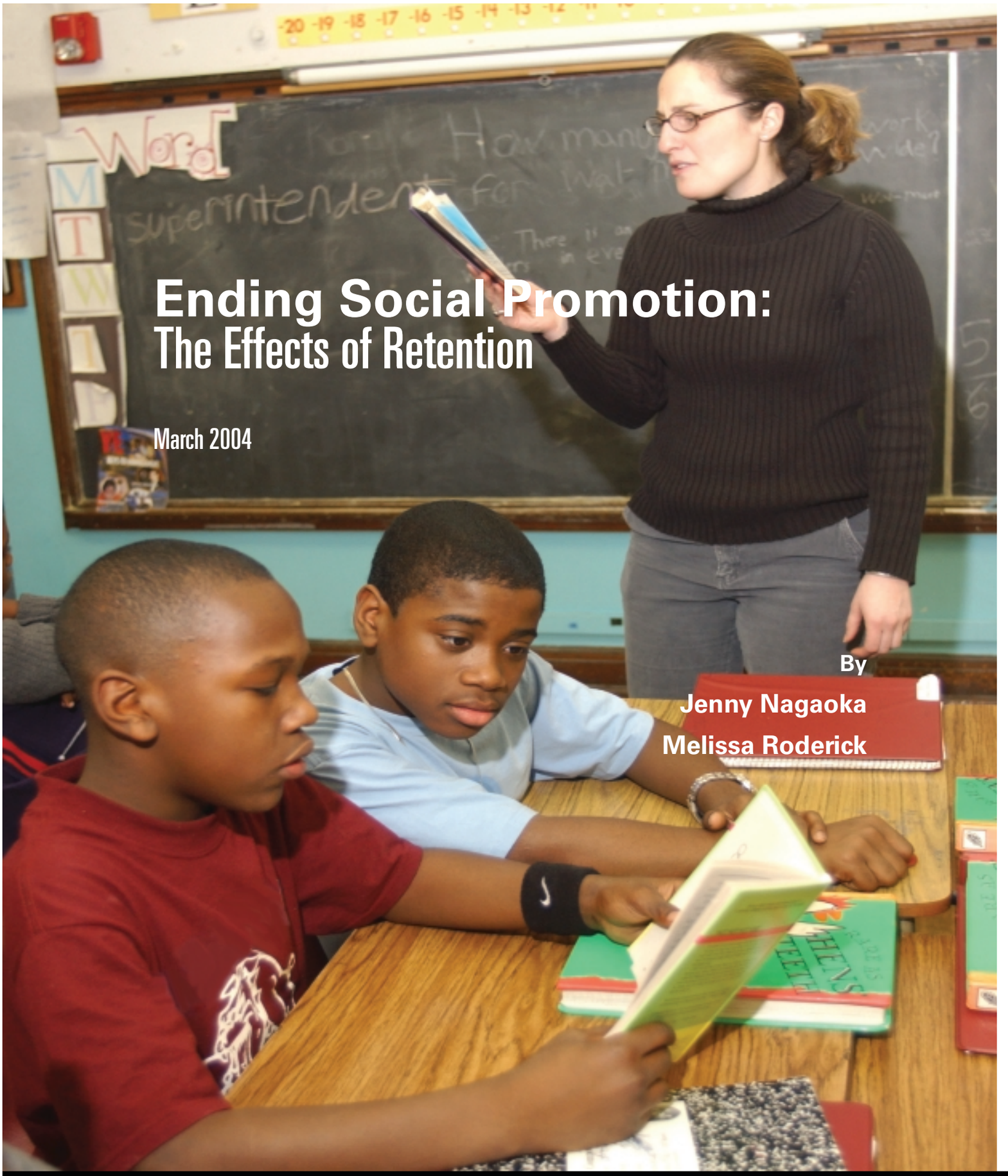


Ending Social Promotion:
The Effects of Retention

March 2004

By

Jenny Nagaoka

Melissa Roderick

CHARTING REFORM IN CHICAGO SERIES

CONSORTIUM ON CHICAGO SCHOOL RESEARCH

ACKNOWLEDGMENTS

This report is the culmination of many years of research on Chicago's efforts to end social promotion. Many individuals have helped shape this work. We would like to acknowledge the researchers who have contributed to this report, including Brian Jacob, Mimi Engel, Susan Stone, and Alex Orfei. This project was greatly enriched by our valuable feedback from Dick Murnane and Lorrie Shepard. We are indebted to the support of all Consortium directors and analysts who gave helpful comments throughout this process, especially John Q. Easton, Elaine Allensworth, Penny Sebring, and Anthony S. Bryk. Thanks also to the members of the Consortium's Steering Committee for their comments, particularly Fred Hess.

We would like to thank former and current staff at the Chicago Public Schools who provided technical advice and supported our field work in schools throughout the course of this project. We are particularly indebted to the students, teachers, principals, and staff in the Chicago Public Schools who graciously welcomed our researchers into their buildings and classrooms.

The administrative staff at the School of Social Service Administration—Keith Madderom, Beverly Mason, and Suzanne Fournier—provided significant support to this work. SSA's staff made us feel that our work was always important and that we were never asking too much.

The study is funded by grants from The Spencer Foundation, The Joyce Foundation, and The Frye Foundation. The contents of this report were also developed under a grant from the Department of Education. However, those contents do not necessarily represent the policy of the Department of Education, and the reader should not assume endorsement by the Federal Government.

