (1.02, 1.03) | 1.04** (1.03, 1.04) | 1.03** (1.02, 1.03) | 1.04** (1.03, 1.05) | 1.03** (1.02, 1.03) |
| NCLB status | 0.85* (0.73, 0.98) | 0.80* (0.66, 0.97) | 0.85 (0.68, 1.07) | 1.03 (0.81, 1.31) | NA | NA |
| Percentage of students needing services | 0.97** (0.96, 0.98) | 0.99 (0.98, 1.00) | 0.99* (0.98, 1.00) | 0.99 (0.98, 1.00) | 0.97** (0.95, 0.99) | 0.99 (0.97, 1.01) |
| Percentage of students not held to policy | 1.00 (0.99, 1.01) | 1.00 (0.99, 1.01) | 0.99 (0.98, 1.01) | 0.99* (0.97, 0.99) | 0.98 (0.96, 1.00) | 0.98 (0.97, 1.00) |

* = statistically significant at the 0.05 level.

** = statistically significant at the 0.01 level.

**Table B.12**
**Influence of School Characteristics and Strategies on the Odds of In-Need Students Scoring Level 2 on the Spring Assessment, ELA and Mathematics, Logistic Model Estimates**

| | 2005–2006 | | 2006–2007 | | 2007–2008 | |
|---|---|---|---|---|---|---|
| **Variable** | **ELA** | **Mathematics** | **ELA** | **Mathematics** | **ELA** | **Mathematics** |
| Prior year's achievement | 1.04** (1.04, 1.05) | 1.03** (1.02, 1.04) | 1.03** (1.02, 1.04) | 1.03** (1.02, 1.04) | 1.05** (1.04, 1.06) | 1.02** (1.01, 1.03) |
| NCLB status | 0.91 (0.71, 1.16) | 0.85 (0.63, 1.15) | 0.94 (0.63, 1.42) | 0.83 (0.52, 1.34) | NA | NA |
| Percentage of students needing services | 0.97** (0.96, 0.99) | 0.99 (0.97, 1.01) | 0.99 (0.97, 1.02) | 0.99 (0.96, 1.01) | 0.98 (0.95, 1.02) | 0.97 (0.94, 1.01) |
| Percentage of students not held to policy | 1.00 (0.99, 1.02) | 1.00 (0.98, 1.01) | 1.00 (0.98, 1.02) | 0.99 (0.96, 1.01) | 0.97* (0.94, 0.99) | 0.99 (0.96, 1.01) |
| Serving all students needing services | 1.12 (0.91, 1.37) | 0.83 (0.64, 1.07) | 1.28 (0.87, 1.88) | 0.78 (0.53, 1.15) | 1.01 (0.60, 1.72) | 1.34 (0.83, 2.16) |
| One-on-one tutoring | 1.27 (0.98, 1.63) | 1.21 (0.87, 1.67) | 1.02 (0.70, 1.50) | 1.22 (0.82, 1.81) | 1.09 (0.59, 2.02) | 0.61 (0.34, 1.10) |
| Special teacher assignment | 1.01 (0.80, 1.28) | 1.25 (0.95, 1.64) | NA | NA | 0.67 (0.36, 1.25) | 1.18 (0.65, 2.13) |
| Reduced class size | 0.84 (0.67, 1.04) | 1.03 (0.78, 1.36) | 0.98 (0.66, 1.47) | 0.88 (0.61, 1.27) | 1.06 (0.64, 1.78) | 1.00 (0.63, 1.59) |
| Hindrances | 0.97 (0.82, 1.14) | 0.93 (0.76, 1.14) | 0.87 (0.64, 1.19) | 0.77* (0.59, 0.99) | 1.11 (0.72, 1.72) | 0.79 (0.56, 1.11) |
| Level of support | 0.77* (0.63, 0.94) | 0.85 (0.66, 1.11) | 0.79 (0.55, 1.13) | 0.96 (0.70, 1.33) | 0.94 (0.70, 1.28) | 1.25 (0.95, 1.64) |
| Teacher quality | 1.04 (0.85, 1.27) | 0.96 (0.74, 1.26) | 1.14 (0.81, 1.59) | 0.91 (0.61, 1.37) | NA | NA |
| Principal leadership | NA | NA | 1.19 (0.88, 1.61) | 1.00 (0.75, 1.34) | NA | NA |
| Number of observations | | | | | | |
| Students not passing (n) | 641 | 644 | 161 | 259 | 86 | 145 |
| Students passing (n) | 3,763 | 994 | 2,169 | 1,356 | 2,761 | 1,011 |

* = statistically significant at the 0.05 level.

** = statistically significant at the 0.01 level.

**Table B.13**
**Influence of School Characteristics and Strategies on the Student Performance on the Spring ELA and Mathematics Assessments, Students Needing Services, Regression Model Results**

| | 2005–2006 | | 2006–2007 | | 2007–2008 | |
|---|---|---|---|---|---|---|
| Variable | ELA | Mathematics | ELA | Mathematics | ELA | Mathematics |
| Prior year's achievement | 0.56** (0.04) | 0.30** (0.05) | 0.34** (0.04) | 0.34** (0.04) | 0.41** (0.03) | 0.20** (0.05) |
| NCLB status | –0.44 (1.18) | –0.45 (2.24) | –0.37 (1.18) | –0.53 (1.95) | NA | NA |
| Percentage of students needing services | –0.38** (0.08) | –0.13 (0.14) | –0.19** (0.07) | –0.08 (0.11) | –0.16* (0.08) | –0.14 (0.13) |
| Percentage of students not held to policy | 0.02 (0.06) | –0.07 (0.11) | 0.02 (0.06) | –0.01 (0.10) | –0.09* (0.06) | –0.15 (0.10) |
| Serving all students needing services | –0.18 (0.93) | 0.36 (1.70) | 1.28 (0.98) | –0.17 (1.57) | 0.84 (0.97) | 1.83 (1.76) |
| One-on-one tutoring | 1.64 (1.10) | 3.83 (2.19) | –0.01 (0.99) | 3.84* (1.74) | –0.75 (1.23) | –0.48 (1.95) |
| Special teacher assignments | –0.23 (0.99) | 3.30 (1.75) | NA | NA | 1.29 (1.03) | 1.98 (1.98) |
| Reduced class size | –0.61 (1.01) | 1.30 (2.00) | –0.13 (1.01) | –2.32 (1.82) | –0.12 (0.91) | –1.77 (1.71) |
| Hindrances | –0.85 (0.77) | –1.92 (1.36) | –1.63 (0.87) | –1.82 (1.14) | –1.15* (0.70) | –3.56** (1.26) |
| Level of support | –2.94** (1.10) | –2.91 (2.48) | –1.82* (0.92) | –0.26 (1.53) | 0.13 (0.59) | 2.14* (0.98) |
| Teacher quality | 0.30 (1.00) | 0.11 (1.61) | 1.17 (0.96) | –0.93 (1.89) | NA | NA |
| Principal leadership | NA | NA | –1.27 (0.74) | 0.06 (1.21) | NA | NA |
| Intercept | 296.42 | 447.47 | 436.81 | 438.73 | 390.69 | 519.56 |
| Number of observations | 4,404 | 1,638 | 2,330 | 1,615 | 2,847 | 1,156 |
| Adjusted R-squared | 0.13 | 0.07 | 0.11 | 0.08 | 0.15 | 0.05 |

* = statistically significant at the 0.05 level.

** = statistically significant at the 0.01 level.

**Table B.14**
**Influence of School Characteristics and Strategies on the Student Performance on the Spring ELA and Mathematics Assessments, All Students Held to Policy, Regression Model Results**

| | 2005–2006 | | 2006–2007 | | 2007–2008 | |
|---|---|---|---|---|---|---|
| **Variable** | **ELA** | **Mathematics** | **ELA** | **Mathematics** | **ELA** | **Mathematics** |
| Prior year's achievement | 0.63** (0.01) | 0.73** (0.01) | 0.58** (0.01) | 0.69** (0.01) | 0.54** (0.01) | 0.66** (0.01) |
| NCLB status | –0.31 (0.92) | –1.36 (1.55) | –1.64 (1.09) | –0.26 (1.48) | NA | NA |
| Percentage of students needing services | –0.68** (0.08) | –0.69 (0.13) | –0.25** (0.05) | –0.25** (0.08) | –0.31** (0.06) | –0.36** (0.10) |
| Percentage of students not held to policy | 0.01 (0.04) | –0.03 (0.06) | –0.07 (0.04) | 0.01 (0.06) | –0.09* (0.04) | –0.20** (0.06) |
| Serving all students needing services | –0.19 (0.69) | 0.54 (1.00) | 1.14 (0.75) | 0.24 (0.98) | 0.94 (0.66) | 0.34 (1.01) |
| One-on-one tutoring | 0.97 (0.83) | 1.44 (1.30) | –0.50 (0.82) | 2.91** (1.06) | 0.18 (0.66) | 1.33 (1.11) |
| Special teacher assignments | –0.20 (0.75) | 0.06 (1.03) | NA | NA | –0.31 (0.69) | –0.35 (1.19) |
| Reduced class size | 0.49 (0.80) | –0.05 (1.17) | 0.21 (0.88) | –0.19 (1.09) | 0.59 (0.71) | –0.16 (1.05) |
| Hindrances | –1.15 (0.63) | –1.09 (0.90) | –0.65 (0.58) | –1.56* (0.71) | –1.58** (0.56) | –2.78** (0.83) |
| Level of support | –1.03 (0.82) | –1.03 (1.19) | –1.59* (0.64) | –2.88** (0.87) | 0.46 (0.43) | –0.30 (0.63) |
| Teacher quality | 0.06 (0.79) | 0.70 (1.11) | 1.81* (0.79) | 0.94 (1.13) | NA | NA |
| Principal leadership | NA | NA | –0.65 (0.69) | 0.26 (0.96) | NA | NA |
| Intercept | 251.75 | 184.37 | 283.45 | 217.00 | 315.71 | 241.89 |
| Number of observations | 26,371 | 28,308 | 17,564 | 18,973 | 20,699 | 20,758 |
| Adjusted R-squared | 0.52 | 0.58 | 0.47 | 0.60 | 0.47 | 0.58 |

* = statistically significant at the 0.05 level.

** = statistically significant at the 0.01 level.

APPENDIX C

# Supporting Data for Chapter Seven

The appendix provides supporting tables and figures for Chapter Seven, which addresses the composition, status, and performance of 5th-grade students in our four NYC analysis cohorts, as well as performance trends in grades 3–8 from 2005–2006 through 2007–2008 in NYC and the rest of New York.

**Table C.1**
**Performance of 5th Graders Held to Policy on the 4th-Grade Spring Assessments, by Subject**

| Performance and Retention Status | 2003–2004[a] | | 2004–2005 | | 2005–2006 | | 2006–2007 | |
|---|---|---|---|---|---|---|---|---|
| | Number | Percentage | Number | Percentage | Number | Percentage | Number | Percentage |
| ELA | | | | | | | | |
| Retained from prior year[b] | 1,175 | 1.8 | 1,057 | 1.7 | 1,575 | 2.7 | 1,540 | 2.7 |
| Level 1 | 2,158 | 3.2 | 1,657 | 2.6 | 938 | 1.6 | 2,362 | 4.1 |
| Low Level 2 | 10,515 | 15.8 | 10,698 | 17.0 | 7,542 | 12.9 | 4,519 | 7.8 |
| High Level 2 | 11,864 | 17.8 | 12,255 | 19.4 | 7,755 | 13.2 | 9,413 | 16.3 |
| Level 3 or Level 4 | 35,864 | 53.9 | 32,800 | 52.0 | 35,074 | 59.8 | 33,911 | 58.7 |
| Data missing | 4,968 | 7.5 | 4,634 | 7.3 | 5,811 | 9.9 | 6,017 | 10.4 |
| Total held to policy | 66,544 | 100.0 | 63,101 | 100.0 | 58,695 | 100.0 | 57,762 | 100.0 |
| Mathematics | | | | | | | | |
| Retained from prior year[b] | 1,175 | 1.8 | 1,057 | 1.7 | 1,575 | 2.7 | 1,540 | 2.7 |
| Level 1 | 2,061 | 3.1 | 1,217 | 1.9 | 507 | 0.9 | 2,090 | 3.6 |
| Low Level 2 | 4,908 | 7.4 | 3,835 | 6.1 | 1,609 | 2.7 | 1,979 | 3.4 |
| High Level 2 | 8,499 | 12.8 | 8,736 | 13.8 | 5,232 | 8.9 | 6,578 | 11.4 |
| Level 3 or Level 4 | 46,355 | 69.7 | 46,034 | 73.0 | 47,821 | 81.5 | 43,640 | 75.6 |
| Data missing | 3,546 | 5.3 | 2,222 | 3.5 | 1,951 | 3.3 | 1,935 | 3.4 |
| Total held to policy | 66,544 | 100.0 | 63,101 | 100.0 | 58,695 | 100.0 | 57,762 | 100.0 |

[a] The comparison cohort was not held to the policy. For this cohort, the number of students held to policy was estimated using policy guidelines.

[b] These were students who had been in the 5th grade in the prior year and were retained in grade, forming part of the current cohort.

**Table C.2**
**Students Held to Policy Who Needed Services at the Beginning of the School Year, by Subject and Cohort**

| Subject and Retention Status | Students Needing Services at the Beginning of the School Year[a] (% of total students held to policy) | | | |
|---|---|---|---|---|
| | C0 Cohort | P1 Cohort | P2 Cohort | P3 Cohort |
| ELA only | 12.3 | 14.5 | 12.4 | 8.8 |
| Mathematics only | 3.8 | 2.9 | 1.6 | 3.9 |
| Both ELA and mathematics | 6.7 | 5.1 | 2.0 | 3.1 |
| Retained in 5th grade | 1.8 | 1.7 | 2.7 | 2.7 |
| Total needing services | 24.6 | 24.2 | 18.7 | 18.5 |
| Total held to policy | 66,544 | 63,101 | 58,695 | 57,762 |

NOTE: The comparison cohort was not held to the policy. For this cohort, the number of students held to the policy was estimated using policy guidelines.

[a] Defined as those who scored Level 1 or low Level 2 on the 4th-grade assessments and retained 5th graders.

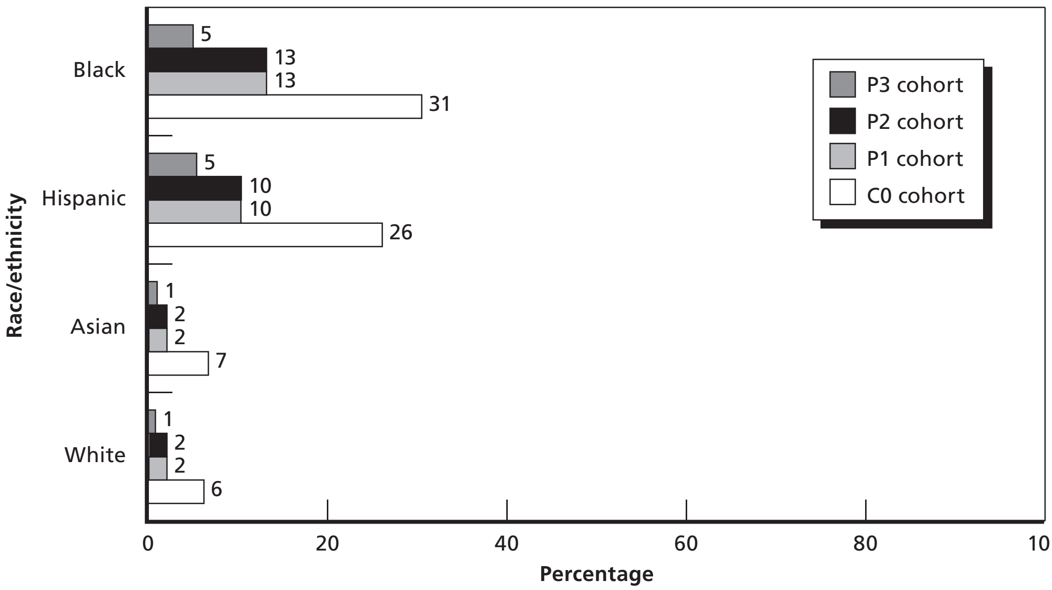
**Table C.3**
**Students Held to Policy Who Needed Services at the Beginning of the School Year, by Race/Ethnicity and Cohort**

| | Students Needing Services at the Beginning of the School Year[a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | C0 Cohort | | P1 Cohort | | P2 Cohort | | P3 Cohort | |
| **Selected Characteristics** | **Number** | **Percentage** | **Number** | **Percentage** | **Number** | **Percentage** | **Number** | **Percentage** |
| Race/ethnicity | | | | | | | | |
| American Indian | 107 | 31.0 | 97 | 29.4 | 50 | 17.6 | 59 | 22.8 |
| Asian | 817 | 9.9 | 736 | 8.7 | 520 | 6.2 | 536 | 6.2 |
| Hispanic | 7,515 | 29.9 | 6,980 | 29.4 | 4,992 | 22.6 | 4,849 | 22.6 |
| Black | 7,074 | 30.3 | 6,702 | 31.1 | 4,926 | 25.4 | 4,663 | 25.4 |
| White | 792 | 8.4 | 710 | 7.9 | 483 | 5.7 | 568 | 6.4 |
| ELL status | | | | | | | | |
| ELL | 1,930 | 63.4 | 1,577 | 50.7 | 1,320 | 32.9 | 1,409 | 38.0 |
| Non-ELL | 14,426 | 22.7 | 13,669 | 22.8 | 9,671 | 17.7 | 9,278 | 17.2 |

NOTE: The comparison cohort was not held to policy. For this cohort, the number of students held to the policy was estimated using policy guidelines.

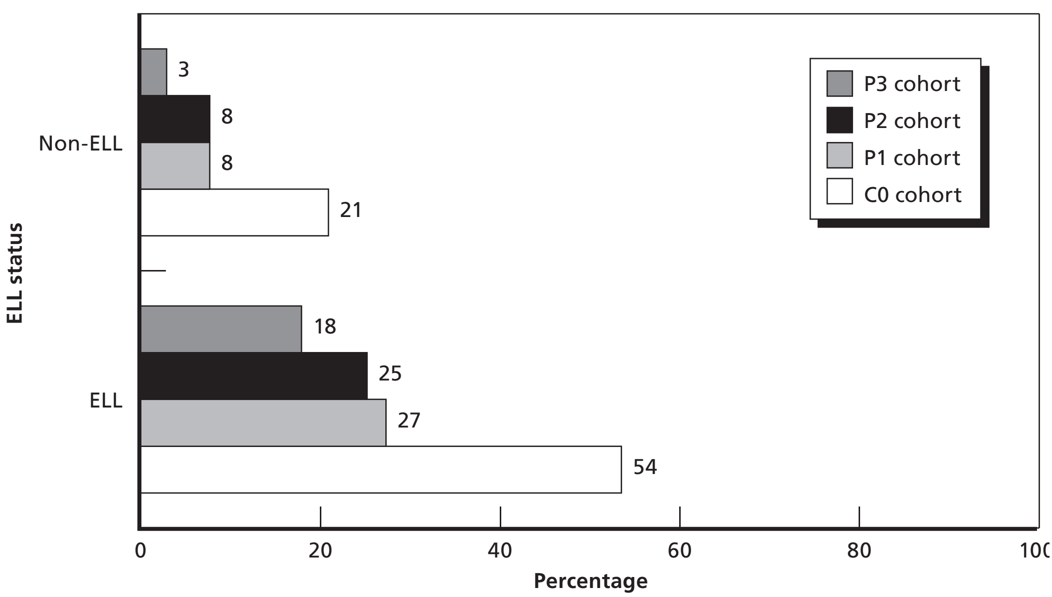
[a] Defined as those who scored Level 1 or low Level 2 on the 4th-grade assessments and retained 5th graders.