Ending Social Promotion:

The Effects of Retention

March 2004

Jenny Nagaoka

Melissa Roderick

TABLE OF CONTENTS

CONSORTIUM ON CHICAGO SCHOOL RESEARCH

John Booz

Executive Summary

In 1996, the Chicago Public Schools made national headlines when it declared an "end to social promotion" and instituted promotional requirements based on students' scores on the Iowa Tests of Basic Skills in the third, sixth, and eighth grades. Since then, CPS has been retaining between 7,000 and 10,000 students per year in these three grades.

Over the past five years, Consortium researchers have been evaluating the effects of the CPS initiative. In this report, we focus on those students who did not meet CPS's promotional cutoff, the majority of whom were retained. We examine how the administration of the policy shaped the experience of retained students and evaluate the impact of retention on students' achievement growth and their progress in school.

CPS adopted a rather traditional approach to retention with one twist—an additional opportunity in January for retained students to pass the promotional requirements and then rejoin their age-appropriate classmates. The central idea behind retention was that students who did not learn the material the first time around would benefit from an additional year of instruction. However, Chicago provided little structure or additional resources for schools to support the learning needs of retained students, and the retained year amounted to students going through the grade a second time with the same curricular and instructional experience.

What Happened to Retained Students during Their Retained Year?

A first and most basic approach to evaluating the success of retaining students under high-stakes testing is to ask: How successful were retained students in raising their test scores to the promotional standard during their retained year?

- ***Retained students struggled their second time trying to meet the promotional standards***. In the second and third year of the policy, the system added an extra chance to pass the promotional gate in January and an opportunity for students to rejoin their classmates. Even with this extra chance, less than 60 percent of retained third and sixth graders in 1998 and 1999 were able to raise their test scores to the promotional cutoff.

- ***Close to 20 percent of retained third and sixth graders were placed in special education within two years of the retention decision.*** Part of the reason for the low passing rates was that a high proportion of these students were placed in special education during their retained year. In addition, retained students who did not meet the cutoff continued to face high rates of special education placement. On average, 18 percent of third and sixth graders retained in the 1997-1998, 1998-1999, and 1999-2000 school years were placed in special education within two years of the retention decision—a rate three times that of low-achieving students prior to the introduction of the policy and nearly three times higher than other low-achieving students who were promoted under the policy.

- *The January test resulted in many retained students rejoining their age-appropriate classmates.* Over one-quarter of third graders and one-third of sixth graders who were retained in 1998 and 1999 rejoined their age-appropriate classmates, ultimately skipping a grade. However, in 2001, the district discontinued the January test and the rate of double promotions fell.

Did Retention and the Various Experiences of Retained Students Ultimately Benefit or Harm Their Academic Achievement?

The central focus of this report is estimating whether retention had a positive impact on students' achievement growth. Evaluating the effect of retention means that we have to find a comparison group of low-achieving students who might represent what would have happened if those students who failed to make the promotional cutoff had not been retained. In this report, we do this by comparing the achievement growth of third and sixth graders whose reading test scores fell just below the promotional cutoff in 1998 and 1999, the majority of whom were retained, to the achievement growth of two comparison groups: (1) students who had reading test scores just above the test-score cutoff in those years, the majority of whom were promoted; and (2) third graders in 2000 who had similar test scores just below the test-score cutoff but who were promoted because of changes in the administration of the policy. In general, students who have test scores within a narrow range around the test-score cutoff should be more similar in terms of their underlying achievement than students with either very low scores or achievement closer to grade level.

We present three alternative methods of estimating the effects of retention on students' achievement growth both one and two years after retention:

Comparison 1 compares the achievement growth of students who just missed meeting the promotional cutoff at the end of the summer of 1998 and 1999 (the majority of whom were retained) to comparison groups of low-achieving students who just passed the promotional cutoff at the end of the summer (the majority of whom were promoted). We compare the achievement growth of students in the first year after the retention or promotion decision and then two years after. For third graders, we also do a comparison with third graders in 2000 who had similar test scores just below the test-score cutoff, but who were promoted because of changes in the administration of the policy.

Comparison 2 estimates the pre- to post-gate achievement growth of students in our sample by their status two years after retention: whether they (1) were promoted at the end of the summer, (2) experienced a full year of retention and remained one grade below their age-appropriate counterparts two years later, (3) rejoined their age-appropriate classmates after initially being retained, (4) were placed in special education after failing to meet the cutoff within two years of the retention decision, or (5) experienced a second retention and were two grades behind the age-appropriate classmates two years later.

Comparison 3 utilizes a statistical model to address selection effects that occurred because some of the students who failed at the end of the summer were promoted anyway. The model first estimates the probability of retention at the end of the summer and then estimates the effect of retention on students' learning growth based on the predicted probability of retention.

- *In the third grade, we find little evidence that students who were retained did better than their low-achieving counterparts who were promoted.*

Third graders in 1998 and 1999 whose reading test scores placed them just below the promotional cutoff (Comparison 1) and those who experienced a full year of retention (Comparison 2) had a small boost in performance the year after the retention decision with no substantial positive effects two years after. Our more sophisticated estimates differ slightly, showing no small positive effect one year after the post-gate grade when we use a predicted probability of retention (Comparison 3). In all three cases, we reach similar conclusions, that retention did not proffer any sustained academic benefits to third graders who were retained nor did it have any substantial negative effects.

- ***In the sixth grade, the question is how much did retention hurt?*** In all three comparisons, we find that sixth graders who were retained or were in groups that were predominantly retained had lower achievement growth than their low-achieving counterparts who were promoted, with that effect remaining two years after the initial promotion or retention decision. Estimates from a statistical model that controls for selection effects (Comparison 3) suggests that achievement growth of retained sixth graders was 6 percent lower than that of their low-achieving counterparts who were promoted.