**Figure C.1**
**Percentage of Students Who Scored Level 1 on the Proximal-Year 5th-Grade Spring Assessments in ELA and/or Mathematics, by Race/Ethnicity**



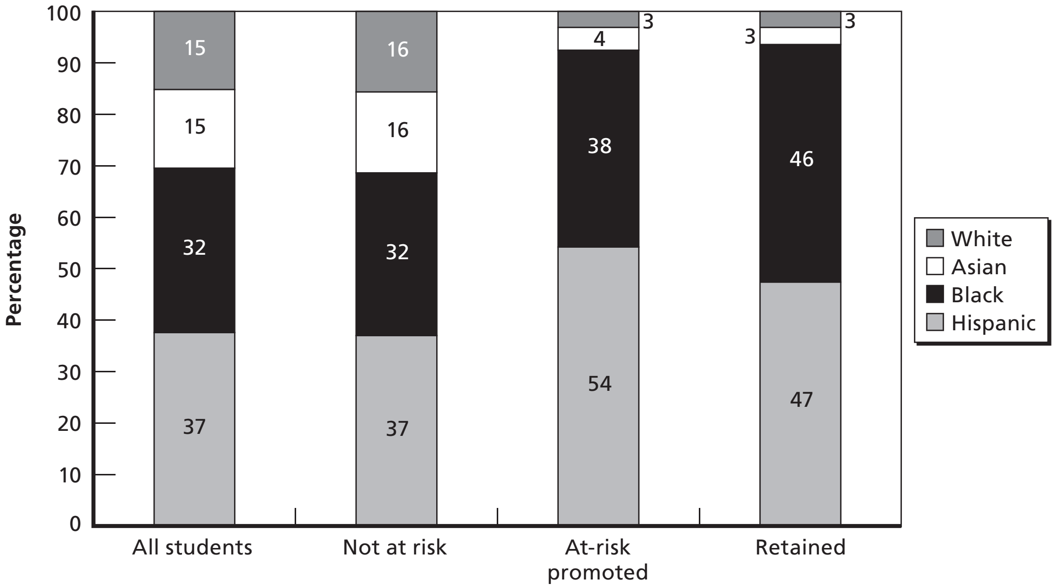

RAND *MG894-C.1*

**Figure C.2**
**Percentage of Students Who Scored Level 1 on the Proximal-Year 5th-Grade Spring Assessments in ELA and/or Mathematics, by ELL Status**



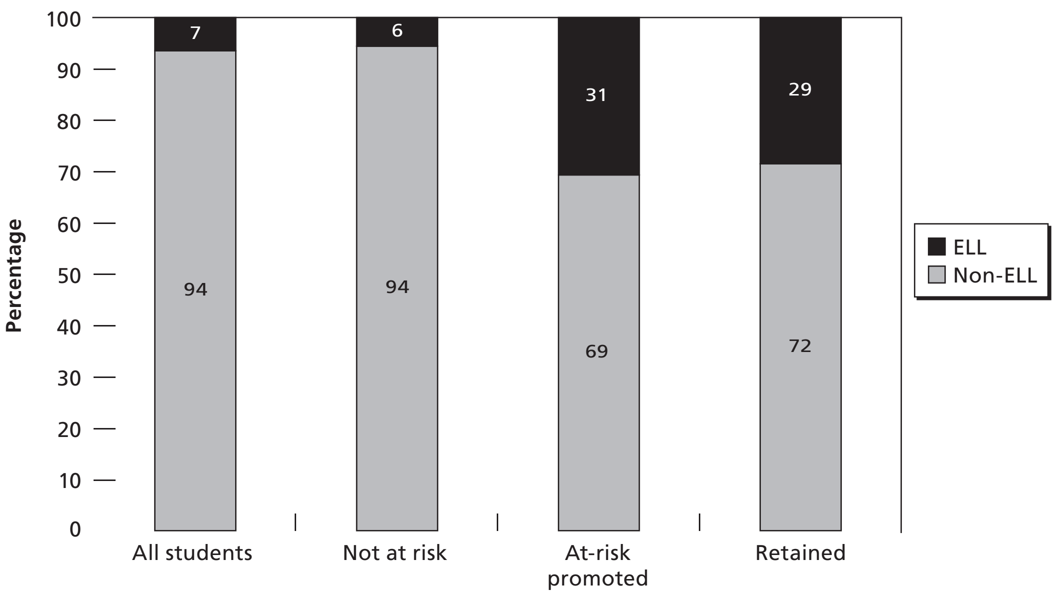

RAND *MG894-C.2*

**Figure C.3**
**Racial/Ethnic Profile of Students, by Final Promotion or Retention Status, 2006–2007 (P3 Cohort)**



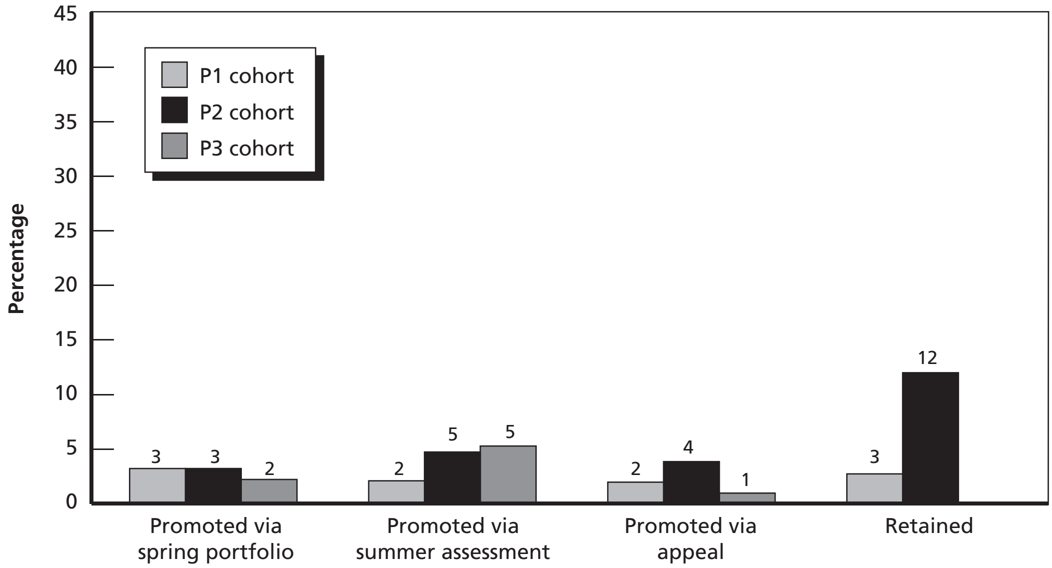

RAND *MG894-C.3*

**Figure C.4**
**ELL Status of Students, by Final Promotion or Retention Status, 2006–2007 (P3 Cohort)**



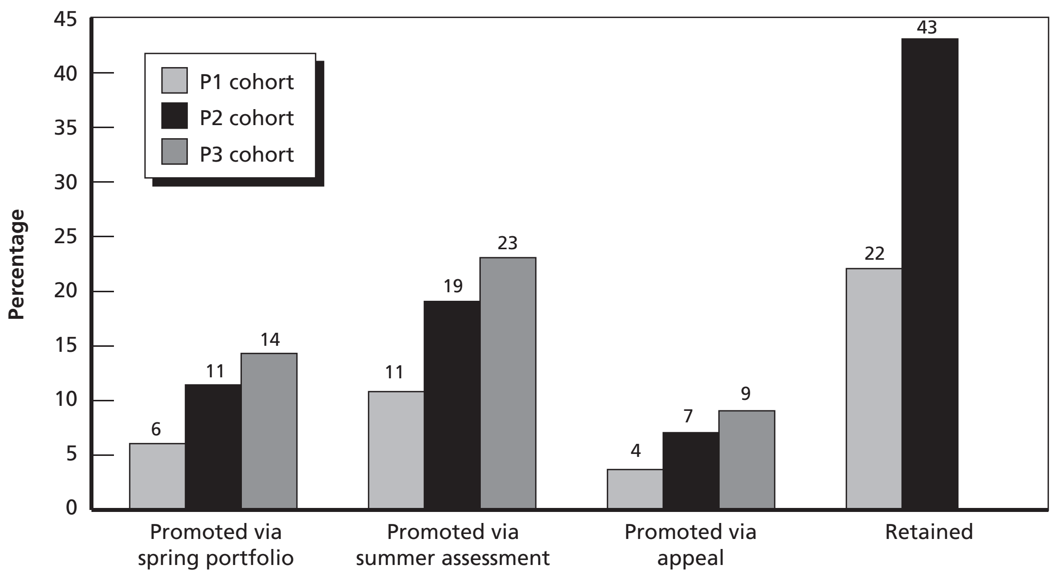

RAND *MG894-C.4*

**Figure C.5**
**Percentage of At-Risk Promoted and Retained Students Scoring Level 3 or Higher on the 6th-Grade ELA Assessment**



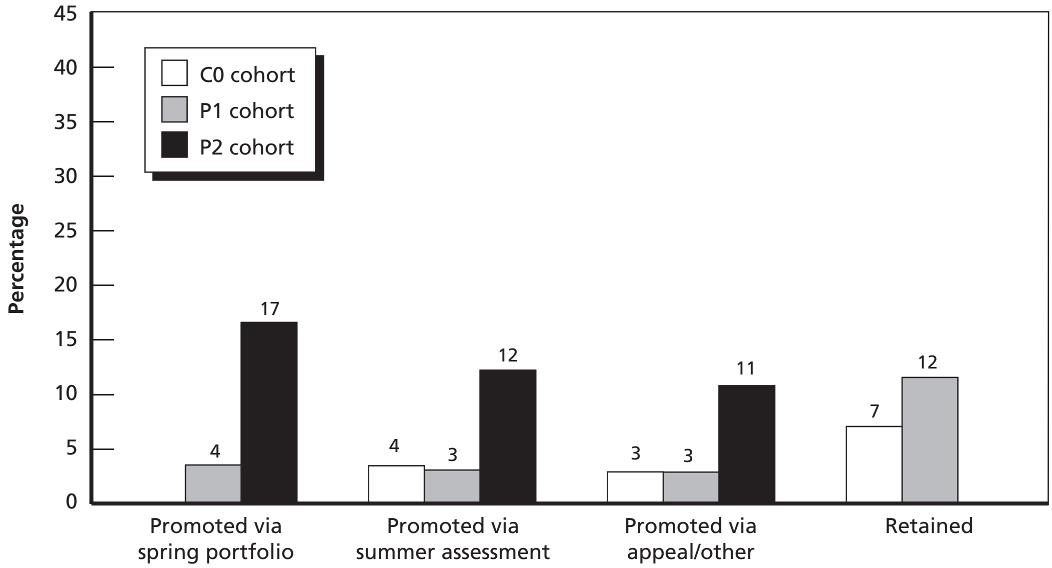

RAND *MG894-C.5*

**Figure C.6**
**Percentage of At-Risk Promoted and Retained Students Scoring Level 3 or Higher on the 6th-Grade Mathematics Assessment**



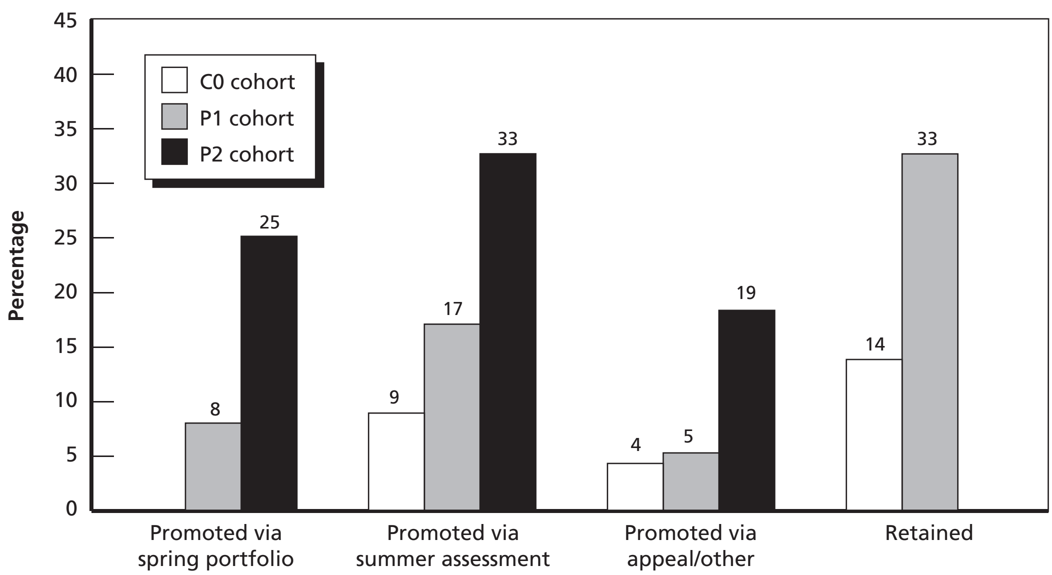

RAND *MG894-C.6*

**Figure C.7**
**Percentage of At-Risk Promoted and Retained Students Scoring Level 3 or Higher on the 7th-Grade ELA Assessment**



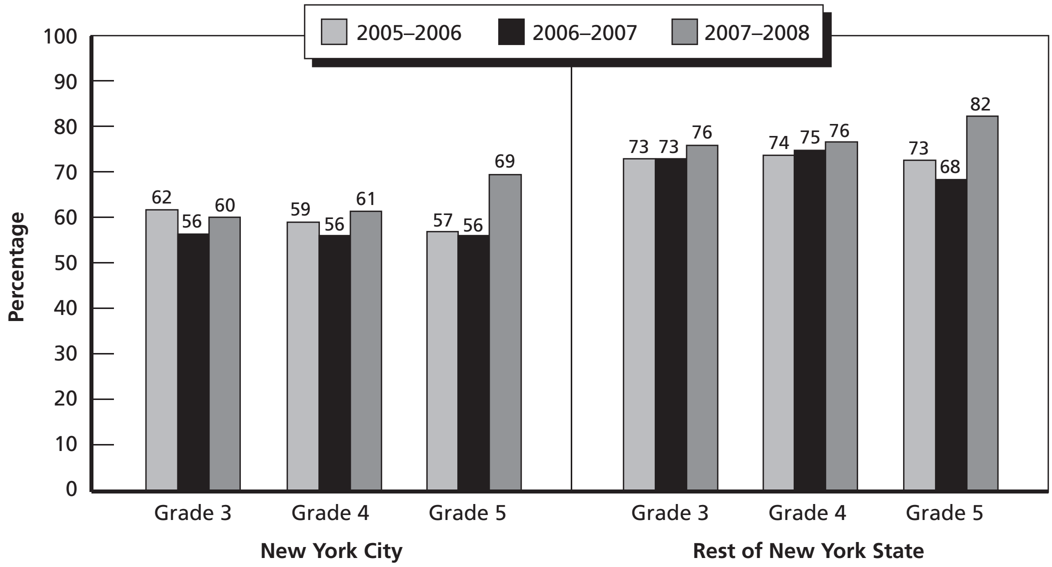

RAND *MG894-C.7*

**Figure C.8**
**Percentage of At-Risk Promoted and Retained Students Scoring Level 3 or Higher on the 7th-Grade Mathematics Assessment**



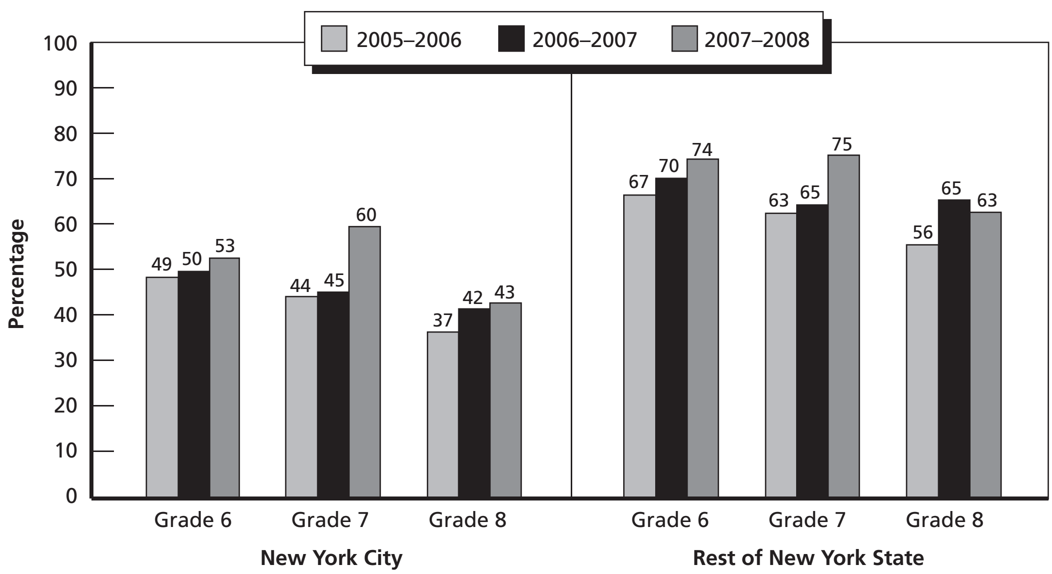

RAND *MG894-C.8*

**Figure C.9**
**Percentage of Students Scoring Level 3 or Higher on the Spring ELA Assessments, Grades 3–5, NYC and Rest of State**



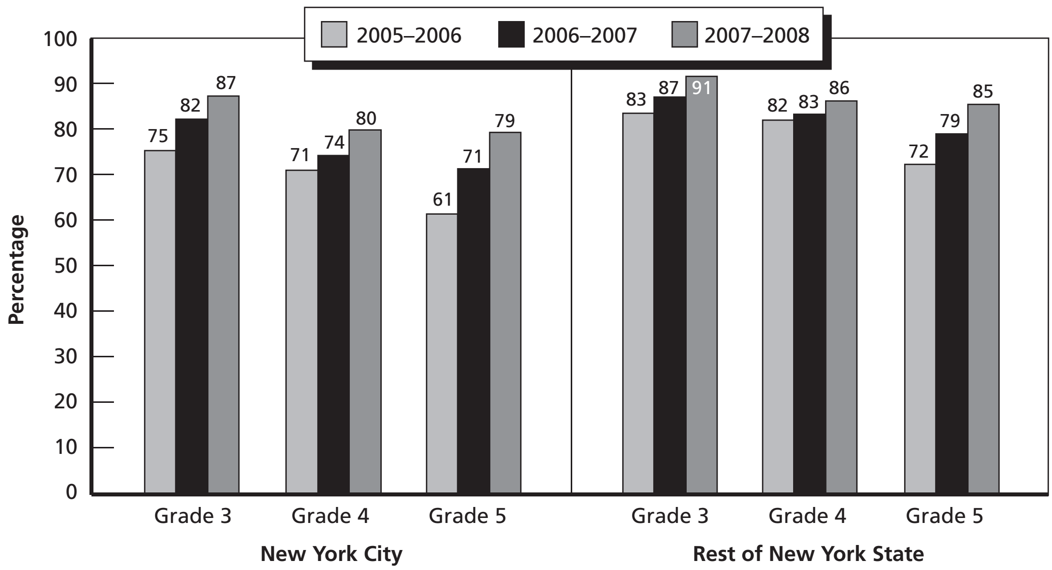

**RAND** *MG894-C.9*

**Figure C.10**
**Percentage of Students Scoring Level 3 or Higher on the Spring ELA Assessments, Grades 6–8, NYC and Rest of State**