- ***Retained students who were placed in special education after retention were struggling during their retained year and continued to struggle.*** Students who were placed in special education after retention by virtue of that status were having difficulty in the year after retention and continued to struggle two years later. These effects were most pronounced among sixth graders. Consistent with other research, there is little evidence that these students benefited academically from being placed in special education.

- ***There is no evidence that mid-year promotions either helped or harmed students' tested achievement in basic skills.*** In 1998 and 1999, the district allowed many students who were retained to rejoin their age-appropriate classmates. This largely untested policy was controversial. Third graders who were initially retained and then rejoined their classmates had slightly higher learning gains between second and fifth grade than students in our sample who were promoted at the end of the summer and thus attended fourth grade. There were no differences in the achievement growth of sixth graders who were promoted to seventh grade and retained sixth graders who then skipped seventh grade. While our results should be interpreted with caution, there is no evidence that these students' tested achievement in basic skills was harmed by their short-term experience of retention and their resultant lack of instructional time in the subsequent grade. There is also no substantial evidence of any benefit.

Placing Results in Context: The effects of retention or social promotion in closing the achievement gap of low-achieving students

As a final look at the effects of retention, we examine the achievement gap (the difference from the average test scores of their cohort from first grade to two years after retention or promotion) for (1) students who had very low test scores at the end of Summer Bridge (very low-achieving), (2) students who scored just below the test-score cutoff (up to .3 grade equivalents below) in reading, and (3) students who scored just above the test-score cutoff (up to .3 grade equivalents

above) in the gate grades. The majority of students in the first two groups were retained while the majority of students in our third group were promoted.

- ***Low-achieving students in all our groups started school substantially behind their classmates and had already fallen further behind before the promotional gate grade.*** Neither promotion nor retention led to a significant closing of this achievement gap after the gate grade for those low-achieving students who had test scores close to the cutoff.

- ***Students who were the lowest achievers in the school system (the majority of whom were retained) experienced a deterioration in their relative performance after retention.*** Two years after retention, the average achievement of sixth graders in our very low-achieving group was nearly one and one-quarter standard deviations below the average reading achievement of their cohort. Without a comparison group, we cannot assess how much of their post-gate performance represents a negative effect of retention. But there is no evidence that retention slowed the significant deterioration in their learning growth.

Summary Points

In the era of No Child Left Behind, virtually every major school system in the United States is struggling with the question of how to motivate students to achieve and address the needs of students who persistently struggle. Chicago's effort to end social promotion was intended to address persistently low performance in two ways. First, the initiative was aimed at decreasing the number of very low-achieving students prior to the retention decision by using a combination of incentives for students to work harder and for teachers and parents to direct attention to students who would be at risk as well as providing focused intervention during Summer Bridge. But many students, over one-quarter of third graders and approximately 15 percent of sixth and eighth graders, did not raise their test scores to the promotional gate requirements. For these students, Chicago's approach was retention—a second dose of the material with which they had struggled and continued incentives to work hard through additional chances to pass the promotional requirements.

In this report, we focused on the question: Did retaining these low-achieving students help? The answer to this question is definitely no. In the third grade, there is no evidence that retention led to greater achievement growth two years after the promotional gate, and in the sixth grade, we find significant evidence that retention was associated with lower achievement growth. There is also evidence that retaining students under CPS's promotional policy significantly increased the likelihood of placement in special education.

If a second year in the same grade or placement in special education is not an effective strategy for remediating very poor reading performance, then what is an alternative approach? Surely, social promotion alone is also ineffective. There is no evidence that low-achieving students did significantly better when promoted to the next grade. We hope that the evidence presented in this report will spur debate both in Chicago and nationally over alternatives to social promotion and retention as well as to more broadly identifying and managing the needs of low-achieving students who are persistently struggling.

One approach that is supported by the evidence presented in this report is to focus on earlier identification of learning problems. The average low-achieving student in our sample started substantially behind the average CPS student in first grade and

the achievement gap for these groups widened most significantly between first and third grade, before the CPS's promotional policy took effect. Waiting until third or sixth grade to identify these students and intervene does not seem to be a judicious use of resources. This does not mean that high-stakes testing should occur in the first grade. It does mean, however, that school systems must invest in developing effective early assessment, instruction, and intervention approaches that identify students who are not moving forward and provide appropriate supports.

The Ending Social Promotion Study

This report is one in a series that began in 1999 with the publication of *Ending Social Promotion: Results from the first two years*. An update to this report was published in 2000. The first two reports tracked important indicators of student progress, including the proportion of students who met the test-score cutoff at the end of the school year, at the end of the summer, and, for those who were retained, after a second time through the grade. These reports found that after the institution of the policy, the proportion of students who had very low test scores (test scores below the promotional cutoff) declined significantly. Declines in the proportion of students with very low test scores were much greater in the sixth and eighth grades than in the third grade. Early data suggested that the Summer Bridge program and the second chance it afforded students to pass the test and avoid retention had been one of the most successful aspects of the policy, accounting for a large proportion of the improvements in passing rates. At the same time, early data suggested that retained students were struggling their second time through the promotional gate grade and that their post-retention achievement growth was no different than previously socially promoted youths.