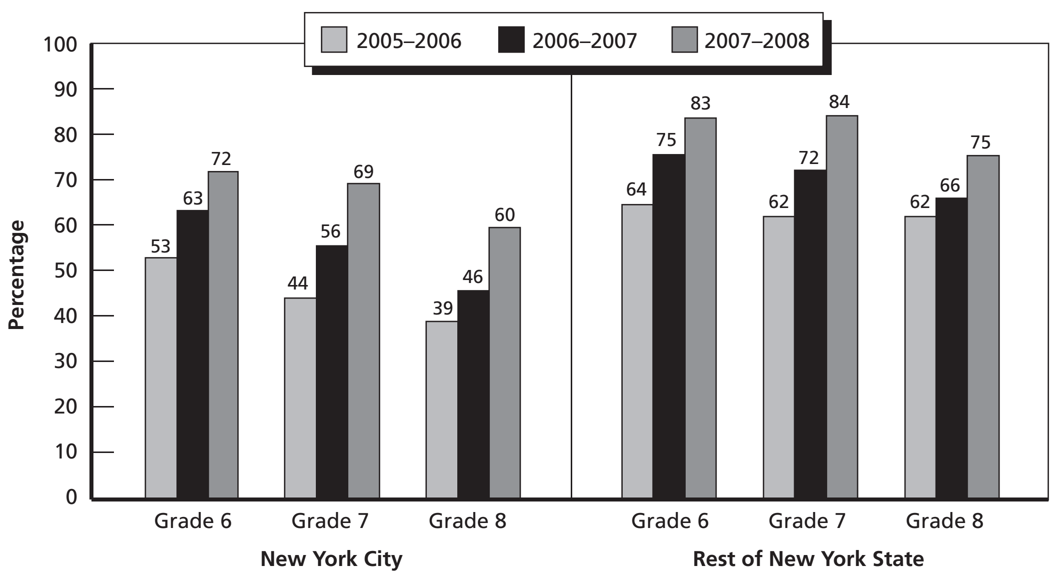

**RAND** *MG894-C.10*

**Figure C.11**
**Percentage of Students Scoring Level 3 or Higher on the Spring Mathematics Assessments, Grades 3–5, NYC and Rest of State**




RAND *MG894-C.11*

**Figure C.12**
**Percentage of Students Scoring Level 3 or Higher on the Spring Mathematics Assessments, Grades 6–8, NYC and Rest of State**




RAND *MG894-C.12*

APPENDIX D

# Supporting Data for Chapter Ten

This appendix presents supporting tables for Chapter Ten, which used responses to our student survey to examine students' socioemotional outcomes. That chapter sought to answer the following research questions:

1. Are there differences among retained, at-risk promoted, and not-at-risk students with respect to school belonging or self-confidence in mathematics and reading? Do these differences change over time?
2. After controlling for student-level characteristics, is retention status related to students' later socioemotional outcomes?
3. Among at-risk students, are socioemotional responses related to final promotion outcomes?
4. Are students' socioemotional responses predictive of future retention status?

**Table D.1**
**Descriptive Statistics for the 5th-Grade Cohort**

| Group | Variable | Number | Minimum | First Quartile | Mean | Median | Third Quartile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|
| Not at risk, 1 year after | School belonging | 7,527 | 1.00 | 2.60 | 2.97 | 2.99 | 3.33 | 4.00 | 1.52 |
| | Mathematics confidence | 7,527 | 1.00 | 2.60 | 3.23 | 3.31 | 3.79 | 4.00 | 1.32 |
| | Reading confidence | 7,527 | 1.00 | 2.80 | 3.30 | 3.31 | 3.72 | 4.00 | 1.50 |
| At-risk promoted, 1 year after | School belonging | 412 | 1.39 | 2.60 | 2.93 | 2.98 | 3.25 | 3.94 | 0.84 |
| | Mathematics confidence | 412 | 1.00 | 2.13 | 2.90 | 2.85 | 3.58 | 4.00 | 1.22 |
| | Reading confidence | 412 | 1.25 | 2.48 | 3.12 | 3.04 | 3.62 | 4.00 | 1.02 |
| Retained, 1 year after | School belonging | 199 | 1.78 | 2.66 | 3.01 | 2.99 | 3.36 | 3.83 | 0.44 |
| | Mathematics confidence | 199 | 1.00 | 2.10 | 2.82 | 2.79 | 3.38 | 4.00 | 0.95 |
| | Reading confidence | 199 | 1.00 | 2.71 | 3.33 | 3.35 | 3.77 | 4.00 | 0.61 |
| Not at risk, 2 years after | School belonging | 6,147 | 1.00 | 2.49 | 2.87 | 2.87 | 3.24 | 4.00 | 1.57 |
| | Mathematics confidence | 6,147 | 1.00 | 2.59 | 3.22 | 3.27 | 3.79 | 4.00 | 1.93 |
| | Reading confidence | 6,147 | 1.00 | 2.80 | 3.32 | 3.31 | 3.72 | 4.00 | 1.35 |
| At-risk promoted, 2 years after | School belonging | 333 | 1.00 | 2.55 | 2.90 | 2.90 | 3.24 | 3.94 | 0.84 |
| | Mathematics confidence | 333 | 1.00 | 2.30 | 2.91 | 2.82 | 3.45 | 4.00 | 0.81 |
| | Reading confidence | 333 | 1.25 | 2.68 | 3.25 | 3.22 | 3.66 | 4.00 | 0.78 |

Table D.1—Continued

| Group | Variable | Number | Minimum | First Quartile | Mean | Median | Third Quartile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|
| Retained, 2 years after | School belonging | 291 | 1.61 | 2.77 | 3.05 | 3.08 | 3.37 | 4.00 | 0.55 |
| | Mathematics confidence | 291 | 1.00 | 2.27 | 2.95 | 2.95 | 3.50 | 4.00 | 1.04 |
| | Reading confidence | 291 | 1.00 | 2.72 | 3.27 | 3.28 | 3.76 | 4.00 | 0.85 |
| Not at risk, 3 years after | School belonging | 2,696 | 1.00 | 2.49 | 2.86 | 2.87 | 3.22 | 4.00 | 1.41 |
| | Mathematics confidence | 2,696 | 1.00 | 2.54 | 3.19 | 3.24 | 3.77 | 4.00 | 2.07 |
| | Reading confidence | 2,696 | 1.00 | 2.82 | 3.31 | 3.30 | 3.69 | 4.00 | 1.24 |
| At-risk promoted, 3 years after | School belonging | 160 | 1.28 | 2.46 | 2.80 | 2.85 | 3.11 | 3.78 | 0.58 |
| | Mathematics confidence | 160 | 1.00 | 2.12 | 2.83 | 2.76 | 3.38 | 4.00 | 0.84 |
| | Reading confidence | 160 | 1.00 | 2.78 | 3.28 | 3.20 | 3.55 | 4.00 | 0.39 |
| Retained, 3 years after | School belonging | 152 | 1.78 | 2.66 | 2.98 | 2.99 | 3.28 | 3.94 | 0.48 |
| | Mathematics confidence | 152 | 1.00 | 2.39 | 3.07 | 3.12 | 3.61 | 4.00 | 0.73 |
| | Reading confidence | 152 | 1.00 | 2.69 | 3.18 | 3.10 | 3.61 | 4.00 | 0.64 |

**Table D.2**
**Standardized Regression Coefficients for the 5th-Grade Cohort One to Three Years After the Retention Decision**

| Variable | 1 Year | 2 Years | 3 Years |
|---|---|---|---|
| School belonging | | | |
| Intercept | −0.06 (0.25) | 0.42 (0.10)** | 0.28 (0.13) |
| Free-lunch eligibility | 0.17 (0.25) | 0.01 (0.06) | 0.04 (0.12) |
| Reduced-price-lunch eligibility | 0.17 (0.26) | −0.06 (0.13) | 0.01 (0.13) |
| English language learner | 0.00 (0.10) | 0.02 (0.07) | 0.07 (0.08) |
| Male | −0.12 (0.04)** | −0.05 (0.05) | −0.10 (0.06) |
| Days absent | −0.14 (0.03)** | −0.03 (0.02) | −0.04 (0.03) |
| Special education student | NA | 0.02 (0.11) | NA |
| Black | −0.14 (0.08) | −0.24 (0.05)** | −0.01 (0.06) |
| White | 0.24 (0.09)* | 0.00 (0.11) | 0.05 (0.06) |
| Other race | 0.06 (0.09) | −0.24 (0.09)** | −0.21 (0.15) |
| At-risk promoted | −0.09 (0.13) | −0.29 (0.09)** | −0.34 (0.11)** |
| Not at risk | −0.06 (0.09) | −0.36 (0.08)** | −0.23 (0.09)** |
| Mathematics confidence | | | |
| Intercept | −0.58 (0.14)** | −0.34 (0.10)** | −0.15 (0.11) |
| Free-lunch eligibility | 0.09 (0.11) | −0.09 (0.07) | −0.08 (0.07) |
| Reduced-price-lunch eligibility | 0.04 (0.13) | −0.13 (0.07) | 0.00 (0.09) |
| English language learner | −0.07 (0.08) | −0.05 (0.05) | 0.05 (0.07) |
| Male | 0.28 (0.03) | 0.24 (0.04)** | 0.21 (0.05)** |
| Days absent | −0.10 (0.02)** | −0.09 (0.02)** | −0.09 (0.03)** |
| Special education student | NA | −0.09 (0.10) | NA |
| Black | 0.02 (0.07) | 0.08 (0.05) | 0.11 (0.07) |
| White | 0.11 (0.13) | 0.08 (0.09) | 0.07 (0.08) |
| Other race | 0.24 (0.07)** | 0.14 (0.08) | 0.29 (0.07)** |
| At-risk promoted | −0.04 (0.14) | −0.06 (0.10) | −0.33 (0.12)** |
| Not at risk | 0.39 (0.11)** | 0.30 (0.08)** | 0.09 (0.09) |

**Table D.2—Continued**

| Variable | 1 Year | 2 Years | 3 Years |
|---|---|---|---|
| Reading confidence | | | |
| Intercept | 0.14 (0.11) | 0.15 (0.11) | –0.03 (0.11) |
| Free-lunch eligibility | –0.06 (0.09) | –0.11 (0.06) | –0.14 (0.07) |
| Reduced-price-lunch eligibility | –0.12 (0.12) | –0.01 (0.07) | –0.05 (0.14) |
| English language learner | –0.22 (0.07)** | –0.29 (0.05)** | –0.22 (0.08)** |
| Male | –0.13 (0.03)** | –0.10 (0.04)* | –0.19 (0.05)** |
| Days absent | –0.03 (0.03) | 0.02 (0.02) | 0.00 (0.02) |
| Special education student | NA | –0.28 (0.08)** | NA |
| Black | 0.17 (0.05)** | 0.05 (0.06) | 0.27 (0.05)** |
| White | 0.10 (0.10) | 0.24 (0.07)** | 0.27 (0.07)** |
| Other race | 0.06 (0.10) | 0.01 (0.06) | 0.20 (0.12) |
| At-risk promoted | –0.47 (0.13)** | –0.04 (0.11) | 0.10 (0.10) |
| Not at risk | –0.07 (0.49) | 0.02 (0.09) | 0.14 (0.09) |

NOTE: Some held-to-policy students were referred to special education in later years; when such students were captured in the sample, we controlled for special education status in the model. Standard errors are in parentheses.

* = statistically significant at the 0.04 level.

** = statistically significant at the 0.01 level.

**Table D.3**
**Standardized Regression Coefficients by Promotion Type, One to Four Years After the Retention Decision**

| Variable | 1 Year | 2 Years | 3 Years | 4 Years |
|---|---|---|---|---|
| School belonging | | | | |
| Intercept | 0.51 (0.24)* | 0.52 (0.40) | 0.09 (0.20) | 0.51 (0.19)** |
| Free-lunch eligibility | −0.36 (0.23) | −0.44 (0.38) | 0.11 (0.18) | −0.43 (0.19)* |
| Reduced-price-lunch eligibility | −0.48 (0.32) | 0.09 (0.38) | −0.02 (0.29) | −0.26 (0.24) |
| English language learner | −0.07 (0.13) | 0.11 (0.17) | 0.12 (0.10) | 0.18 (0.09) |
| Male | −0.04 (0.11) | 0.12 (0.13) | −0.03 (0.07) | 0.06 (0.06) |
| Special education student | NA | 0.09 (0.31) | 0.28 (0.13)* | 1.71 (0.14)** |
| Days absent | −0.12 (0.08) | −0.05 (0.06) | −0.06 (0.04) | −0.04 (0.04) |
| Black | −0.19 (0.12) | 0.06 (0.20) | −0.08 (0.06) | −0.08 (0.07) |
| White | −0.32 (0.23) | 0.22 (0.36) | 0.47 (0.24) | 0.07 (0.14) |
| Other race | 0.30 (0.25) | −0.18 (0.21) | 0.25 (0.14) | −0.25 (0.15) |
| Spring portfolio | −0.40 (0.25) | −0.29 (0.28) | −0.18 (0.12) | −0.25 (0.12) |
| Summer performance | −0.03 (0.14) | −0.47 (0.20)* | −0.33 (0.07)** | −0.16 (0.07)* |
| Appeal | 0.02 (0.18) | −0.58 (0.22)* | −0.44 (0.14)** | −0.27 (0.11)* |
| Mathematics confidence | | | | |
| Intercept | 0.47 (0.32) | 0.19 (0.39) | −0.33 (0.20) | 0.05 (0.18) |
| Free-lunch eligibility | −0.66 (0.32)* | −0.48 (0.35) | 0.16 (0.18) | −0.19 (0.17) |
| Reduced-price-lunch eligibility | −0.09 (0.43) | −0.12 (0.50) | −0.12 (0.25) | −0.23 (0.19) |
| English language learner | 0.14 (0.14) | 0.15 (0.16) | 0.00 (0.11) | −0.05 (0.10) |
| Male | 0.37 (0.12) | 0.18 (0.16) | 0.33 (0.06)** | 0.34 (0.05)** |
| Special education student | NA | 0.30 (0.37) | −0.22 (0.38) | 1.09 (0.13)** |
| Days absent | −0.12 (0.04) | −0.10 (0.07) | −0.06 (0.03) | −0.06 (0.03) |
| Black | 0.01 (0.11) | 0.15 (0.14) | 0.21 (0.08)** | 0.01 (0.07) |
| White | 0.49 (0.61) | 0.71 (0.41) | −0.01 (0.52) | 0.08 (0.18) |
| Other race | 0.33 (0.22) | −0.59 (0.22)** | 0.06 (0.16) | 0.02 (0.18) |
| Spring portfolio | −0.06 (0.16) | 0.67 (0.23)** | −0.19 (0.13) | −0.01 (0.11) |
| Summer performance | −0.06 (0.14) | 0.14 (0.17) | −0.13 (0.07) | −0.03 (0.06) |

**Table D.3—Continued**

| Variable | 1 Year | 2 Years | 3 Years | 4 Years |
|---|---|---|---|---|
| Appeal | −0.11 (0.19) | −0.11 (0.24) | −0.26 (0.13) | −0.24 (0.12) |
| Reading confidence | | | | |
| Intercept | 0.32 (0.35) | −0.41 (0.42) | 0.02 (0.14) | 0.19 (0.17) |
| Free-lunch eligibility | −0.14 (0.30) | 0.33 (0.41) | −0.08 (0.10) | −0.32 (0.15)* |
| Reduced-price-lunch eligibility | −0.89 (0.49) | 0.46 (0.60) | −0.07 (0.21) | −0.37 (0.21) |
| English language learner | −0.19 (0.22) | −0.19 (0.19) | 0.11 (0.08) | −0.02 (0.11) |
| Male | −0.12 (0.14) | 0.21 (0.13) | −0.03 (0.06) | −0.05 (0.05) |
| Special education student | NA | −0.42 (0.23) | 0.47 (0.15)** | 0.78 (0.14)** |
| Days absent | −0.05 (0.06) | 0.05 (0.07) | 0.01 (0.04) | 0.00 (0.03) |
| Black | 0.24 (0.14) | 0.23 (0.19) | 0.15 (0.07)* | 0.23 (0.07)** |
| White | −0.14 (0.38) | 1.28 (0.70) | 0.17 (0.18) | 0.46 (0.14)** |
| Other race | −0.01 (0.19) | −0.21 (0.28) | −0.15 (0.24) | −0.16 (0.19) |
| Spring portfolio | −0.44 (0.17)* | 0.24 (0.20) | −0.01 (0.13) | 0.12 (0.11) |
| Summer performance | −0.49 (0.15)** | 0.06 (0.17) | −0.03 (0.07) | 0.04 (0.05) |
| Appeal | −0.18 (0.13) | −0.02 (0.23) | −0.29 (0.10)** | 0.08 (0.13) |

NOTE: Years 1 and 2 used data from the 5th-grade cohorts; years 3 and 4 used data from the 3rd-grade cohorts. Some held-to-policy students were referred to special education in later years; when such students were captured in the sample, we controlled for special education status in the model. Standard errors are in parentheses.

* = statistically significant at the 0.04 level.