Subsequent reports and research articles have looked more closely at these initial findings. In two articles, "The Grasshopper and the Ant: Motivational responses of low achieving students to high stakes testing" and "High Stakes Testing in Chicago: Effects on achievement in promotional gate grades," we looked both qualitatively and quantitatively at the effects of the policy in shaping students' work effort, experiences in school, and achievement.[1] *Ending Social Promotion: Results from Summer Bridge* presented the results of a multiyear evaluation of Summer Bridge. This study examined the short- and long-term effects of Summer Bridge on student achievement and looked carefully at the nature and quality of instruction and students' reports of their experiences in the program. An additional report, *Ending Social Promotion in Chicago: The response of teachers and students*, used Consortium surveys to examine teachers' assessments of the impact of the policy. This report used surveys from 1994 to 2001 to investigate whether teachers spent more time on test preparation after the policy was implemented and aligned the content of their courses to the ITBS. It also examined trends in students' reports of the academic support they received from teachers and parents, their perceptions of the challenge of their coursework, and their participation in after school programs.

In this final report, *Ending Social Promotion: Effects of retention,* and its companion report, *Ending Social Promotion: Dropout rates in Chicago after implementation of the eighth-grade promotion gate*, we focus on evaluating the effects of retention on student achievement and progress in school and ultimately on dropout rates. More technical presentations of these results can be found in two accompanying research papers by Elaine Allensworth and Melissa Roderick and Jenny Nagaoka.[2] Pre-publication versions of these papers can be found on the Consortium website (www.consortium-chicago.org). In addition, a paper by Susan Stone and Mimi Engel examined qualitatively the experience of a group of students who were retained in Chicago, how teachers approached the retention year, and the degree to which retained students experienced different educational supports.[3]

---

[1] Roderick and Engel (2001); Roderick, Jacob, and Bryk (2003).

[2] Allensworth (2004); Roderick and Nagaoka (2004).

[3] Stone and Engel (2004).

INTRODUCTION

There is perhaps no more controversial policy in education today than the decision to retain students on the basis of their performance on standardized tests. Proponents of "ending social promotion" argue that promoting students who have not mastered basic skills sets these students up for failure later on and sends the message that achievement does not matter. Opponents argue that avoiding failure by failing students is a misguided approach and point to research evidence that retaining students does not help their achievement, negatively affects self-esteem, and increases their risk of dropping out. Since 1996, the Chicago Public Schools (CPS) has been at the center of this national debate. In that year, CPS declared an "end to social promotion" and instituted promotional requirements based on students' scores on the Iowa Tests of Basic Skills (ITBS) in the third, sixth, and eighth grades. As a result of this policy, CPS has retained between 7,000 and 10,000 students per year in these three grades—nearly one in five third graders and one in ten sixth and eighth graders subject to the policy.

Over the past five years, Consortium researchers have been evaluating the effects of the CPS initiative. In previous reports, we examined trends in overall student achievement under the policy, the effects of CPS's second-chance summer program, Summer Bridge, teachers' assessments of the impact of the policy, and changes in instruction and students' reports of their experiences in school (see *The Ending Social Promotion Study* on page 6). In this final report, we focus on those students who did not

meet CPS's promotional cutoff, the majority of whom were retained. We look at how the administration of the policy shaped the retention experience and evaluate the impact of retention on students' achievement growth and their progress in school. Specifically, this report focuses on three central questions:

1. Is there evidence that the extra year of instruction allowed low-achieving retained students to catch up and meet the same promotional standards they failed the previous year?

2. How did retention under high-stakes testing and Chicago's use of mid-year promotions shape retained students' subsequent progress? In other words, what was the probability of students being placed in special education, being retained again, or of rejoining their age-appropriate classmates?

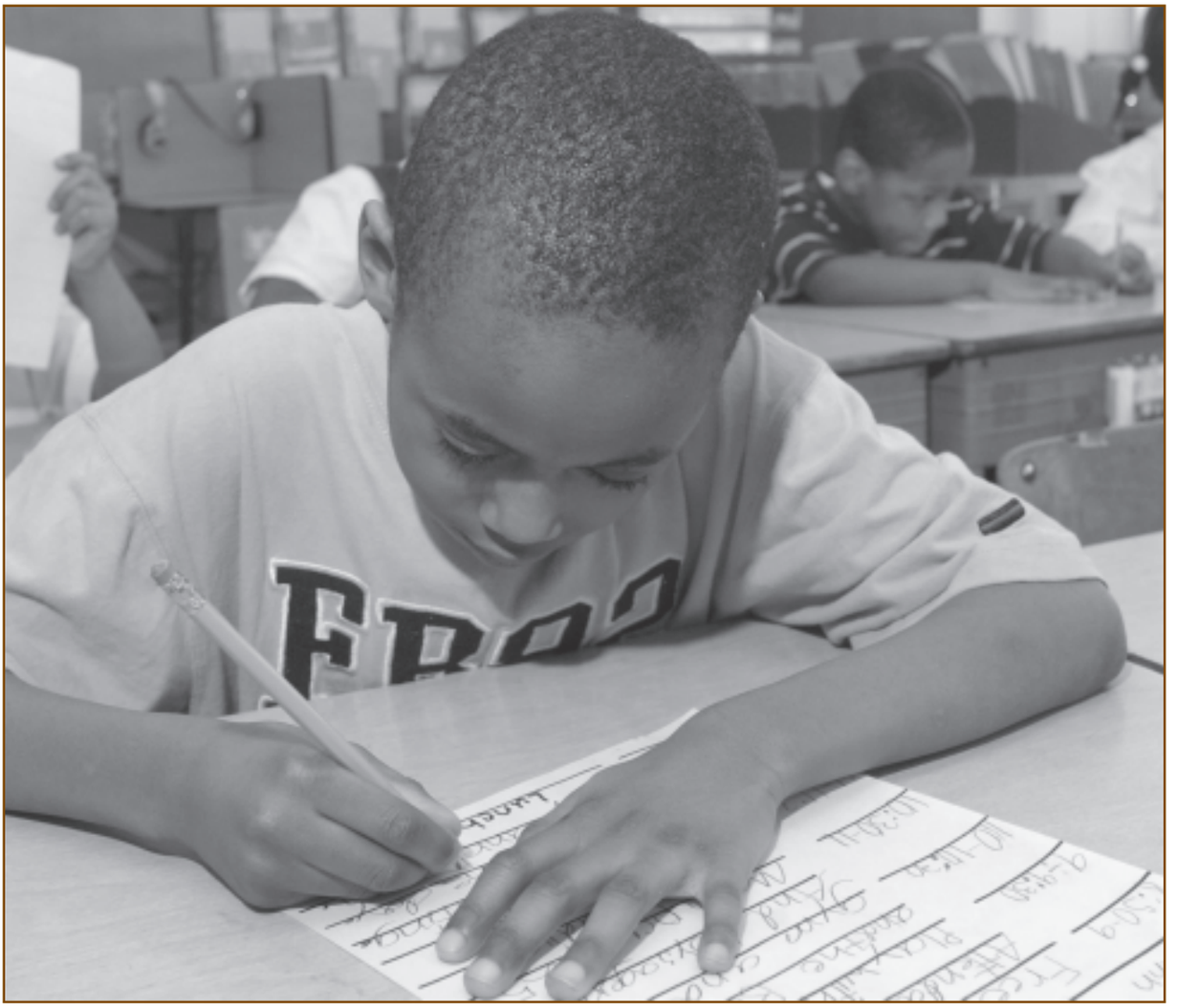
3. Is there evidence that retention led to higher achievement for these students than if they had been promoted to the next grade?

Chicago's ending social promotion policy used a combination of strong incentives and resources to promote student achievement prior to the promotional decision. The central theory of action of such policies is that by setting standards for performance and using the threat of retention, students would work harder, their parents would pay more attention to their education, their teachers would focus on students at risk of retention, and students would be required to demonstrate a mastery of basic skills before progressing to the next grade. The administration provided substantial resources to schools through an after school program and a summer program, Summer Bridge, that gave students at risk of retention extra support and a second chance to pass the test. Presumably, then, much of the effects of high-stakes testing would take place before the retention decision, when students are working harder and getting more support in raising their skills to meet the promotional requirements. Our previous reports and research articles have focused on evaluating evidence for these pre-retention effects in the three gate grades.

While Chicago directed substantial resources to increase the number of students meeting the promotional requirements, the administration provided little structure or additional resources for schools in dealing with the learning needs of students who did not meet the requirements and were retained. What they did do was add a third opportunity to pass the test in January of the retained year. The basic approach, then, was to provide retained students a second time through the same material and an additional chance to pass the test. In this way, the Chicago policy differs from many subsequent district and state efforts to end social promotion where school districts adopted alternatives to retention or much more prescribed approaches to retention. The reader needs to be clear that this report does not evaluate the effect of Chicago's approach versus approaches that offer special services as alternatives to retention or substantial intervention during the retained year. The focus of this report is to examine whether retention under the conditions implemented in Chicago shaped students' experiences in school and subsequent achievement. This report also does not seek to evaluate the overall merits of test-based promotion versus social promotion on all CPS students—such an evaluation would require considering the merits associated with high-stakes testing for those promoted students who might in the absence of the policy, have had lower test scores against the effects of the policy on those who were retained.

Chicago's initiative has been one of the most sustained efforts to date to set promotional requirements in elementary schools and to consistently raise the bar to demand even higher student performance. While the basic thrust of the policy has remained the same, over time there have been significant changes in its administration and in the details of the promotional requirements. This report focuses primarily on evaluating the experience of students who faced the pro-

motional cutoffs in the 1997-1998 to 1999-2000 school years. We describe changes in the administration of the policy over time and present descriptive data on passing and retention rates following the 2000-2001 school year, when CPS altered the policy significantly to include a range around the test-score cutoffs and other measures of performance in making retention decisions. In addition, this report focuses only on evaluating the progress of retained students in the third and sixth grades. Eighth graders are the focus of a companion to this report, (*Ending Social Promotion: Dropout rates in Chicago after implementation of the eighth-grade promotion gate*) by Elaine Allensworth, that evaluates the effect of retention on the likelihood that students retained in eighth grade would later drop out of school.



John Booz

CHAPTER 1

The Policy Debate over High-Stakes Testing and Retention

Chicago's initiative to end social promotion is not the first time a school district has turned to high-stakes testing as a means of raising student achievement. Over the past several decades, large urban school systems have been caught in a cycle of a new administration "ending social promotion" followed by the next administration declaring an end to the deleterious practice of retention.[1] This pendulum swing captures a central conundrum facing administrators of how to motivate teachers and students and to set high expectations while dealing with the problem of persistent poor student performance and the responsibilities of large urban school systems towards their most vulnerable students.