** = statistically significant at the 0.01 level.

**Table D.4**
**Descriptive Statistics for the 2006–2007 5th-Grade Cohort**

| Group | Variable | Number | Minimum | First Quartile | Mean | Median | Third Quartile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|
| Not at risk, before decision | School belonging | 3,104 | 1.17 | 2.62 | 2.98 | 3.01 | 3.35 | 4.00 | 1.18 |
| | Mathematics confidence | 3,104 | 1.00 | 2.58 | 3.20 | 3.26 | 3.79 | 4.00 | 1.33 |
| | Reading confidence | 3,104 | 1.00 | 2.81 | 3.31 | 3.34 | 3.75 | 4.00 | 0.97 |
| At-risk promoted, before decision | School belonging | 179 | 1.67 | 2.67 | 2.97 | 3.01 | 3.29 | 3.83 | 0.59 |
| | Mathematics confidence | 179 | 1.00 | 2.28 | 2.91 | 2.83 | 3.41 | 4.00 | 1.36 |
| | Reading confidence | 179 | 1.00 | 2.61 | 3.05 | 3.02 | 3.45 | 4.00 | 0.72 |
| Retained, before decision | School belonging | 67 | 1.72 | 2.48 | 2.84 | 2.80 | 3.15 | 3.89 | 0.49 |
| | Mathematics confidence | 67 | 1.00 | 2.07 | 2.72 | 2.56 | 3.21 | 4.00 | 0.76 |
| | Reading confidence | 67 | 1.25 | 2.48 | 3.11 | 3.08 | 3.75 | 4.00 | 0.57 |
| Not at risk, 2 years after | School belonging | 4,452 | 1.00 | 2.52 | 2.90 | 2.91 | 3.27 | 4.00 | 1.46 |
| | Mathematics confidence | 4,452 | 1.00 | 2.58 | 3.20 | 3.26 | 3.78 | 4.00 | 1.92 |
| | Reading confidence | 4,452 | 1.00 | 2.78 | 3.29 | 3.29 | 3.69 | 4.00 | 1.12 |
| At-risk promoted, 2 years after | School belonging | 150 | 1.83 | 2.65 | 2.96 | 2.97 | 3.30 | 3.94 | 0.54 |
| | Mathematics confidence | 150 | 1.00 | 2.42 | 3.06 | 3.03 | 3.62 | 4.00 | 0.84 |
| | Reading confidence | 150 | 1.25 | 2.59 | 3.14 | 3.06 | 3.48 | 4.00 | 0.62 |

**Table D.4—Continued**

| Group | Variable | Number | Minimum | First Quartile | Mean | Median | Third Quartile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|
| Retained, 2 years after | School belonging | 122 | 1.83 | 2.89 | 3.16 | 3.22 | 3.51 | 4.00 | 0.62 |
| | Mathematics confidence | 122 | 1.25 | 2.35 | 2.99 | 2.90 | 3.45 | 4.00 | 0.77 |
| | Reading confidence | 122 | 1.00 | 2.52 | 3.14 | 3.05 | 3.66 | 4.00 | 0.65 |

**Table D.5**
**Standardized Regression Coefficients Predicting Retention Status at the End of the Academic Year**

| Variable | Coefficient |
|---|---|
| Intercept | –4.93 (0.78)** |
| School belonging | –0.04 (0.16) |
| Mathematics confidence | –0.46 (0.15)** |
| Reading confidence | –0.20 (0.18) |
| Free-lunch eligibility | 0.06 (0.75) |
| Reduced-price-lunch eligibility | –0.60 (0.95) |
| English language learner | 1.82 (0.37)** |
| Male | –0.13 (0.30) |
| Days absent | 0.40 (0.11)** |
| Previously retained | 0.92 (0.46) |
| Black | 0.63 (0.35) |
| White | –12.79 (0.66)** |
| Other race | 0.37 (0.77) |

NOTE: Standard errors are in parentheses.
* = statistically significant at the 0.04 level.
** = statistically significant at the 0.01 level.

APPENDIX E

# Data and Analyses for 3rd-Grade Cohorts

This appendix provides an overview of analyses conducted on the 3rd-grade cohorts as a supplement to our 5th-grade promotion policy study. The 3rd-grade cohorts consist of 3rd graders in the 2002–2003, 2003–2004, and 2004–2005 cohorts; the first is the comparison cohort not held to the promotion policy. The 3rd-grade promotion policy was implemented in the spring of 2004, so the two later cohorts were subject to the policy. In the text, we often refer to these cohorts as the comparison, first policy, and second policy cohorts, respectively. It is important to remember that this analysis concerns an earlier period than our analyses of the 5th-grade cohorts, so we retain the cohort year in tables and figures.

The appendix is organized into three sections. The first provides an overview of the cohorts and their proximal- and higher-grade outcomes (similar to the data presented in Chapter Seven). As in the body of this monograph, our focus here is on low-performing students. The second section presents results from the RDD and doubly robust models on the effects of the policy on proximal and future outcomes (similar to those presented in Chapters Eight and Nine). The third section examines the socio-emotional outcomes of 3rd graders based on data from the student surveys (similar to results presented in Chapter Ten). Recall that data on implementation was collected only for the 5th-grade promotion policy, so we do not have parallel data for the 3rd-grade policy.

## Performance of Students in the 3rd-Grade Cohorts on Proximal- and Higher-Grade Assessments

This section details the 3rd-grade cohorts in terms of total numbers, those who were held to policy, those who were at risk of retention after the spring assessments, and achievement trends over time.

### Students Who Were Subject to the Promotion Policy

Table E.1 shows the total number of students in the 3rd-grade cohorts and the percentage held to policy in each cohort (estimated in the case of the comparison cohort).

**Table E.1**
**Percentage of Students Classified as English Language Learner or Special Education and Held to Policy, by 3rd-Grade Cohort**

| ELL or Special Education Status | 2002–2003 | 2003–2004 | 2004–2005 |
|---|---|---|---|
| Number of students | 85,749 | 82,579 | 79,797 |
| Number of students held to policy | 71,325 | 63,054 | 61,413 |
| As a percentage of total number of students | 83.2 | 76.4 | 77.0 |

NOTE: The comparison cohort (2002–2003) was not held to policy. For this cohort, the number of students held to policy was estimated using the policy guidelines.

Special education students accounted for about 13–14 percent of the cohort, while students classified as ELL for three or fewer years accounted for another 5–7 percent. As with the 5th-grade cohorts, a number of students were exempt from the policy based on criteria other than those outlined in the regulations.

We estimated that 83 percent of the comparison cohort would have been held to the policy if it had been in place that year. This is somewhat higher than the 76–77 percent of students in the 3rd-grade cohorts who were actually subject to the promotion policy.

### Performance on the Spring Assessments

**Overall.** In both policy cohorts, more than 50 percent of students held to the policy achieved a Level 3 or Level 4 score on the ELA exam, indicating proficiency under NCLB requirements (Table E.2). The percentage of students scoring Level 1 in ELA decreased over time, from a high of 16 percent of the comparison cohort to a low of 10 percent in 2004–2005. Performance at Level 3 and Level 4 simultaneously increased, from a low of 48-percent proficient in the comparison year to a high of 60-percent proficient in the second policy cohort.

In mathematics, the percentage of students scoring Level 3 or Level 4 increased substantially over time, from 58 percent in 2002–2003 to 73 percent in 2004–2005 (Table E.2). The percentage of students scoring Level 1 in mathematics stayed relatively steady over the same period, but the percentage scoring at Level 2 decreased significantly, contributing to the higher proficiency rate of the policy cohorts.

**Students at Risk of Retention.** Students who scored Level 1 on the spring ELA or mathematics assessments were considered at risk of retention in the 3rd grade. Table E.3 shows the percentages of at-risk students in the three 3rd-grade cohorts. As we saw with the 5th-grade cohorts, the percentage of students held to the promotion policy who were at risk of retention after the spring assessments declined over time, from 19 percent in the comparison cohort to 14 percent in the second policy cohort. Unlike the 5th-grade cohorts, there was no immediate decline in the percentage of students at risk of retention when the policy was first implemented.

**Table E.2**
**Performance on the Proximal 3rd-Grade Spring Subject Assessments, by 3rd-Grade Cohort**

| | 2002–2003 | | 2003–2004 | | 2004–2005 | |
|---|---|---|---|---|---|---|
| **Performance** | **Number** | **Percentage** | **Number** | **Percentage** | **Number** | **Percentage** |
| ELA | | | | | | |
| Level 1 | 10,385 | 15.5 | 8,024 | 13.4 | 5,566 | 10.0 |
| Level 2 | 24,288 | 36.3 | 21,363 | 35.5 | 16,766 | 30.1 |
| Level 3–4 | 32,304 | 48.2 | 30,736 | 51.1 | 33,378 | 59.9 |
| Mathematics | | | | | | |
| Level 1 | 6,941 | 10.1 | 6,597 | 10.5 | 5,107 | 8.6 |
| Level 2 | 22,003 | 32.0 | 15,473 | 24.6 | 11,068 | 18.7 |
| Level 3–4 | 39,739 | 57.9 | 40,906 | 65.0 | 43,108 | 72.7 |

NOTE: The comparison cohort (2002–2003) was not held to the policy. For this cohort, the number of students held to the policy was estimated using policy guidelines.

**Table E.3**
**Students at Risk of Retention, by Subject and 3rd-Grade Cohorts**

| Subject | 2002–2003 | 2003–2004 | 2004–2005 |
|---|---|---|---|
| Total (any subject) | 18.5 | 17.5 | 13.6 |
| ELA only | 8.8 | 7.0 | 5.3 |
| Mathematics only | 4.0 | 4.7 | 4.5 |
| Both ELA and mathematics | 5.8 | 5.7 | 3.8 |
| Total held to policy | 71,325 | 63,054 | 61,413 |

NOTE: The comparison cohort (2002–2003) was not held to the policy. For this cohort, the number of students held to policy was estimated using the policy guidelines.

Unlike the 5th-grade cohorts, 3rd graders were initially at greater risk of retention because of their ELA scores. For example, the percentage at risk of being retained because of ELA was 15 percent and 13 percent for the first two cohorts, respectively, compared to 10 percent for mathematics. However, by the 2004–2005 cohort, the percentage of students at risk of retention was similar for the two subjects (9 percent for ELA and 8 percent for mathematics).

**Difference by Race/Ethnicity and ELL Status.** The 3rd-grade cohorts resemble the 5th-grade cohorts in terms of differences in the risk of retention by race/ethnicity (Figure E.1). For example, in the 2004–2005 cohort, 15 percent of Hispanic 3rd graders and 21 percent of black 3rd graders were at risk of being retained, compared to 3 percent of Asian and white students.

Figure E.2 shows differences in the risk of retention for ELL and non-ELL students. As with the 5th graders, the percentage of ELL students at risk of being retained fell markedly, from 35 percent in the 2003–2004 cohort to 23 percent in the 2004–2005 cohort; there was a smaller decline for non-ELL students (from 16 percent to 13 percent).

## Promotion and Retention Outcomes of At-Risk Students

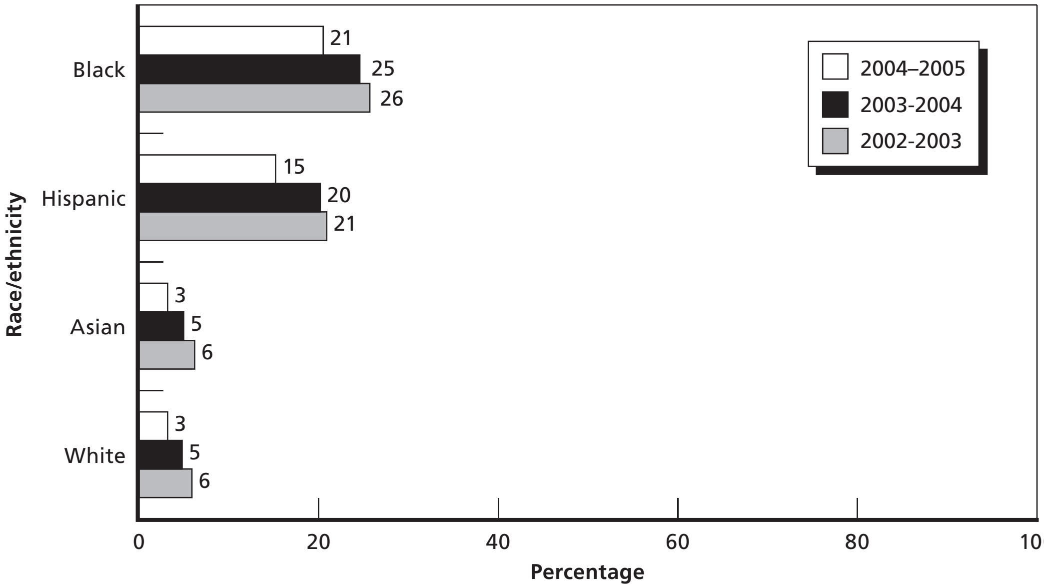
Overall, in the 2004–2005 cohort (the second policy cohort), 8,337 students were at risk of retention in either ELA or mathematics—significantly fewer than the 11,007 students at risk of retention in the first policy cohort. However, we were able to track outcomes only for 7,882 students in the second policy cohort and for 10,427 students in the first policy cohort.

Figure E.3 shows data on who was promoted and retained disaggregated by subject in which students were at risk of retention for the 2004–2005 cohort. Overall trends in promotion and retention for the two cohorts are shown in Figure E.4. In 2004–2005, 37 percent of students at risk of retention in the spring were retained in grade. This was higher than the percentage of students at risk of retention and retained in 2003–2004 (29 percent). However, in both policy cohorts, the number of students retained in grade was about the same (3,000), and these students accounted for a little under 5 percent of each cohort. As with the 5th-grade cohorts, the percentage retained varied considerably by subject and whether students were at risk in one or both subjects. Among those at risk in ELA only, 20 percent were retained in grade; this increased to 33 percent among those at risk in mathematics only and 64 percent among those at risk in both subjects. In the 2003–2004 cohort, the percentage of retained students who had been at risk in both subjects was somewhat lower, at 55 percent.

About 8–9 percent of at-risk students in the two policy cohorts were promoted on the basis of the spring portfolio, and nearly all of these students were at risk in one subject only.

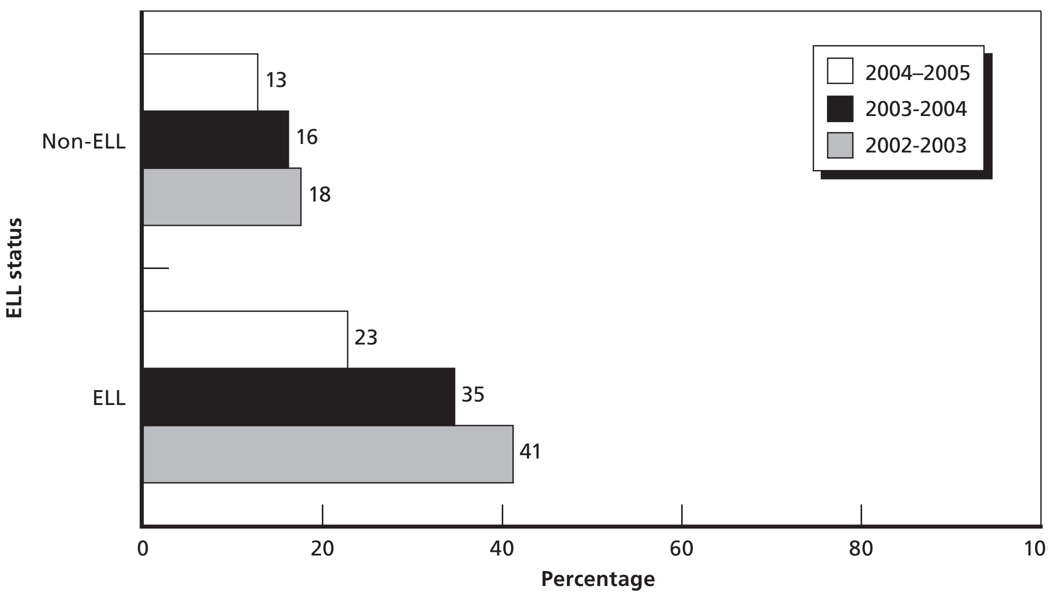
In 2004–2005, 42 percent of at-risk students were promoted by meeting the promotion standards on the summer assessments. Among students at risk in ELA only, 58 percent were promoted on this basis, and among those at risk in mathematics only, 40 percent were promoted by meeting summer assessment standards. The overall summer assessment promotion rate was similar for the 2003–2004 cohort (40 percent), with students at risk in mathematics only achieving a higher passing rate on the summer assessment than the 2004–2005 cohort and students at risk in ELA only achieving a lower passing rate than students in the 2004–2005 cohort. Across the two cohorts, between 19 and 21 percent of those at risk in both ELA and mathematics were able to meet the standards on the summer assessments.

**Figure E.1**
**Percentage of Students Who Scored Level 1 on the Proximal-Year 3rd-Grade Spring Assessments in ELA and Mathematics, by Race/Ethnicity**



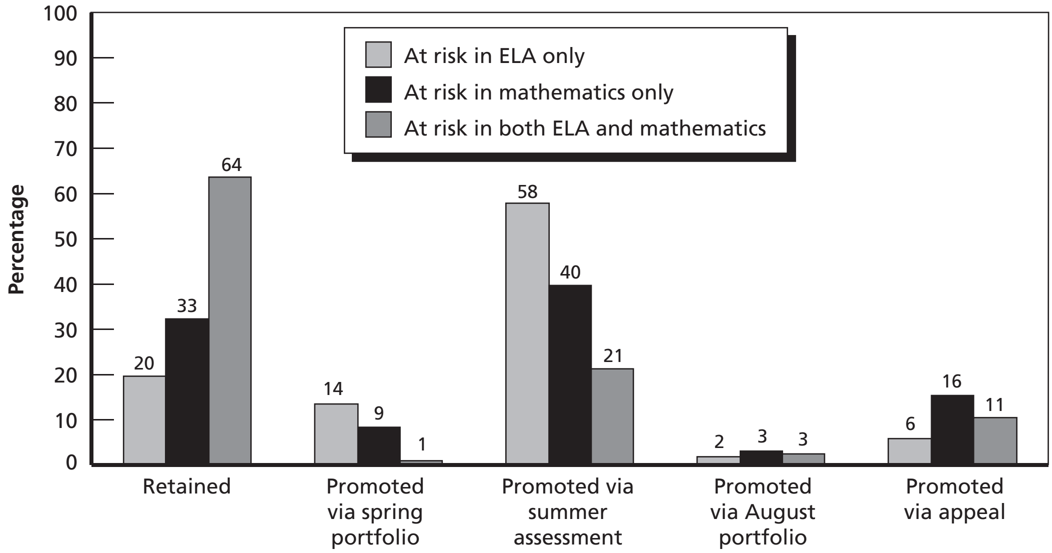

RAND *MG894-E.1*

**Figure E.2**
**Percentage of Students Who Scored Level 1 on the Proximal-Year 3rd-Grade Spring Assessments in ELA and Mathematics, by ELL Status**



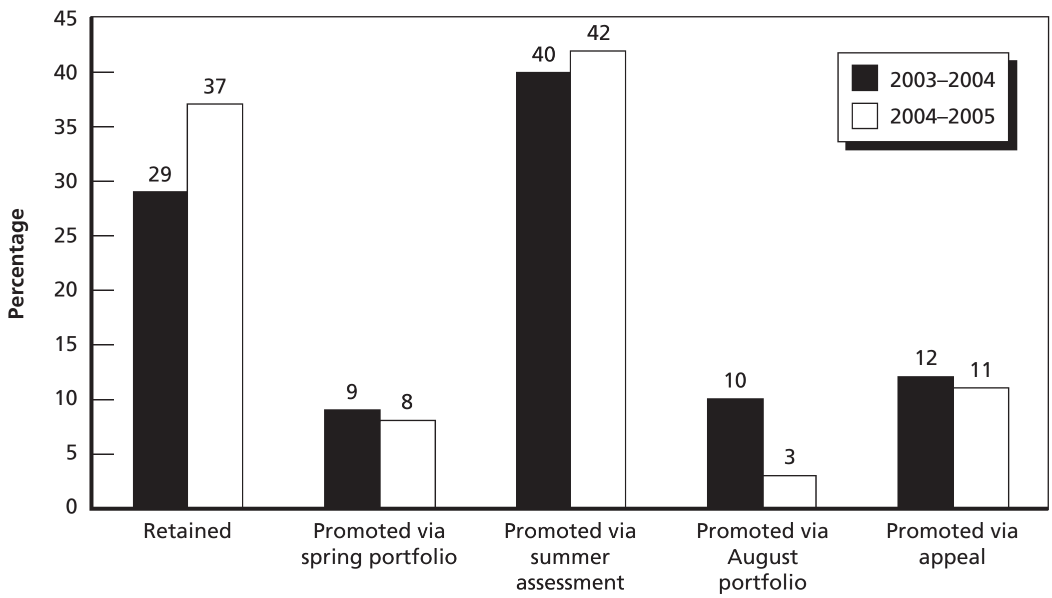

RAND *MG894-E.2*

**Figure E.3**
**Final Promotion and Retention Outcomes of At-Risk Students, by Subject, 2004–2005 3rd-Grade Cohort**




RAND *MG894-E.3*

**Figure E.4**
**Trends in Final Promotion and Retention Outcomes of At-Risk Students, by 3rd-Grade Policy Cohort**




RAND *MG894-E.4*

In 2004–2005, only 3 percent of all at-risk students were promoted based on the August portfolio. This was a decline from the 2003–2004 cohort, when 10 percent of at-risk students achieved promotion via the August portfolio. In 2004–2005, 11 percent of at-risk students were promoted on the basis of appeals, versus 12 percent the year before.