The central tension, however, is that high-stakes testing policies are premised on the idea that it is the threat of retention as much as retention itself that will lead to higher performance. Proponents argue that by setting standards, being clear that achievement matters, and enforcing negative consequences, students will work harder and teachers and parents will pay attention to the needs of the lowest-performing students, leading to higher achievement.[2] Thus, such policies assume that retention and the academic failure it signifies is a negative experience that students, their teachers, and their parents will work to avoid. However, for such policies to work, not only must students' achievement improve because they are

trying to avoid retention, but those who are retained must receive strong enough achievement gains to mitigate the negative impact of the initial failure. Many policymakers and educators believe strongly that an extra year of instruction could give low-achieving students the extra time they need to raise their skills and that the extra time will lay the foundation for more positive achievement later on.[3] In one longitudinal study, Karl Alexander and his colleagues concluded that investing in an extra year of education for low-achieving students may provide an effective catch-up approach.[4]

Opponents argue that using high-stakes testing as a motivational tool ultimately makes sacrificial lambs of the most vulnerable students, wrongly holds students accountable for being in schools and classrooms where they are disadvantaged by poor instruction, and ignores substantial research evidence that grade retention is not an effective strategy for addressing the needs of students with significant learning deficits. Few studies have been able to examine the long-term effects of retention on student achievement and engagement in school and these studies have generally found that, even when there are short-term benefits, they are not sustained over time.[5] There is also strong evidence that students who are overage for grade, particularly retained students, are more at risk of dropping out.[6] An evaluation of a prior New York initiative concluded that even when retained students were given substantial extra support during the retained year, they did not perform better than a comparison group of low-achieving students and were more likely to drop out.[7]

This policy debate highlights the importance of taking a balanced and rigorous approach to examining the effect of retention under Chicago's policy in shaping students' chances of meeting the minimum standards for promotion, their subsequent progress in school, and ultimately, their achievement growth. Before examining these questions, we first outline the details of the Chicago policy and passing and retention rates during the 1996-2002 school years.

CHAPTER 2

Chicago Public Schools' Initiative to End Social Promotion

The centerpiece of Chicago's high-stakes testing program for students is a set of minimum test-score standards in reading and mathematics on the Iowa Tests of Basic Skills (ITBS) for third, sixth, and eighth graders. The promotional test-score cutoffs were set using the grade-equivalent (GE) metric. A student is considered on grade level at national norms if, when taking the test in the eighth month of the school year, he or she obtains a score of that grade plus eight months (e.g., 3.8 for the third grade). From 1996 to 2000, CPS's promotional test-score cutoff for third graders was set at 2.8 GEs, one year below grade level. The sixth-grade cutoff was set at 5.3 GEs, 1.5 years below grade level; and the eighth-grade cutoff was initially set at 7.0 GEs, 1.8 years below grade level.[1] (See *Changes over Time in the Administration of the CPS Promotion Policy* on page 14). These test-score cutoffs roughly correspond to scoring at the 20th percentile on national norms. Two groups of students– special education students and students in a bilingual education program for three years or less—were "exempted" from the promotional criteria, meaning that their test scores did not count in making promotional decisions.[2]

Students who did not meet the test-score cutoffs at the end of the school year were required to participate in a special summer school program called Summer Bridge and retake the test in August. Those who failed again were supposed to be retained in their grade, or if they were 15 years or older, be sent to an alternative program for overage eighth graders.[3] Students who

Changes over Time in the Administration of the CPS Promotion Policy

Chicago Public Schools' (CPS) effort to end social promotion began in the 1995-1996 school year for eighth graders and in the next year for third and sixth graders. From the 1996-1997 to 1999-2000 school years, the CPS promotional standards were set using strict cutoffs for promotion and made no initial provisions for exceptions. But in the first year, 1997, about one-third of students who did not meet the cutoff at the end of the summer were promoted, receiving what we term "waivers." Some of these waivers reflected initial problems in implementation and may represent students and/or schools simply circumventing the policy.[1] In 1998 and 1999, the district tightened significantly its monitoring of promotional decisions and waiver rates fell. Retention decisions were handled by the Office of Schools and Regions. At the time, CPS organized itself into six geographic regions, each of which was led by a region officer. Region officers handled principals' requests for waivers for students who had extenuating circumstances or who had test scores very close to the cutoffs. Our analysis across these two years found that there was significant variation in the number of waivers granted by the region offices, suggesting that region officers differed greatly in their willingness to grant waivers.

Initially, the Chicago policy gave little guidance as to what would happen if retained students did not meet the cutoff for promotion during the next school year or what would happen during the retained year. In the spring of 1998, when the first group of retained students reached the promotional gate again, the district was faced with the decision of whether to double retain students, and did so. This practice, however, was discontinued in the spring of 1999 when the administration moved to waiving retained students if they did not meet the promotional cutoff the second time through the grade. In the second year of the full policy, 1997-1998, the district gave retained students another chance to reach the promotional cutoff in January. This policy was discontinued in the 2000-2001 school year.

From the beginning, CPS leadership received significant criticism for its strict reliance on the test-score cutoff. The policy was challenged in a civil rights complaint, and in 2000, the district began to discuss broadening the requirements for promotion which were adopted for the 2000-2001 school year. In the summer of 1999-2000, before the policy was formally adopted, the district signaled its intention to move to a more inclusive standard and promoted third graders with ITBS scores above 2.2 grade equivalents.

In the 2000-2001 school year, the district officially adopted an amended promotion policy for all three promotional gate grades that specified the use of a wider range around the cutoffs and the use of multiple measures to make promotional decisions at the end of the summer. These multiple measures included attendance, conduct, homework completion, and grades. At the same time, the district raised the test-score cutoffs that would automatically qualify a student for promotion to 3.0 for third graders, 6.0 for sixth graders and 8.0 for eighth graders. In the 2001-2002 school year, CPS discontinued the use of the grade-equivalent metric, instead adopting percentile ranks as benchmarks for the retention policy. The cutoff score for all three grades was the 35th percentile and the band for consideration of other criteria for promotion was between the 24th percentile and the 34th percentile. Because of these changes in the policy and the testing system, the changes in passing and retention rates in 2000-2001 and 2001-2002 are likely be driven by changes in the policy rather than by changes in the overall performance of students.