### Performance of At-Risk Promoted and Retained Students in Later Years

We now turn to the performance of at-risk students on assessments given in subsequent years. As mentioned earlier, comparing performance over time is difficult because some earlier assessments were administered by the city and later assessments were administered by the state. For the comparison cohort, the city administered the 3rd-grade assessments in spring 2003 and the 5th-grade assessments in spring 2005. For students retained for one year, the state administered the 5th-grade assessment in spring 2006. From then on, the state administered the annual assessments in all grades (3–8). For the first policy cohort, we note that the city administered the 3rd-grade assessment in spring 2004 and, for promoted students, the 4th-grade assessment in spring 2005. From then on, the assessments were administered by the state. For the second policy cohort, the city administered the 3rd-grade assessment in spring 2005, after which the state took over all the remaining assessments for both promoted and retained students.

**Retained Students' Performance on the Repeated 3rd-Grade Assessments.** Table E.4 presents the outcomes of students retained in the 3rd grade in the 2002–2003 comparison cohort and the 2003–2004 and 2004–2005 policy cohorts in terms of their performance on the 3rd-grade assessments that they repeated the following year.

In the comparison cohort, about 35 percent of retained students who had been at risk in ELA failed to meet the promotion standards in that subject the following year, and 31 percent of those who had been at risk in mathematics failed to do so in that subject. Others improved considerably over the repeated year, however. For example, about 11 percent of those at risk in ELA were able to achieve Level 3 or higher on the repeated assessment, and this was true of 23 percent of students who had been at risk in mathematics.

The first policy cohort resembled the comparison cohort, with 31 percent of students at risk in ELA and 27 percent at risk in mathematics failing to meet the promotion standards the following year. However, the percentage of students who improved their performance to the proficiency level was higher in this cohort, with 14 percent of those at risk in ELA and 32 percent of those at risk in mathematics scoring Level 3 or higher on the repeated 3rd-grade test.

For the second policy cohort, the first test was given by the city and the second by the state. There was a marked improvement in the performance of retained students. Only 21 percent of retained students failed to meet promotion standards in ELA, and this was true of only 12 percent of retained students in mathematics. Remarkably,

**Table E.4**
**Retained Students' Performance on the Repeated 3rd-Grade Assessment, by Subject and 3rd-Grade Cohort**

| Cohort and Subject | Performance Level on Repeated 3rd-Grade Assessments (% who scored Level 1 in subject in previous year) Level 1 | Level 2 | Level 3–4 | Number of Students |
|---|---|---|---|---|
| 2002–2003 (assessed in spring 2004) | | | | |
| ELA | 35.4 | 54.0 | 10.7 | 2,505 |
| Mathematics | 30.6 | 46.1 | 23.3 | 1,953 |
| 2003–2004 (assessed in spring 2005) | | | | |
| ELA | 30.9 | 54.8 | 14.3 | 2,347 |
| Mathematics | 26.8 | 41.5 | 31.7 | 2,309 |
| 2004–2005 (assessed in spring 2006) | | | | |
| ELA | 20.9 | 54.0 | 25.2 | 1,938 |
| Mathematics | 12.3 | 33.6 | 54.1 | 2,207 |

NOTE: The 2004 (2002–2003 cohort) and 2005 (2003–2004 cohort) assessments were administered by the city. The 2006 (2004–2005 cohort) assessment represents a shift in administration from the city to the state. For this cohort, the original 3rd-grade assessment was administered by the city, while the repeated 3rd-grade assessment was administered by the state.

one-quarter of students achieved proficiency level or higher in ELA, and more than half (54 percent) did so in mathematics. Again, we cannot determine whether the differences are caused by actual differences in the performance of this cohort or in the benchmarking procedure between the city and state.

**Retained Students' Performance on the 4th- Through 7th-Grade Assessments.** Table E.5 tracks the performance of the retained students in later grades. While all 4th-grade assessments were given by the state, a new state testing program was introduced in the spring of 2006, when retained students from the first policy cohort were in 4th grade. While not directly comparable, it is interesting to note that the retained students from the two policy cohorts who took the new test did worse than students from the comparison cohort, with higher percentages of students scoring Level 1. Interestingly, the ELA proficiency rates were similar across the three cohorts, but mathematics proficiency rates fell and then rose again over the two policy cohorts compared to the comparison cohort.

Students from all three cohorts were subject to the 5th-grade promotion policy. The comparison cohort took the state-administered 5th-grade test under the new program, so we cannot compare performance between the 4th and 5th grades. About

**Table E.5**
**Performance of Retained Students on the 4th- Through 7th-Grade Assessments, by 3rd-Grade Cohort**

| Cohort and Subject | Performance on the 4th-Grade Assessments: Level 1 | Level 2 | Level 3–4 | Number of Students | Performance on the 5th-Grade Assessments: Level 1 | Level 2 | Level 3–4 | Number of Students |
|---|---|---|---|---|---|---|---|---|
| 2002–2003 | | | | | | | | |
| ELA | 14.0 | 65.5 | 20.5 | 2,169 | 17.3 | 62.9 | 19.7 | 2,048 |
| Mathematics | 9.5 | 42.0 | 48.5 | 1,638 | 30.5 | 46.3 | 23.2 | 1,530 |
| 2003–2004 | | | | | | | | |
| ELA | 25.2 | 56.4 | 18.4 | 2,085 | 25.0 | 56.2 | 18.8 | 1,943 |
| Mathematics | 19.8 | 43.4 | 36.9 | 1,954 | 18.9 | 43.3 | 37.9 | 1,825 |
| 2004–2005 | | | | | | | | |
| ELA | 18.7 | 61.9 | 19.5 | 1,727 | 2.2 | 61.0 | 36.8 | 1,635 |
| Mathematics | 13.8 | 45.1 | 41.0 | 1,933 | 8.6 | 41.9 | 49.5 | 1,840 |

| Cohort and Subject | Performance on the 6th-Grade Assessments: Level 1 | Level 2 | Level 3–4 | Number of Students | Performance on the 7th-Grade Assessments: Level 1 | Level 2 | Level 3–4 | Number of Students |
|---|---|---|---|---|---|---|---|---|
| 2002–2003 | | | | | | | | |
| ELA | 4.4 | 80.5 | 15.1 | 1,803 | 2.6 | 72.2 | 25.2 | 1,683 |
| Mathematics | 30.5 | 45.4 | 24.1 | 1,323 | 11.9 | 52.8 | 35.3 | 1,222 |
| 2003–2004 | | | | | | | | |
| ELA | 23.6 | 57.0 | 19.4 | 1,804 | NA | NA | NA | NA |
| Mathematics | 17.3 | 43.7 | 39.0 | 1,690 | NA | NA | NA | NA |
| 2004–2005 | | | | | | | | |
| ELA | NA | NA | NA | NA | NA | NA | NA | NA |
| Mathematics | NA | NA | NA | NA | NA | NA | NA | NA |

NOTE: Shading indicates that the assessment was administered by the state under the old testing program.

17 percent failed the ELA assessment, and 31 percent failed the mathematics assessment. The percentages not meeting the promotion standards remained the same for the first policy cohort between 4th and 5th grades (both given by the state under the new testing program), with 25 percent not meeting the standards in ELA and 19 percent in mathematics. The failure rate fell dramatically for the second policy cohort, reflecting the general trend in the city and state in 2008. Proficiency rates were remarkably high

for this cohort—37 percent of these retained students achieved proficiency in ELA and about half did so in mathematics.

Performance in the 6th grade was similar to that in the 5th grade for the comparison and first policy cohorts. The performance on the 5th- and 6th-grade assessments was similar to that of the retained 5th graders in these two grades.

**At-Risk Promoted Students' Performance.** The comparison cohort and the first policy cohort took the city assessment in the 4th grade, and the second policy cohort took the state assessment in the spring of 2006. By the time these students were in the 5th grade, they became part of the 5th-grade cohorts that are the focus of our study and whose outcomes are discussed in detail in Chapter Seven. As a result, we do not analyze their outcomes in later grades separately here.

## Effect of Supportive Interventions and Retention on Proximal-Year and Future Student Achievement, 3rd-Grade Cohorts

Chapters Eight and Nine presented results of the models measuring the effect of the supportive interventions and being retained in grade on proximal and future student outcomes for the 5th-grader cohorts. This section presents analogous results for the 3rd-grade cohorts.

### Effects on Proximal-Year Outcomes

Table E.6 shows the research questions we examine with respect to proximal-year outcomes, mapped to our methods.

The analytical methods used to analyze the effects of supportive interventions on proximal- and later-grade outcomes are discussed in Chapters Four, Eight, and Nine, as well as in Appendix A. Thus, in this section, we simply present results from our models.

Table E.7 shows the available comparison groups for the 3rd-grade models. The 2002–2003, 2003–2004, and 2004–2005 cohorts all took city-administered assessments in the 3rd grade. Thus, we can use the 2002–2003 cohort to define comparison

**Table E.6**
**Proximal-Year Outcomes: Research Questions and Methods**

| Research Question | Method |
|---|---|
| Students at risk of retention based on performance on proximal-year spring assessments, relative to a comparison group | |
| How did these students perform on the summer assessments? | Propensity score weighting and doubly robust regression |
| What was the relationship between SSA attendance and performance on the summer assessments? | GAMM |

**Table E.7**
**Timing of Assessments, by Promotion Status and City- Versus State-Administered Tests, 3rd-Grade Cohort**

| Cohort | Spring and Summer 3rd-Grade Assessment | Promotion Status | 3rd-Grade Repeated Assessment | 4th-Grade Assessment | 5th-Grade Assessment |
|---|---|---|---|---|---|
| 2002–2003 | 2002–2003 | Promoted | — | 2003–2004 | 2004–2005 |
| | | Retained | 2003–2004 | 2004–2005 | 2005–2006 |
| 2003–2004 | 2003–2004 | Promoted | — | 2004–2005 | 2005–2006 |
| | | Retained | 2004–2005 | 2005–2006 | 2006–2007 |
| 2004–2005 | 2004–2005 | Promoted | — | 2005–2006 | 2006–2007 |
| | | Retained | 2005–2006 | 2006–2007 | 2007–2008 |

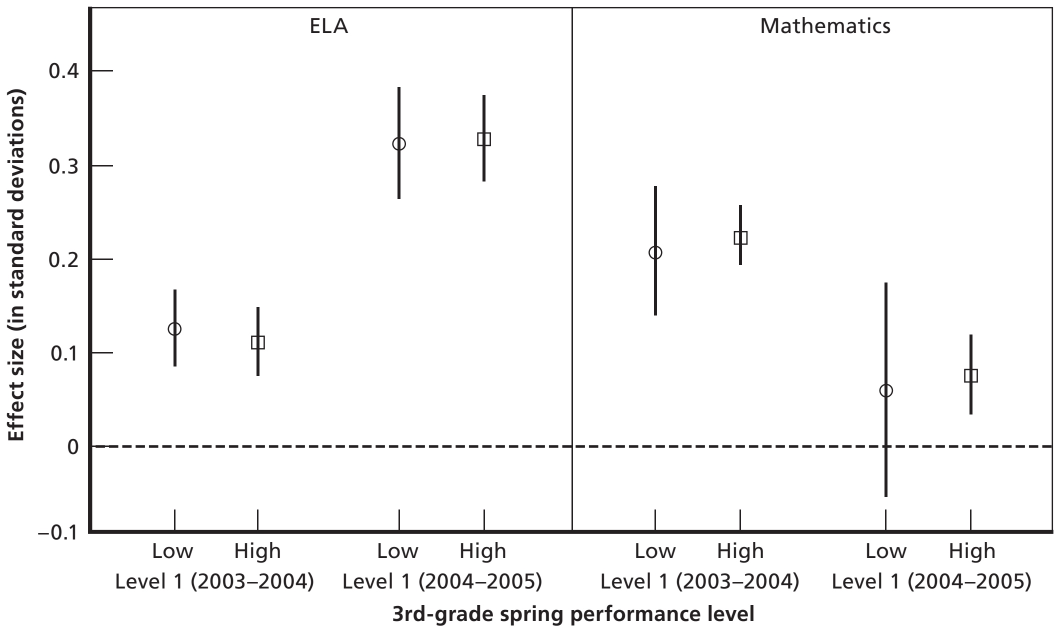
NOTE: Shading indicates tests administered by the city (3rd and 5th grades) or under the previous state assessment program (4th grade).

groups of students who were (1) at risk of retention because of failure to score Level 2 or higher on the spring assessments and (2) retained in grade.

For the study period, we cannot compare any future outcomes for 3rd graders in the 2002–2003 comparison cohort for two reasons. First, until the 5th grade, there was no single grade in which the comparison and policy cohorts took assessments administered by the same entity. Second, the 6th-grade assessment was the first administered entirely by the state for all subgroups in these cohorts and, thus, available for comparison. However, in the 5th grade, the 3rd-grade cohorts were again subject to the policy, and some students were retained; as a result, some students in the 2004–2005 cohort who were retained in grade twice would not have taken the 6th-grade assessment by the spring of 2008, the last year for which we have data.

Figure E.5 shows the estimated treatment effect sizes on performance on the summer assessments for 3rd graders at risk of retention in the 2003–2004 and 2004–2005 cohorts, respectively. The ELA results are all positive. On average, 3rd graders in the policy cohorts who were at risk of retention after the spring assessments did better than they would have had they not been subject to the promotion policy, regardless of where they scored in the Level 1 range. The 2003–2004 cohort shows much smaller effect sizes than the 2004–2005 cohort (0.11–0.13 versus 0.32–0.33). These effect sizes are all small in magnitude. In mathematics, we see a reverse trend, with larger effect sizes in the earlier cohort. For example, effect sizes in the 2003–2004 cohort were 0.21–0.22 versus 0.07–0.08 for the 2004–2005 cohort. The effect size for low Level 1 students in the 2004–2005 cohort was not statistically significant.

**Figure E.5**
**Estimated Treatment Effects on Performance on the Summer Assessment, 3rd-Grade Cohorts**



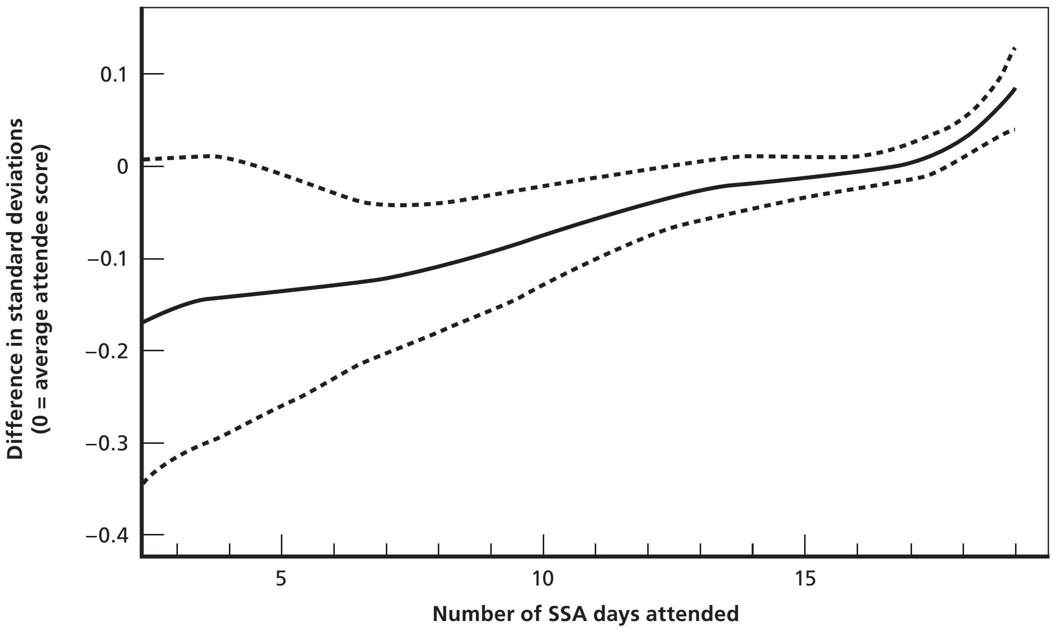

RAND *MG894-E.5*

These outcomes are substantially higher than those for the 5th-grade cohorts. Summer school existed prior to the promotion policy; these results show the marginal effect of the SSA under the promotion policy relative to that of the prior program.

### Relationship Between Frequency of SSA Attendance and Performance on the Summer Assessments

We used GAMM to examine the relationship between the frequency of SSA attendance and performance on the summer assessments. The relationship for ELA was not significant and is not shown here. Figure E.6 presents the results for mathematics. Similar to the 5th-grade graphs presented in Chapter Eight, this figure is centered at the average summer scale score of the students attending the SSA, and the *y*-axis is in the number of standard deviations of the scale score. As with the 2006 mathematics results, there is an increasing relationship between SSA attendance and performance in the midrange of attendance; in this case, the expected summer mathematics outcome is expected to increase by about 0.10 standard deviations between six and 13 days of attendance. After 13 days, the relationship flattens; however, unlike the 5th-grade results, there is an additional range of increasing relationship, between 17 and 19 days of SSA attendance, where the expected mathematics summer outcome increases by roughly an additional 0.08 standard deviations.

**Figure E.6**
**Relationship Between Number of SSA Sessions Attended and 3rd-Grade Summer Mathematics Assessment Scale Score, 2003–2004 3rd-Grade Cohort**




RAND *MG894-E.6*

### Future Outcomes of Students at Risk of Retention

We now examine how students at risk of retention or who were retained in grade performed in subsequent grades, comparing them to similar groups of students falling just above the treatment thresholds. We use RDD models to address this question.

Table E.8 presents the results from the RDD models for the two 3rd-grade cohorts on the 4th- and 5th-grade assessments. For the 4th grade, because of the switch between city and state administrations of the test, we report outcomes only for the 2004–2005 cohort. Again, the two treatments are (1) being mandated to SSA based on 3rd-grade spring assessment performance and (2) retained on the basis of the summer assessment. As with the 5th-grade results, the RDD estimates produced using the 3rd-grade spring assessment as the treatment assignment variable are an indication of the effect of being mandated to SSA for all students just below the spring Level 2 threshold plus the effect of retention for those students just below the spring threshold who ultimately were retained; in the top half of Table E.8, we show estimated net SSA effects that have been adjusted to remove the retained component (see Chapter Nine for a more detailed analogous discussion for the 5th-grade cohorts). Being mandated to SSA does not produce a significant effect on future 4th- or 5th-grade ELA outcomes for 3rd-grade Level 1 students just below the Level 2 threshold. In mathematics, we found a small

**Table E.8**
**Effect of Being Mandated to SSA and Being Retained in Grade on Future-Grade Spring Assessments, 3rd-Grade Cohorts**

| Grade, Subject, and Cohort Year | Estimate | 95% Confidence Interval | Number of Observations Below Cutoff | Number of Observations Above Cutoff |
|---|---|---|---|---|
| Effect of being mandated to SSA | | | | |
| 4th-grade ELA | | | | |
| 2004–2005 | 0.00 | (–0.05, 0.04) | 2,851 | 22,900 |
| 4th-grade mathematics | | | | |
| 2004–2005 | 0.11* | (0.04, 0.19) | 2,320 | 12,549 |
| 5th-grade ELA | | | | |
| 2003–2004 | 0.02 | (–0.05, 0.09) | 3,296 | 39,030 |
| 2004–2005 | –0.04 | (–0.08, 0.01) | 2,642 | 21,561 |
| 5th-grade mathematics | | | | |
| 2003–2004 | 0.02 | (–0.04, 0.09) | 1,729 | 19,353 |
| 2004–2005 | 0.08* | (0.01, 0.15) | 2,121 | 11,706 |
| Effect of being retained in 3rd grade | | | | |
| 4th-grade ELA | | | | |
| 2004–2005 | 0.72* | (0.64, 0.80) | 678 | 2,211 |
| 4th-grade mathematics | | | | |
| 2004–2005 | 0.67* | (0.58, 0.77) | 1,293 | 1,527 |
| 5th-grade ELA | | | | |
| 2003–2004 | 0.81* | (0.59, 1.03) | 1,056 | 1,885 |
| 2004–2005 | 0.58* | (0.46, 0.70) | 668 | 1,977 |
| 5th-grade mathematics | | | | |
| 2003–2004 | 0.76* | (0.64, 0.88) | 589 | 1,454 |
| 2004–2005 | 0.48* | (0.38, 0.58) | 1,215 | 1,322 |

NOTE: The estimates are reported as effect sizes. We standardized using the theoretic standard deviation for the state-administered tests (40 for both subjects).

* = statistically significant at the 0.05 level.

effect of SSA of 0.08–0.11 standard deviations on future 4th- and 5th-grade outcomes for the 2004–2005 cohort, but no significant effect for the prior cohort.

We found large and significant effects of being retained into the subsequent 3rd-grade cohort (about 0.70 standard deviations in both subjects in the 4th grade). In the 5th grade, the effects are larger for the earlier cohort, but even in the second cohort, we show that students who were retained in grade outperformed their peers who were just above the Level 2 cutoff on the 3rd-grade summer assessments by 0.58 standard deviations in ELA and 0.48 standard deviations in mathematics.

Note that, as with the 5th-grade RDD estimates, the estimate for retention into the subsequent 3rd-grade cohort includes both a retention component and a component for programmatic differences between the original and subsequent cohort. In Chapter Nine, we adjusted these RDD estimates to account for an estimate of the programmatic differences using nontreated students and a baseline score from the spring assessment in the prior grade. However, in this case, there are no standardized grade 2 test scores to use as a baseline. The retention estimates in Table E.8 are unadjusted for programmatic differences between cohorts. In other words, the estimates in the table are an accurate reflection of how much better students just below the Level 2 threshold in the summer performed in later grades relative to their peers just above the threshold, but these differences cannot necessarily be attributed to the effect of retention alone.

## Analyses of Socioemotional Outcomes of 3rd Graders

We presented results from our student surveys of 5th graders in Chapter Ten. Here, we present our results on the socioemotional outcomes of students held to the 3rd-grade promotion policy. Table E.9 shows the timing of the survey administrations relative to the retention decision for these two cohorts, and Table E.10 shows the sample size of respondents by promotion decision subgroup for these two 3rd-grade cohorts.

### Distributional Differences in Socioemotional Responses Among Retained, At-Risk Promoted, and Not-at-Risk Students

The 3rd-grade cohorts allowed us to extend the analysis to examine socioemotional responses over a longer period—namely, four and five years removed from the retention decision. Table E.11 presents the descriptive statistics for the 3rd-grade cohorts. Figure E.7 presents the school belonging and reading and mathematics confidence for students who were two to five years removed from their 3rd-grade retention decision.[1] Like the 5th-grade cohort, the 3rd-grade retained students expressed a significantly higher sense of school belonging than both the not-at-risk and at-risk promoted

[1] As with the 5th-grade cohorts, there were few differences among the cohorts between years, and all the effect sizes were within the small range (*d* ranged from 0.09 to 0.25).

**Table E.9**
**Student Surveys, by Number of Years Since the Retention Decision, by 3rd-Grade Cohort**

| Cohort | 2006–2007 | 2007–2008 | 2008–2009 |
|---|---|---|---|
| 2004–2005 | 2 | 3 | 4 |
| 2003–2004 | 3 | 4 | 5 |

**Table E.10**
**Sample Sizes for Each Subgroup, by 3rd-Grade Cohort and Survey Year**

| Cohort | 2006–2007 | 2007–2008 | 2008–2009 |
|---|---|---|---|
| 2004–2005 | | | |
| Not at risk | 2,412 | 3,278 | 3,722 |
| At-risk promoted | 389 | 385 | 420 |
| Retained | 304 | 309 | 423 |
| 2003–2004 | | | |
| Not at risk | 2,918 | 1,554 | 2,384 |
| At-risk promoted | 622 | 402 | 457 |
| Retained | 285 | 312 | 283 |

students two, three, and four years after the retention decision. However, the mean differences were small, with effect sizes no larger than 0.22, and by the fifth year, there were no differences among any of the groups.

With respect to mathematics attitudes, the not-at-risk students were significantly more confident than the at-risk promoted group two, three, and four years after the retention decision. Effect sizes ranged from 0.08 to 0.19, indicating that the mean differences were all within the small range. By the fifth year, there were no differences between the not-at-risk and at-risk promoted students. This pattern of diminishing advantage can be contrasted with that between the not-at-risk and retained students, in which retained students were as confident as the not-at-risk group in the second and third years but expressed significantly lower mathematics confidence in the fourth and fifth years ($d = 0.05$ and 0.10, respectively). This finding was due to the declining confidence of retained students, who experienced a significant decrease in mathematics attitudes over time, especially between the third and fifth years. In contrast, the mathematics confidence scores of the not-at-risk and at-risk promoted students remained stable across the years.

**Table E.11**
**Descriptive Statistics for the 3rd-Grade Cohorts**

| Group | Variable | Number | Minimum | First Quartile | Mean | Median | Third Quartile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|
| Not at risk, 2 years after | School belonging | 2,412 | 1.17 | 2.64 | 3.00 | 3.03 | 3.36 | 4.00 | 0.95 |
| | Mathematics confidence | 2,412 | 1.00 | 2.61 | 3.22 | 3.27 | 3.79 | 4.00 | 1.14 |
| | Reading confidence | 2,412 | 1.00 | 2.84 | 3.34 | 3.37 | 3.76 | 4.00 | 0.89 |
| At-risk promoted, 2 years after | School belonging | 389 | 1.33 | 2.60 | 2.93 | 2.98 | 3.27 | 4.00 | 0.55 |
| | Mathematics confidence | 389 | 1.00 | 2.30 | 3.02 | 3.04 | 3.76 | 4.00 | 0.91 |
| | Reading confidence | 389 | 1.00 | 2.45 | 3.06 | 3.01 | 3.53 | 4.00 | 0.86 |
| Retained, 2 years after | School belonging | 304 | 1.56 | 2.83 | 3.13 | 3.18 | 3.40 | 4.00 | 0.54 |
| | Mathematics confidence | 304 | 1.00 | 2.51 | 3.14 | 3.10 | 3.75 | 4.00 | 0.81 |
| | Reading confidence | 304 | 1.00 | 2.39 | 3.09 | 3.07 | 3.64 | 4.00 | 1.13 |
| Not at risk, 3 years after | School belonging | 6,196 | 1.11 | 2.62 | 2.98 | 3.01 | 3.35 | 4.00 | 1.29 |
| | Mathematics confidence | 6,196 | 1.00 | 2.59 | 3.22 | 3.30 | 3.79 | 4.00 | 1.24 |
| | Reading confidence | 6,196 | 1.00 | 2.86 | 3.35 | 3.35 | 3.74 | 4.00 | 1.77 |
| At-risk promoted, 3 years after | School belonging | 1,007 | 1.17 | 2.55 | 2.91 | 2.94 | 3.25 | 4.00 | 0.74 |
| | Mathematics confidence | 1,007 | 1.00 | 2.33 | 3.06 | 3.08 | 3.71 | 4.00 | 1.18 |
| | Reading confidence | 1,007 | 1.00 | 2.46 | 3.06 | 3.00 | 3.48 | 4.00 | 0.85 |

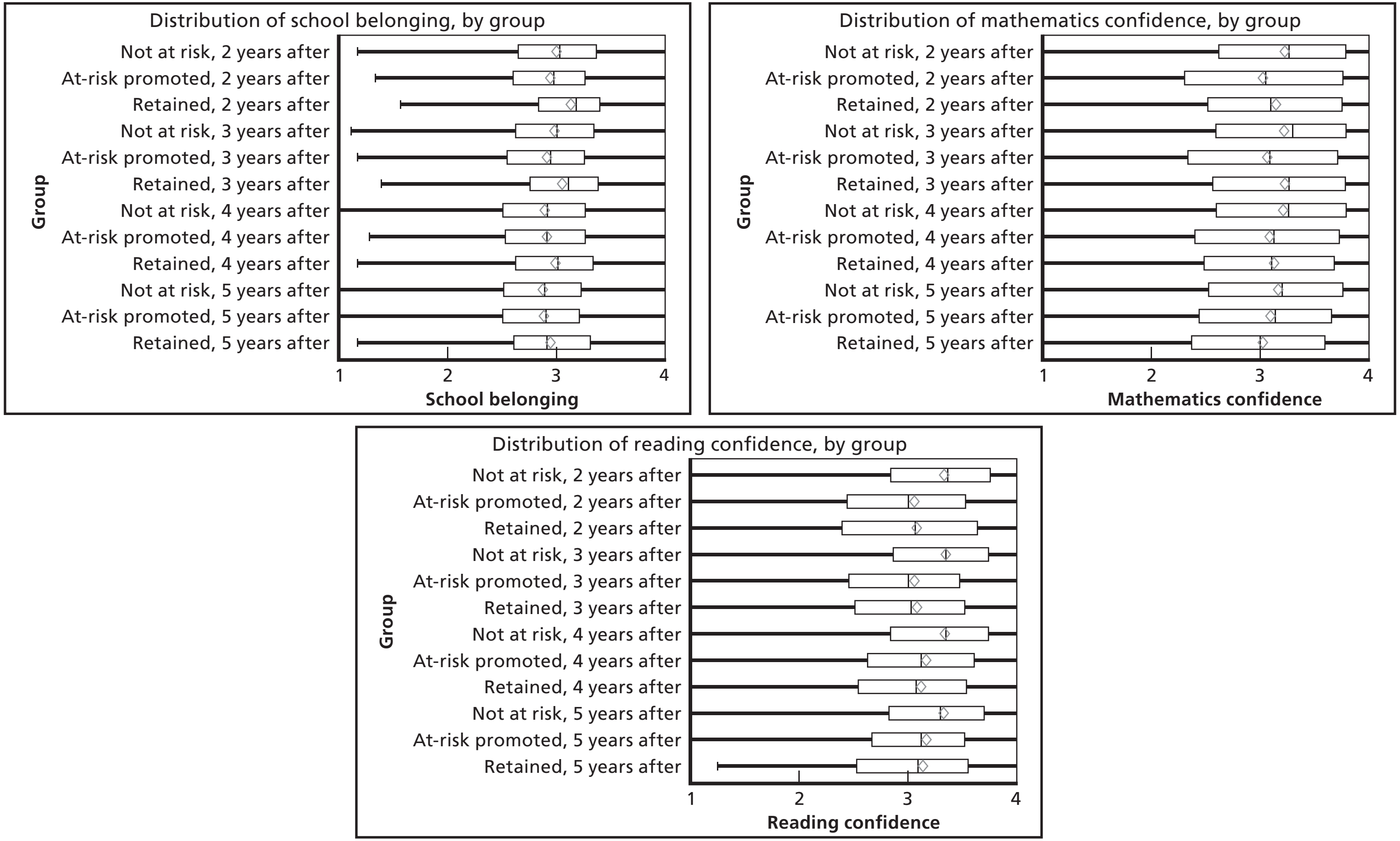
**Table E.11—Continued**

| Group | Variable | Number | Minimum | First Quartile | Mean | Median | Third Quartile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|
| Retained, 3 years after | School belonging | 594 | 1.39 | 2.75 | 3.06 | 3.10 | 3.38 | 4.00 | 0.65 |
| | Mathematics confidence | 594 | 1.00 | 2.56 | 3.22 | 3.26 | 3.78 | 4.00 | 0.93 |
| | Reading confidence | 594 | 1.00 | 2.51 | 3.09 | 3.03 | 3.52 | 4.00 | 0.92 |
| Not at risk, 4 years after | School belonging | 5,276 | 1.00 | 2.51 | 2.89 | 2.91 | 3.27 | 4.00 | 1.53 |
| | Mathematics confidence | 5,276 | 1.00 | 2.59 | 3.21 | 3.26 | 3.79 | 4.00 | 1.79 |
| | Reading confidence | 5,276 | 1.00 | 2.84 | 3.34 | 3.35 | 3.74 | 4.00 | 0.89 |
| At-risk promoted, 4 years after | School belonging | 822 | 1.28 | 2.53 | 2.91 | 2.91 | 3.26 | 4.00 | 0.73 |
| | Mathematics confidence | 822 | 1.00 | 2.39 | 3.08 | 3.13 | 3.73 | 4.00 | 1.10 |
| | Reading confidence | 822 | 1.00 | 2.63 | 3.17 | 3.12 | 3.61 | 4.00 | 0.75 |
| Retained, 4 years after | School belonging | 735 | 1.17 | 2.62 | 2.99 | 3.01 | 3.33 | 4.00 | 0.57 |
| | Mathematics confidence | 735 | 1.00 | 2.48 | 3.13 | 3.10 | 3.68 | 4.00 | 0.78 |
| | Reading confidence | 735 | 1.00 | 2.54 | 3.12 | 3.07 | 3.54 | 4.00 | 0.68 |
| Not at risk, 5 years after | School belonging | 2,384 | 1.00 | 2.51 | 2.87 | 2.89 | 3.23 | 4.00 | 1.21 |
| | Mathematics confidence | 2,384 | 1.00 | 2.52 | 3.17 | 3.20 | 3.76 | 4.00 | 1.56 |
| | Reading confidence | 2,384 | 1.00 | 2.83 | 3.32 | 3.30 | 3.70 | 4.00 | 1.15 |

**Table E.11—Continued**

| Group | Variable | Number | Minimum | First Quartile | Mean | Median | Third Quartile | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|
| At-risk promoted, 5 years after | School belonging | 457 | 1.00 | 2.51 | 2.88 | 2.90 | 3.21 | 4.00 | 0.58 |
| | Mathematics confidence | 457 | 1.00 | 2.43 | 3.10 | 3.14 | 3.66 | 4.00 | 1.26 |
| | Reading confidence | 457 | 1.00 | 2.67 | 3.17 | 3.12 | 3.52 | 4.00 | 0.85 |
| Retained, 5 years after | School belonging | 283 | 1.17 | 2.61 | 2.94 | 2.91 | 3.31 | 4.00 | 0.63 |
| | Mathematics confidence | 283 | 1.00 | 2.36 | 3.02 | 3.00 | 3.59 | 4.00 | 0.79 |
| | Reading confidence | 283 | 1.25 | 2.53 | 3.14 | 3.09 | 3.56 | 4.00 | 0.71 |

**Figure E.7**
**Third-Grade Students' Socioemotional Responses One, Two, and Three Years After the Retention Decision**




RAND *MG894-E.7*

The results for reading attitudes showed a clear advantage for the not-at-risk students over both the retained and at-risk promoted students across all years examined. The effect sizes favoring the not-at-risk students ranged from 0.14 to 0.31 and were, therefore, considered small. There were no differences in reading confidence between the at-risk promoted and retained students, both in the short and longer terms.

Overall, the results suggest a mixed picture of the relationship between retention status and students' socioemotional well-being. Retention was positively related to school belonging, as retained students expressed a greater sense of school belonging two, three, and four years after the retention decision. It is important to note that, although there were statistically significant mean differences between the retained and nonretained students, the differences translated to small effect sizes in every instance.

Retention did not have a consistent relationship with mathematics or reading attitudes. For example, mathematics confidence decreased over the years for 3rd-grade retained students but increased over the years for 5th-grade retained students. This resulted in diminishing confidence gaps between the retained and nonretained students in the 3rd-grade cohorts, but increasing gaps in the 5th-grade cohorts. Similarly, for the first two years after the retention decision, retention status was unrelated to 5th-grade students' reading attitudes, as retained students expressed similar levels of reading confidence to the not-at-risk students. In contrast, across all four years, there was a negative relationship between retention and reading confidence in the 3rd-grade cohorts, with 3rd-grade students reporting lower levels of reading confidence than their not-at-risk peers. It is difficult to explain these contradictory results, but it is possible that the analyses were influenced by the specific characteristics of the sample of participating students in both cohorts.

### Relationship Between Retention and Students' Later Socioemotional Outcomes

We analyzed how prior retention status is related to later socioemotional outcomes, after controlling for student demographics. Table E.12 provides the results of the full model for the not-at-risk and at-risk promoted subgroups for the 3rd-grade cohort two to five years after the retention decision. For ease of discussion, we pull out the standardized coefficients from the three models for the two groups in Table E.13.

These results showed virtually no changes from the descriptive analyses reported earlier. The one exception involved the mathematics confidence scale. While the unadjusted mean differences suggested that the not-at-risk students expressed a higher level of mathematics confidence than did the retained students four and five years after the retention decision, the regression results indicated no differences between the two groups for these years after controlling for student demographics.