---

[1] These waivers also reflected last minute decisions about how to handle particular subgroups of students and special cases brought by principals and region heads. Analysis found that Latino students and students in school with high proportions of Latinos were much more likely in this first year to receive waivers, suggesting initial uncertainty about how to implement the policy among schools with high Latino populations, and thus high numbers of bilingual students. Roderick et al. (1999).

were retained were still held to the policy during their retained year, and if they did not raise their test scores to the cutoffs by the following May, they were required to attend a second Summer Bridge. In the third and fourth years of the policy (1998-1999 and 1999-2000), retained students were provided a third chance to meet the promotional test-score cutoffs in January (mid-year). These "mid-year promotes," along with retained students who passed the promotional requirement at the end of the school year with scores well above the cutoff, were allowed to rejoin their age-appropriate classmates after an intensive double dose of summer school in a program called "Making the Grade."[4]

In the Summer Bridge program, CPS adopted a highly structured and prescribed approach to addressing the needs of students who did not initially meet the promotional test-score cutoffs, including providing significant reduction in class size and a mandatory curriculum.[5] Aside from additional promotional opportunities, however, the district gave little structure to the retained year. Decisions about how to group retained students for instruction, whether retained students would receive the same teacher, or whether they would be given special supports were left to each principal's discretion. Retained students were required to participate, along with other students at risk of retention, in an after school program, Lighthouse. And, initially, the district directed some class-size-reduction monies to provide extra support to schools with very high retention rates, but there was little prescription as to how these extra resources should be used.[6] As documented by Susan Stone and Mimi Engel, most retained students received few extra supports.[7] Thus, in Chicago, the educational experience of retention amounted to going through the policy a second time with the same curricular and instructional experience.

John Booz

CHAPTER 3

Passing and Retention Rates in the Third, Sixth, and Eighth Grades, 1997–2002

The test-score standards initially set by CPS were relatively low—a year below grade level in the third grade, 1.5 years below grade level in the sixth grade, and 1.8 years below grade level in the eighth grade. Yet, in the first year of the policy, only one-half of third graders, two-thirds of sixth graders, and about three-fourths of eighth graders met the promotional standards at the end of the school year (see Table 3-1). Summer Bridge and the second chance it provided were important in raising the proportion of students who met minimum test-score cutoffs. By the end of summer 1997, 69 percent of third graders, 80 percent of sixth graders, and 84 percent of eighth graders had reached the cutoff for promotion. The proportion of students who met the criteria for promotion during the school year increased significantly in the third and sixth grades in both 1998 and 1999. Passing rates remained steady among eighth graders, despite the fact that eighth graders faced a steadily increasing cutoff (7.2 and 7.4, versus 7.0) each year. The proportion of eighth graders meeting the promotional cutoff declined significantly, however, in May 2000, when the standard for eighth graders was raised to 7.7.

Chicago's administration of the policy evolved over time, causing significant shifts in the number of students retained each year. (See *Changes over Time in the Administration of the CPS Promotion Policy* on page 14.) The most important change occurred in the 2000-2001 school year when CPS adopted a revised promotional policy that included the use of a wider range around the cutoff and the inclusion of other measures of performance. In the previous year, May 2000, the district signaled its intention to move to a more inclusive standard by increasing waiver rates in all three grades and, in particular, promoting third graders with ITBS scores above

Table 3-1: About 20 Percent of Third Graders and 11 Percent of Sixth and Eighth Graders Were Retained Between 1997 and 1999

Passing and retention rates of first-time third, sixth, and eighth graders included under the policy, 1997-2000[1]

	Third Grade				Sixth Grade				Eighth Grade			
	1996-97	1997-98	1998-99	1999-00	1996-97	1997-98	1998-99	1999-00	1996-97	1997-98	1998-99	1999-00
Cutoff score	2.8	2.8	2.8	2.8	5.3	5.3	5.3	5.5	7.0	7.2	7.5	7.7
Total tested	23,483	23,299	24,277	24,680	24,833	24,196	24,208	22,973	22,229	22,890	21,804	22,719
Percent passing in spring	52%	61%	64%	62%	65%	72%	74%	72%	73%	70%	73%	63%
Percent passing by summer	69%	73%	76%	74%	80%	84%	85%	85%	84%	83%	86%	74%
Percent promoted [1]	75%	74%	75%	84%	83%	84%	83%	87%	77%	76%	79%	78%
Total retained fall [2]	4,644	4,878	4,522	2,657	3,047	2,789	2,768	1,778	2,217	2,531	1,791	2,003
— eighth grade									1,389	1,566	1,000	1,022
— APC									828	965	791	981
Percent retained	20%	21%	19%	11%	12%	12%	11%	8%	10%	11%	8%	9%

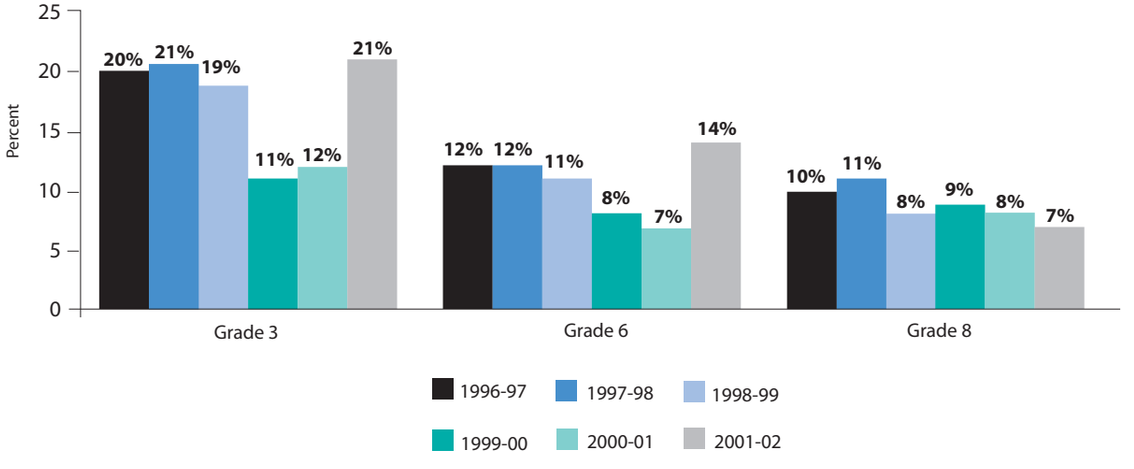
[1] Students are counted as promoted if they are in the next grade and enrolled in CPS the following fall. Some students were promoted even though they did not meet the cutoff scores.