**Table E.12**
**Standardized Regression Coefficients for the 3rd-Grade Cohort Two to Five Years After the Retention Decision**

| Variable | 2 Years | 3 Years | 4 Years | 5 Years |
|---|---|---|---|---|
| School belonging | | | | |
| Intercept | 0.57 (0.16)** | 0.33 (0.10)** | 0.39 (0.09)** | 0.25 (0.12) |
| Free-lunch eligibility | −0.15 (0.13)** | −0.06 (0.08) | −0.10 (0.07) | −0.06 (0.09) |
| Reduced-price-lunch eligibility | −0.09 (0.16) | −0.07 (0.10) | −0.15 (0.11) | −0.10 (0.11) |
| English language learner | −0.06 (0.07) | 0.08 (0.04) | 0.11 (0.06) | 0.11 (0.08) |
| Male | −0.12 (0.04)** | −0.12 (0.03)** | −0.10 (0.04)* | −0.06 (0.06) |
| Days absent | 0.06 (0.03)* | −0.09 (0.02)** | −0.05 (0.02)* | 0.00 (0.03) |
| Special education student | −0.95 (0.06)** | −0.25 (0.25) | −0.09 (0.50) | −0.97 (0.89) |
| Black | −0.15 (0.06)** | −0.10 (0.04)* | −0.13 (0.06)* | 0.03 (0.06) |
| White | 0.28 (0.20) | 0.25 (0.10)* | 0.09 (0.11) | 0.02 (0.07) |
| Other race | −0.25 (0.10)* | 0.02 (0.05) | −0.23 (0.07)** | −0.01 (0.13) |
| At-risk promoted | −0.40 (0.08)** | −0.32 (0.08)** | −0.19 (0.06)** | −0.11 (0.09) |
| Not at risk | −0.29 (0.07)** | −0.21 (0.06)** | −0.25 (0.05)** | −0.17 (0.09) |
| Mathematics confidence | | | | |
| Intercept | −0.04 (0.14) | −0.10 (0.08) | −0.12 (0.08) | −0.13 (0.10) |
| Free-lunch eligibility | −0.15 (0.12) | −0.03 (0.05) | −0.08 (0.07) | −0.07 (0.09) |
| Reduced-price-lunch eligibility | −0.17 (0.15) | −0.07 (0.06) | 0.01 (0.06) | 0.02 (0.16) |
| English language learner | −0.17 (0.06)** | −0.03 (0.03) | −0.03 (0.05) | −0.13 (0.06) |
| Male | 0.34 (0.04)** | 0.24 (0.03)** | 0.21 (0.04)** | 0.23 (0.06)** |
| Days absent | −0.06 (0.02)** | −0.10 (0.02)** | −0.11 (0.02)** | −0.07 (0.02) |
| Special education student | −0.73 (0.04)** | −0.58 (0.35) | −0.79 (0.09)** | −0.32 (0.07)** |
| Black | −0.02 (0.05) | 0.04 (0.04) | 0.07 (0.05) | 0.05 (0.06) |
| White | 0.17 (0.17) | 0.14 (0.09) | 0.08 (0.09) | 0.08 (0.08) |
| Other race | 0.22 (0.06)** | 0.22 (0.03)** | 0.17 (0.05)** | 0.22 (0.07)** |
| At-risk promoted | −0.15 (0.09) | −0.17 (0.07)* | −0.08 (0.06) | 0.06 (0.10) |
| Not at risk | 0.08 (0.06) | −0.01 (0.06) | 0.09 (0.05) | 0.07 (0.08) |

**Table E.12—Continued**

| Variable | 2 Years | 3 Years | 4 Years | 5 Years |
|---|---|---|---|---|
| Reading confidence | | | | |
| Intercept | –0.08 (0.13) | –0.22 (0.08)** | –0.12 (0.09) | –0.19 (0.12) |
| Free-lunch eligibility | –0.24 (0.08)** | –0.12 (0.05)* | –0.17 (0.07)* | –0.05 (0.08) |
| Reduced-price-lunch eligibility | –0.14 (0.10) | –0.06 (0.06) | –0.14 (0.08) | 0.11 (0.09) |
| English language learner | –0.30 (0.05)** | –0.16 (0.05)** | –0.17 (0.04)** | –0.20 (0.10) |
| Male | –0.03 (0.03) | –0.08 (0.03)** | –0.12 (0.04)** | –0.14 (0.05)** |
| Days absent | 0.04 (0.02)** | 0.00 (0.02) | 0.02 (0.01) | 0.00 (0.02) |
| Special education student | –1.60 (0.05)** | 0.18 (0.27) | –0.78 (0.79) | 0.74 (0.10)** |
| Black | 0.13 (0.05)** | 0.20 (0.04)** | 0.15 (0.05)** | 0.29 (0.05)** |
| White | 0.10 (0.24) | 0.10 (0.06) | 0.27 (0.05)** | 0.32 (0.08)** |
| Other race | 0.19 (0.07)** | 0.12 (0.05)* | 0.00 (0.06) | 0.28 (0.15) |
| At-risk promoted | –0.03 (0.10) | –0.07 (0.07) | 0.07 (0.06) | 0.03 (0.09) |
| Not at risk | 0.36 (0.09)** | 0.37 (0.06)** | 0.33 (0.05)** | 0.22 (0.08)** |

NOTE: Standard errors are in parentheses.

* = statistically significant at the 0.04 level.

** = statistically significant at the 0.01 level.

**Table E.13**
**Standardized Regression Coefficients for the 3rd-Grade Cohort Two to Five Years After the Retention Decision**

| Variable | 2 Years | 3 Years | 4 Years | 5 Years |
|---|---|---|---|---|
| School belonging | | | | |
| At-risk promoted | –0.40 (0.08)** | –0.32 (0.08)** | –0.19 (0.06)** | –0.11 (0.09) |
| Not at risk | –0.29 (0.07)** | –0.21 (0.06)** | –0.25 (0.05)** | –0.17 (0.09) |
| Mathematics confidence | | | | |
| At-risk promoted | –0.15 (0.09) | –0.17 (0.07)* | –0.08 (0.06) | 0.06 (0.10) |
| Not at risk | 0.08 (0.06) | –0.01 (0.06) | 0.09 (0.05) | 0.07 (0.08) |
| Reading confidence | | | | |
| At-risk promoted | –0.03 (0.10) | –0.07 (0.07) | 0.07 (0.06) | 0.03 (0.09) |
| Not at risk | 0.36 (0.09)** | 0.37 (0.06)** | 0.33 (0.05)** | 0.22 (0.08)** |

NOTE: Standard errors are in parentheses.
* = statistically significant at the 0.04 level.
** = statistically significant at the 0.01 level.

# References

Alexander, Karl L., Doris R. Entwisle, and Susan L. Dauber. (1994). *On the Success of Failure: A Reassessment of the Effects of Retention in the Primary School Grades*. Cambridge, UK: Cambridge University Press, 1st edition.

———. (2003). *On the Success of Failure: A Reassessment of the Effects of Retention in the Primary School Grades*. Cambridge, UK: Cambridge University Press, 2nd edition.

Alexander, Karl L., Doris R. Entwisle, and Nader Kabbani. (2001). "The Dropout Process in Life Course Perspective: Early Risk Factors at Home and School." *Teachers College Record*, 103(5), 760–822.

Allensworth, Elaine M. (2004). *Ending Social Promotion: Dropout Rates in Chicago After Implementation of the Eighth-Grade Promotion Gate*. Chicago, IL: Consortium for Chicago School Research.

———. (2005). "Dropout Rates After High-Stakes Testing in Elementary School: A Study of the Contradictory Effects of Chicago's Efforts to End Social Promotion." *Educational Evaluation and Policy Analysis*, 27(4), 341–364.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for Educational Psychological Testing.* Washington, DC: American Psychological Association.

Baenen, Nancy R. (April 1988). *Perspectives After Five Years—Has Grade Retention Passed or Failed?* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Baker, Dwayne, and Peter A. Witt. (1996). "Evaluation of the Impact of Two After-School Recreation Programs." *Journal of Park and Recreation Administration*, 14(3), 23–44.

Bang, Heejung, and James M. Robins. (2005). "Doubly Robust Estimation and Missing Data and Causal Inference Models." *Biometrics*. 61(4), 962–972.

Battistin, Erich, and Enrico Rettore. (2002). *Another Look at the Regression Discontinuity Design*. As of August 5, 2009:
http://www.cepr.org/meets/wkcn/4/4528/papers/rettore.pdf

Benjamini, Yoav, and Yosef Hochberg. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1), 289–300.

Berends, Mark, Susan J. Bodilly, and Sheila Nataraj Kirby. (2002). *Facing the Challenges of Whole-School Reform: New American Schools After a Decade*. Santa Monica, CA: RAND Corporation. MR-1498-EDU. As of August 5, 2009:
http://www.rand.org/pubs/monograph_reports/MR1498/

Blair, Clancy (2001). "The Early Identification of Risk for Grade Retention Among African American Children at Risk for School Difficulty." *Applied Developmental Science*, 5(1), 37–50.

Bodilly, Susan J., and Megan K. Beckett. (2005). *Making Out-of-School Time Matter: Evidence for an Action Agenda*. Santa Monica, CA: RAND Corporation. MG-242-WF. As of August 5, 2009: http://www.rand.org/pubs/monographs/MG242/

Bonvin, Patrick, Gerard Bless, and Marianne Schuepbach. (2008). "Grade Retention: Decision-Making and Effects on Learning as Well as Social and Emotional Development." *School Effectiveness and School Improvement*, 19(1), 1–19.

Bryk, Anthony S., Valerie E. Lee, and Peter B. Holland. (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.

Byrd, Robert S., and Michael Weitzman. (1994). "Predictors of Early Grade Retention Among Children in the United States." *Pediatrics*, 93(3), 481–487.

Carnegie Council on Adolescent Development. (1989). *Turning Points: Preparing American Youth for the 21st Century*. New York: Carnegie Corporation.

Chaput, Sandra Simpkins. (2004). "Characterizing and Measuring Participation in Out-of-School Time Programs." *Evaluation Exchange*, 10(1), 2–3, 29.

Cohen, Jacob. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 2nd edition.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. (1966). *Equality of Educational Opportunity*. Washington, DC: Government Printing Office.

Cook, Thomas D. (2008). "'Waiting for Life to Arrive': A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics." *Journal of Econometrics*, 142(2), 636–654.

Cooper, Harris, Kelly Charlton, Jeff C. Valentine, and Laura Muhlenbruck. (2000). *Making the Most of Summer School: A Meta-Analytic and Narrative Review*. Malden, MA: Monographs of the Society for Research in Child Development.

Corman, Hope. (2003). "The Effects of State Policies, Individual Characteristics, Family Characteristics, and Neighbourhood Characteristics on Grade Repetition in the United States." *Economics of Education Review*, 22(4), 409–420.

Dauber, Susan L., Karl L. Alexander, and Doris R. Entwisle. (1993). "Characteristics of Retainees and Early Precursors of Retention in Grade: Who Is Held Back?" *Merrill-Palmer Quarterly*, 39(3), 326–343.

Eccles, Jacquelynne S., and Carol Midgley. (1989). "Stage-Environment Fit: Developmentally Appropriate Classrooms for Early Adolescents." In Russell Ames and Carol Ames (Eds.), *Research on Motivation in Education*, Vol. 3 (pp. 139–186). San Diego, CA: Academic Press.

Eide, Eric R., and Dan D. Goldhaber. (2005). "Grade Retention: What Are the Costs and Benefits?" *Journal of Education Finance*, 31(2), 195–214.

Eide, Eric R., and Mark H. Showalter. (2001). "The Effect of Grade Retention on Education and Labor Market Outcomes." *Economics of Education Review*, 20(6), 563–576.

El-Hassan, Karma. (1998). "Relation of Academic History and Demographic Variables to Grade Retention in Lebanon." *Journal of Educational Research*, 91(5), 279–288.

Elmore, Richard F., and Deanna Burney. (1999). "Investing in Teacher Learning: Staff Development and Instructional Improvement." In Linda Darling-Hammond and Gary Sykes (Eds.), *Teaching as the Learning Profession: Handbook of Policy and Practice* (pp. 263–291). San Francisco, CA: Jossey-Bass.

Fager, Jennifer, and Rae Richen. (1999). *When Students Don't Succeed: Shedding Light on Grade Retention*. Portland, OR: Northwest Regional Educational Laboratory.

Ferguson, Phillip, Shane R. Jimerson, and Michael J. Dalton. (2001). "Sorting Out Successful Failures: Exploratory Analyses of Factors Associated with Academic and Behavioral Outcomes of Retained Students." *Psychology in the Schools*, 38(4), 327–341.

Fine, Jodene Goldenring, and John M. Davis. (2003). "Grade Retention and Enrollment in Post-Secondary Education." *Journal of School Psychology*, 41(6), 401–411.

Finn, Jeremy. (1993). *School Engagement and Students at Risk*. Washington, DC: National Center for Education Statistics. No. 93470.

Frederick, Carl B., and Robert M. Hauser. (2006). *Have We Put an End to Social Promotion? Changes in Grade Retention Rates Among Children Aged 6 to 17 from 1972 to 2003*. Madison, WI: Center for Demography of Health and Aging, University of Wisconsin–Madison.

———. (2008). "Have We Put an End to Social Promotion? Changes in School Progress Among Children Aged 6 to 17 from 1972 to 2005." *Demography*, 45(3), 719–740.

Fruchter, Norm. (April 2006). "Mayoral Control in New York City." Paper presented at the American Education Research Association, Montreal.

Frymier, Jack. (1997). "Characteristics of Students Retained in Grade." *High School Journal*, 80(3), 184–192.

Gamoran, Adam. (1987). "The Stratification of High School Learning Opportunities." *Sociology of Education*, 60, 135–155.

———. (1992). "The Variable Effects of High School Tracking." *American Sociological Review*, 57, 812–828.

GAO—*see* U.S. Government Accountability Office.

Gilman, Rich. (2001). "The Relationship Between Life Satisfaction, Social Interest, and Frequency of Extracurricular Activities Among Adolescent Students." *Journal of Youth and Adolescence*, 30, 749–767.

Goodenow, Carol. (1993). "Classroom Belonging Among Early Adolescent Students: Relationships to Motivation and Achievement." *Journal of Early Adolescence*, 13(1), 21–43.

Goldschmidt, Pete, and Jia Wang. (1999). "When Can Schools Affect Dropout Behavior? A Longitudinal Multilevel Analysis." *American Educational Research Journal*, 36(4), 715–738.

Gottfredson, Denise C., Carolyn Molden Fink, and Nanette Graham. (1994). "Grade Retention and Problem Behavior." *American Educational Research Journal*, 31(4), 761–784.

Greene, Jay P., and Marcus A. Winters. (2004). *An Evaluation of Florida's Program to End Social Promotion*. New York: Center for Civic Innovation, Manhattan Institute for Policy Research. Education Working Paper No. 7.

———. (2006). *Getting Farther Ahead by Staying Behind: A Second-Year Evaluation of Florida's Policy to End Social Promotion*. New York: Center for Civic Innovation, Manhattan Institute for Policy Research. Civic Report No. 49.

———. (2007). "Revisiting Grade Retention: An Evaluation of Florida's Test-Based Promotion Policy." *Education Finance and Policy*, 2(4), 319–340.

———. (2009). "The Effects of Exemptions to Florida's Test-Based Promotion Policy: Who Is Retained? Who Benefits Academically?" *Economics of Education Review*, 28(1), 135–142.

Grissom, James B., and Lorrie A. Shepard. (1989). “Repeating and Dropping Out of School.” In Lorrie A. Shepard and Mary Lee Smith (Eds.), *Flunking Grades: Research and Policies on Retention* (pp. 34–63). London: The Falmer Press.

Guevremont, Anne, Noralou P. Roos, and Marnie Brownell. (2007). “Predictors and Consequences of Grade Retention: Examining Data from Manitoba, Canada.” *Canadian Journal of School Psychology*, 22(1), 50–67.

Hagborg, Winston J., Gaetano Masella, Phillip Palladino, and Jerry Shepardson. (1991). “A Follow-Up Study of High School Students with a History of Grade Retention.” *Psychology in the Schools*, 28(4), 310–317.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. (2001). “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.” *Econometrica*, 69(1), 201–209.

Hartke, Karen. (1999). “The Misuse of Tests for Retention.” *Thrust for Educational Leadership*, 28(3), 22–25.

Hauser, Robert M. (1999). *Should We End Social Promotion? Truth and Consequences*. Madison, WI: Center for Demography and Ecology, University of Madison–Wisconsin. CDE Working Paper No. 99-06.

Hauser, Robert M., Carl B. Frederick, and Megan Andrew. (2007). “Grade Retention in the Age of Standards-Based Reform.” In Adam Gamoran (Ed.) *Standards-Based Reform and the Poverty Gap: Lessons for No Child Left Behind* (pp. 120–153). Washington, DC: Brookings Institution Press.

Hauser, Robert M., Devah Pager, and Solon J. Simmons. (2000). *Race-Ethnicity, Social Background, and Grade Retention*. Madison, WI: Center for Demography and Ecology, University of Wisconsin–Madison. CDE Working Paper No. 2000-08.

Heubert, Jay P., and Robert M. Hauser, eds. (1999). *High Stakes: Testing for Tracking, Promotion and Graduation*. Washington, DC: National Academies Press.

Holmes, C. Thomas. (1989). “Grade Level Retention Effects: Meta-Analysis of Research Studies.” In Lorrie A. Shepard and Mary Lee Smith (Eds.), *Flunking Grades: Research and Policies on Retention* (pp. 16–33). London: The Falmer Press.

Holmes, C. Thomas, and Kenneth M. Matthews. (1984). “The Effects of Nonpromotion on Elementary and Junior High School Pupils: A Meta-Analysis.” *Review of Educational Research*, 54(2), 225–236.

Holmes, C. Thomas, and Janice Saturday. (2000). “Promoting the End of Retention.” *Journal of Curriculum and Supervision*, 15(4), 300–314.

Hong, Guanglei, and Stephen W. Raudenbush. (2005). “Effects of Kindergarten Retention Policy on Children’s Cognitive Growth in Reading and Mathematics.” *Educational Evaluation and Policy Analysis*, 27(3), 205–224.

Hong, Guanglei, and Bing Yu. (2007). “Early-Grade Retention and Children’s Reading and Math Learning in Elementary Years.” *Educational Evaluation and Policy Analysis*, 29(4), 239–261.

———. (2008). “Effects of Kindergarten Retention on Children’s Social-Emotional Development: An Application of Propensity Score Method to Multivariate, Multilevel Data.” *Developmental Psychology*, 44(2), 407–421.

Imbens, Guido, and Thomas Lemieux. (2008). “Regression Discontinuity Designs: A Guide to Practice.” *Journal of Econometrics*, 142(2), 615–635.

Jacob, Brian A., and Lars Lefgren. (2002). *Remedial Education and Student Achievement: A Regression-Discontinuity Analysis*. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 8918.

———. (2007). *The Effect of Grade Retention on High School Completion*. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 13514.

Jacob, Robin Tepper, and Susan Stone. (2005). "Educators and Students Speak: Those Closest to the Action Like the Retention Policy." *Education Next*, 5(1), 49–53.

Jacob, Robin Tepper, Susan Stone, and Melissa Roderick. (2004). *Ending Social Promotion: The Response of Teachers and Students*. Chicago, IL: Consortium for Chicago School Research.

Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Ginitis, Barbara Heyns and Stephan Michaelson. (1972). *Inequality: A Reassessment of Family and Schooling in America*. New York: Basic Books.

Jimerson, Shane R. (1999). "On the Failure of Failure: Examining the Association Between Early Grade Retention and Education and Employment Outcomes During Late Adolescence." *Journal of School Psychology*, 37(3), 243–272.

———. (2001). "Meta-Analysis of Grade Retention Research: Implications for Practice in the 21st Century." *School Psychology Review*, 30(3), 420–438.

Jimerson, Shane R., Gabrielle E. Anderson, and Angela D. Whipple. (2002). "Winning the Battle and Losing the War: Examining the Relation Between Grade Retention and Dropping Out of High School." *Psychology in the Schools*, 39(4), 441–457.

Jimerson, Shane R., Elizabeth Carlson, Monique Rotert, Byron Egeland, and L. Alan Sroufe. (1997). "A Prospective, Longitudinal Study of the Correlates and Consequences of Early Grade Retention." *Journal of School Psychology*, 35(1), 3–25.

Jimerson, Shane R., and Phillip Ferguson. (2007). "A Longitudinal Study of Grade Retention: Academic and Behavioral Outcomes of Retained Students Through Adolescence." *School Psychology Quarterly*, 22(3), 314–339.

Jimerson, Shane R., Phillip Ferguson, Angela D. Whipple, Gabrielle E. Anderson, and Michael J. Dalton. (2002). "Exploring the Association Between Grade Retention and Dropout: A Longitudinal Study Examining Socio-Emotional, Behavioral, and Achievement Characteristics of Retained Students." *California School Psychologist*, 7, 51–62.

Juvonen, Jaana, Vi-Nhuan Le, Tessa Kaganoff, Catherine H. Augustine, and Louay Constant. (2004). *Focus on the Wonder Years: Challenges Facing the American Middle School*. Santa Monica, CA: RAND Corporation. MG-139-EDU. As of August 5, 2009: http://www.rand.org/pubs/monographs/MG139/

Karweit, Nancy L. (1999). *Grade Retention: Prevalence, Timing, and Effects*. Baltimore, Md.: Center for Research on the Education of Students Placed at Risk. Report No. 33.

Kass, Robert E., and Adrian E. Raftery. (1995). "Bayes Factors." *Journal of the American Statistical Association*, 90, 773–795.

Kazdin, Alan. (1993). "Adolescent Mental Health: Prevention and Treatment Programs." *American Psychologist*, 48, 127–141.

Klonsky, Michael. (2002). "How Smaller Schools Prevent School Violence." *Educational Leadership*, 59(5), 65–69.

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

———. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge, MA: Harvard University Press.

Koretz, Daniel M., Daniel F. McCaffrey, and Laura S. Hamilton. (2001). *Toward a Framework for Validating Gains Under High-Stakes Conditions*. Washington, DC: U.S. Department of Education, Center for the Study of Evaluation. Report No. 551.

Lee, David S., and David E. Card. (2008). "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics*, 142(2), 655–674.

Lee, David S., and Thomas Lemieux. (2009). *Regression Discontinuity Designs in Economics*. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 14723.

Lee, Valerie E., Douglas D. Ready, and Kevin Welner. (2002) *Educational Equity and School Structure: School Size, School Overcrowding, and Alternative Organizational Structures*. Los Angeles, CA: Institute for Democracy, Education, and Access, University of California, Los Angeles. Williams Watch Series: Investigating the Claims of Williams v. State of California, No. WWS-RR015-1002.

Liddell, Christine, and Gordon Rae. (2001). "Predicting Early Grade Retention: A Longitudinal Investigation of Primary School Progress in a Sample of Rural South African Children." *British Journal of Educational Psychology*, 71(3), 413–428.

Lin, Xihong, and Daowen Zhang. (1999). "Inference in Generalized Additive Mixed Models by Using Smoothing Splines." *Journal of the Royal Statistical Society, Series B*, 61, 381–400.

Linn, Robert L. (2000). "Assessments and Accountability." *Educational Researcher*, 29(2), 4–16.

Lorence, Jon, and Anthony Gary Dworkin. (August 2004). "Grade Retention and School Dropout Behavior in Texas." Paper presented at the annual meeting of the American Sociological Association, San Francisco, CA.

———. (2006). "Elementary Grade Retention in Texas and Reading Achievement Among Racial Groups: 1994–2002." *Review of Policy Research*, 23(5), 999–1033.

Lorence, Jon, A. Gary Dworkin, Laurence A. Toenjes, and Antwanette N. Hill. (2002). "Grade Retention and Social Promotion in Texas 1994–99: Academic Achievement Among Elementary School Students." In Diane Ravitch (Ed.), *Brookings Papers on Education Policy 2002* (pp. 13–67). Washington, DC: Brookings Institution Press.

Ludwig, Jens, and Douglas L. Miller. (2005). *Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design*. Cambridge, MA: National Bureau of Economic Research. Working Paper 11702.

Mantzicopoulos, Panayota Y., and Delmont C. Morrison. (1992). "Kindergarten Retention: Academic and Behavioral Outcomes Through the End of Second Grade." *American Educational Research Journal*, 29(1), 182–198.

Mantzicopoulos, Panayota Y., Delmont C. Morrison, Stephen P. Hinshaw, and Estol T. Carte. (1989). "Nonpromotion in Kindergarten: The Role of Cognitive, Perceptual, Visual-Motor, Behavioral, Achievement, Socioeconomic, and Demographic Characteristics." *American Educational Research Journal*, 26(1), 107–121.

Marcon, Rebecca. (March 1993). "At-Risk Preschoolers: Early Predictors of Future Grade Retention." Paper presented at the annual meeting of the Southeastern Psychological Association, Atlanta, GA.

Marsh, Julie A. (2002). "How Districts Relate to States, Schools, and Communities: A Review of Emerging Literature." In Amy M. Hightower, Michael S. Knapp, Julie A. Marsh, and Milbrey W. McLaughlin (Eds.), *School Districts and Instructional Renewal* (pp. 25–40). New York: Teachers College Press.

Marsh, Julie A., Daniel Gershwin, Sheila Nataraj Kirby, and Nailing Xia. (2009). *Retaining Students in Grade: Lessons Learned Regarding Policy Design and Implementation*. Santa Monica, CA: RAND Corporation. TR-677-NYCDOE. As of October 2009:
http://www.rand.org/pubs/technical_reports/TR677

Martin, Michael O., Ina V. S. Mullis, and Steven J. Chrostowski. (2004). *Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston, MA: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College.

Matsudaira, Jordan D. (2008). "Mandatory Summer School and Student Achievement." *Journal of Econometrics*, 142(2), 829–850.

McCaffrey, Daniel F., Greg Ridgeway, and Andrew R. Morral. (2004). "Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies." *Psychological Methods*, 9(4), 403–425.

McCoy, Ann R., and Arthur J. Reynolds. (1999). "Grade Retention and School Performance: An Extended Investigation." *Journal of School Psychology*, 37(3), 273–298.

McLaughlin, Milbrey W., and Joan Talbert. (2002). "Reforming Districts." In Amy M. Hightower, Michael S. Knapp, Julie A. Marsh, and Milbrey W. McLaughlin (Eds.), *School Districts and Instructional Renewal* (pp. 173–192). New York: Teachers College Press.

Meisels, Samuel J., and Fong-Ruey Liaw. (1993). "Failure in Grade: Do Retained Students Catch Up?" *Journal of Educational Research*, 87(2), 69–77.

Nagaoka, Jenny, and Melissa Roderick. (2004). *Ending Social Promotion: The Effects of Retention*. Chicago, IL: Consortium on Chicago School Research.

National Center for Education Statistics. (Undated). "The Nation's Report Card." Web page. As of August 5, 2009:
http://nces.ed.gov/nationsreportcard/

New York City Department of Education. (Undated[a]). "About Our Schools: Children First." Web page. As of August 5, 2009:
http://schools.nyc.gov/AboutUs/schools/childrenfirst.htm

———. (Undated[b]). "Educator Resources: Professional Development Offerings." Web page. Current listings, as of August 5, 2009:
http://schools.nyc.gov/StudentSupport/AcademicInterventionServices/EducatorResources/Office+of+Academic+Intervention+Professional+Development+Offerings.htm

———. (Undated[c]). "Quick Reference Guide." As of April 2009:
http://schools.nyc.gov/NR/rdonlyres/EB063CC7-8CB9-4FB0-B071-9442F73F4C7D/7002/DoEJargonReferenceGuideFINAL.doc

———. (2007). "Schools Chancellor Klein Announces New School Support Organizations to Provide Services to Schools Beginning in the 2007–08 School Year." Press release. April 16.

———. (2008). "Children First Intensive Inquiry Profile Search." Web site, study data accessed November 12. As of August 5, 2009:
http://www.childrenfirstintensive.com/

———. (2009). Regulation of the Chancellor A-501: Promotion Standards, June 29. As of August 5, 2009:
http://docs.nycenet.edu/docushare/dsweb/Get/Document-24/A-501.pdf

New York State Education Department. (Undated). "Elementary, Middle, Secondary and Continuing Education: Technical Reports." Web page. As of August 5, 2009:
http://www.emsc.nysed.gov/osa/pub/reports.shtml

———. (2008). “2008 Grade 3–8 English Language Arts and Math Results.” Briefing. As of August 5, 2009: http://www.emsc.nysed.gov/irts/pressRelease/20080623/Grade3-8MathEng-testsfinal.ppt

———. (2009). “English Language Arts (ELA) and Mathematic [sic] Assessment Results.” Web page, last updated June 30. As of August 5, 2009: http://www.emsc.nysed.gov/irts/ela-math/

Niklason, Lucille B. (1987). “Do Certain Groups of Children Profit from a Grade Retention?” *Psychology in the Schools*, 24(4), 339–345.

NYCDOE—*see* New York City Department of Education.

NYSED—*see* New York State Education Department.

Ou, Suh-Ruu, and Arthur J. Reynolds. (2008). “Predictors of Educational Attainment in the Chicago Longitudinal Study.” *School Psychology Quarterly*, 23(2), 199–229.

Owings, William A., and Susan Magliaro. (1998). “Grade Retention: A History of Failure.” *Educational Leadership*, 56(1), 86–88.

Peterson, Sarah E., James S. DeGracie, and Carol R. Ayabe. (1987). “A Longitudinal Study of the Effects of Retention/Promotion on Academic Achievement.” *American Educational Research Journal*, 24(1), 107–118.

Pierson, Louisa H., and James P. Connell. (1992). “Effect of Grade Retention on Self-System Processes, School Engagement, and Academic Performance.” *Journal of Educational Psychology*, 84(3), 300–307.

Plummer, Diane L., and William G. Graziano. (1987). “Impact of Grade Retention on the Social Development of Elementary School Children.” *Developmental Psychology,* 23(2), 267–275.

Pomplun, Mark. (1988). “Retention: The Earlier, the Better?” *Journal of Educational Research*, 81(5), 281–287.

Public Law 89-10, Elementary and Secondary Education Act of 1965, April 11, 1965.

Public Law 103-382, Improving America’s Schools Act of 1994, October 20, 1994.

Public Law 107-110, No Child Left Behind Act of 2001, January 8, 2002.

Raudenbush, Stephen W., and Anthony S. Bryk. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications, 2nd edition.

Reynolds, Arthur J. (1992). “Grade Retention and School Adjustment: An Explanatory Analysis.” *Educational Evaluation and Policy Analysis*, 14(2), 101–121.

Robinson, Marian A., Patricia Kannapel, Joan Gujarati, Hakim Williams, and Andrea Oettinger. (2008). *A Formative Study of the Implementation of the Inquiry Team Process in New York City Public Schools: 2007–08 Findings*. New York: Consortium for Policy Research in Education, Teachers College, Columbia University.

Robles-Pina, Rebecca A., Emily Defrance, and Deborah L. Cox. (2008). “Self-Concept, Early Childhood Depression and School Retention as Predictors of Adolescent Depression in Urban Hispanic Adolescents.” *School Psychology International*, 29(4), 426–441.

Roderick, Melissa. (1994). “Grade Retention and School Dropout: Investigating the Association.” *American Educational Research Journal*, 31(4), 729–759.

Roderick, Melissa, Anthony S. Bryk, Brian A. Jacob, John Q. Easton, and Elaine Allensworth. (1999). *Ending Social Promotion: Results from the First Two Years*. Chicago, IL: Consortium on Chicago School Research.

Roderick, Melissa, and Mimi Engel. (2001). "The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-Stakes Testing." *Educational Evaluation and Policy Analysis*, 23(3), 197–227.

Roderick, Melissa, Mimi Engel, and Jenny Nagaoka, with Brian A. Jacob, Sophie Degener, Alex Orfei, Susan Stone, and Jen Bacon. (2003). *Ending Social Promotion: Results from Summer Bridge*. Chicago, IL: Consortium on Chicago School Research.

Roderick, Melissa, and Jenny Nagaoka. (2005). "Retention Under Chicago's High-Stakes Testing Program: Helpful, Harmful, or Harmless?" *Educational Evaluation and Policy Analysis*, 27(4), 309–340.

Roderick, Melissa, Jenny Nagaoka, and Elaine Allensworth. (2005). "Is the Glass Half Full or Mostly Empty? Ending Social Promotion in Chicago." *Yearbook of the National Society for the Study of Education*, 104(2), 223–259.

Rosenbaum, Paul R., and Donald R. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70(1), 41–55.

Rosenholtz, Susan J. (1991). *Teacher's Workplace: The Social Organization of Schools*. New York: Teachers College Press.

Rumberger, Russell W., and Katherine A. Larson. (1998). "Student Mobility and the Increased Risk of High School Dropout." *American Journal of Education*, 107(1), 1–35.

Rush, Sheila, and Patrick A. Vitale. (1994). "Analysis for Determining Factors That Place Elementary Students at Risk." *Journal of Educational Research*, 87, 325–333.

Safer, Daniel J. (1986). "Nonpromotion Correlates and Outcomes at Different Grade Levels." *Journal of Learning Disabilities*, 19(8), 500–503.

Schwarz, Gideon. (1978). "Estimating the Dimension of a Model." *Annals of Statistics*, 6(2), 461–464.

Setencich, Jill. (1994). "The Impact of Early Grade Retention on the Academic Achievement and Self-Esteem of Seventh and Eighth Grade Students." Paper presented at the Annual Convention of the National Association of School Psychologists, Seattle, WA, March.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shepard, Lorrie A., and Mary Lee Smith. (1987). "Effects of Kindergarten Retention at the End of First Grade." *Psychology in the Schools*, 24(4), 346–357.

———. (1989). "Academic and Emotional Effects of Kindergarten Retention in One School District." In Lorrie A. Shepard and Mary Lee Smith (Eds.), *Flunking Grades: Research and Policies on Retention* (pp. 79–107). London: The Falmer Press.

———. (1990). "Synthesis of Research on Grade Retention." *Educational Leadership*, 47(8), 84–88.

Silberglitt, Benjamin, James J. Appleton, Matthew K. Burns, and Shane R. Jimerson. (2006). "Examining the Effects of Grade Retention on Student Reading Performance: A Longitudinal Study." *Journal of School Psychology*, 44(4), 255–270.

Silberglitt, Benjamin, Shane R. Jimerson, Matthew K. Burns, and James J. Appleton. (2006). "Does the Timing of Grade Retention Make a Difference? Examining the Effects of Early Versus Later Retention." *School Psychology Review*, 35(1), 134–141.

Simmons, Roberta G., and Dale A. Blyth. (1987). *Moving into Adolescence: The Impact of Pubertal Change and School Context*. Hawthorn, NY: Random House.

Snipes, Jason, Fred Doolittle, and Corrine Herlihy. (2002). *Foundations for Success: Case Studies of How Urban Schools Improve Student Achievement*. New York: MDRC.

Spillane, James P. (1996). "School Districts Matter: Local Educational Authorities and State Instructional Policy. *Educational Policy*, 10(1), 63–87.

Sunderman, Gail L., and Jimmy Kim. (2004). *Increasing Bureaucracy or Increasing Opportunities? School District Experience with Supplemental Education Services*. Cambridge, MA: Civil Rights Project, Harvard University.

Tanner, C. Kenneth, and Susan Allan Galis. (1997). "Student Retention: Why Is There a Gap Between the Majority of Research Findings and School Practice?" *Psychology in the Schools*, 34(2), 107–114.

Thompson, Charles L., and Elizabeth K. Cunningham. (2000). "Retention and Social Promotion: Research and Implications for Policy." *ERIC Clearinghouse on Urban Education Digest*, 161, 1–5.

Togneri, Wendy, and Stephen E. Anderson. (2003). *Beyond Islands of Excellence: What Districts Can Do to Improve Instruction and Achievement in All Schools*. Washington, DC: Learning First Alliance.

U.S. Government Accountability Office. (2006). *The No Child Left Behind Act: Education Actions Needed to Improve Local Implementation and State Evaluation of Supplemental Educational Services*. Washington, DC. GAO-06-758.

Van der Klaauw, Wilbert. (2002). "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach." *International Economic Review*, 43(4), 1249–1287.

Vernez, Georges, Scott Naftel, Karen E. Ross, Kerstin C. Le Floch, Christopher Beighley, Brian Gill, Beatrice Birman, Michael Garet, and Jennifer O'Day. (2009). *State and Local Implementation of the No Child Left Behind Act*, Vol. VII: *Title I School Choice and Supplemental Educational Services: Final Report*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Development. As of August 5, 2009:
http://www.rand.org/pubs/reprints/RP1383/

Wasik, Barbara A., and Robert E. Slavin. (1993). "Preventing Early Reading Failure with One-to-One Tutoring: A Review of Five Programs." *Reading Research Quarterly*, 28(2), 178–200.

Weiss, Carol H. (1972). *Evaluation*. Englewood Cliffs, NJ: Prentice Hall.

West, Stephen D., Jeremy C. Biesanz, and Steven C. Pitts. (2000). "Causal Inference and Generalization in Field Settings: Experimental and Quasi-Experimental Designs." In Harry T. Resi and Charles M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 40–88). New York: Cambridge University Press.

Willson, Victor L., and Jan N. Hughes. (2006). "Retention of Hispanic/Latino Students in First Grade: Child, Parent, Teacher, School, and Peer Predictors." *Journal of School Psychology*, 44(1), 31–49.

Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Wu, Wei, Stephen G. West, and Jan N. Hughes. (2008). "Short-Term Effects of Grade Retention on the Growth Rate of Woodcock-Johnson III Broad Math and Reading Scores." *Journal of School Psychology*, 46(1), 85–105.

Xia, Nailing, and Sheila Nataraj Kirby. (2009). *Retaining Students in Grade: A Literature Review of the Effects of Retention on Students' Academic and Nonacademic Outcomes*. Santa Monica, CA: RAND Corporation. TR-678-NYCDOE. As of October 2009:
http://www.rand.org/pubs/technical_reports/TR678

Zinth, Kyle. (2005). *Student Promotion/Retention Policies*. Denver, CO: Education Commission of the States.