[2] Percent retained of total tested represent the percent of students who did not meet the cutoff who were retained. Some students who did meet the cutoff were retained for other reasons. For example, in third grade in 1997, 137 students who met the cutoff at the end of the summer were retained for other reasons.

Figure 3-1

Retention Rates Declined after 1999, but Rose Again in 2002

Percentage of students retained, 1997 to 2002

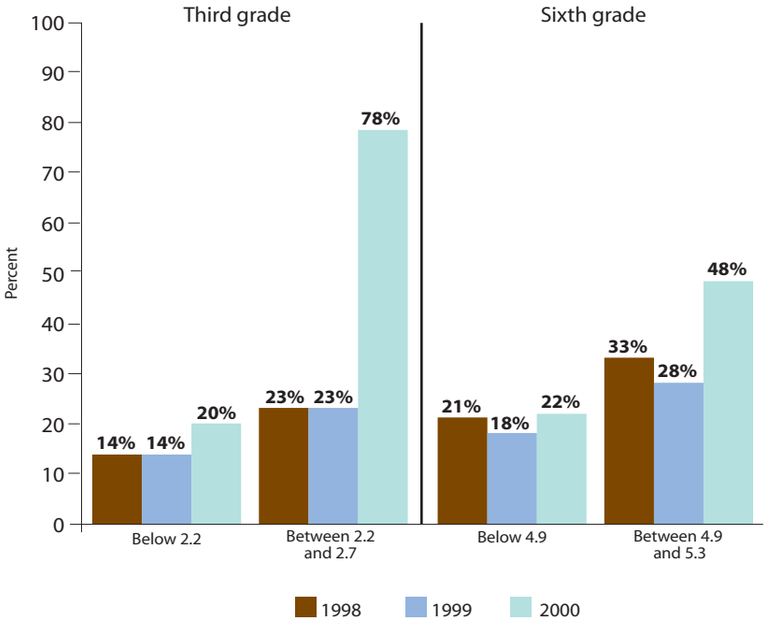


2.2. In that year, the proportion of third graders who were retained dropped from 19 percent in 1999 to 11 percent, as seen in Figure 3-1. Retention rates also decreased in the sixth grade, despite the fact that the sixth-grade cutoff score was increased. Figure 3-2 compares promotion rates for the 1997-1998 and 1999-2000 school years for both third and sixth grades by the students' highest ITBS reading score by the end of Summer Bridge. As seen in Figure 3-2, the rise in third-grade waivers in 1999-2000 was driven largely by the use of a range between 2.2 and 2.7. In the sixth grade, where waiver rates increased much less significantly, the change in the probability of receiving a waiver in 2000 was slight for students with scores close to the old promotional cutoff.

Figure 3-2

In 2000, Most Third Graders With Scores Just Below the Cutoff Score Were Promoted

Percent of students who scored below the cutoff who were promoted

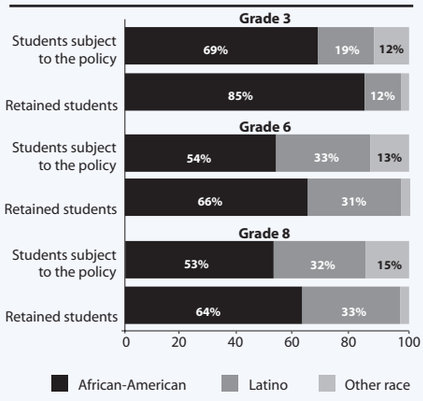


Note: In 1998 and 1999, the cutoff was 2.8 in the third grade and 5.3 in sixth grade. In 2000, the sixth-grade cutoff was raised to 5.5. That year, 76 percent of sixth graders who scored between 5.3 and 5.5 were waived. The scores shown are based on the students' highest ITBS reading score by the end of Summer Bridge.

Who Was Retained: Racial and gender differences in retention under CPS's high-stakes testing

Like many large urban school districts, Chicago serves a predominantly minority student population. In the 1996-1997 school year, 54 percent of the CPS student body was African-American, 32 percent was Latino, and 10 percent was white non-Latino. Many critics of the retention policy worry that if large urban school systems adopt high-stakes testing policies, minority students will be disproportionately affected.

Graph 1
African-American Students Were Disproportionately Retained
On average, demographics of retained students compared to all students subject to the policy, 1997-2002



There are two reasons that we might observe differences in retention rates by race and ethnicity under high-stakes testing. First, minority students might be affected differentially if they have lower test scores. Second, minority students with even the same level of school achievement may be affected differentially if they have fewer resources and supports upon which to draw (i.e., if they attend very low-performing schools or if their teachers and families provide less intervention during the promotional gate grade).

Graph 1 shows the percentage of students included under the policy in each promotional gate grade and the percentage of students retained. In the third grade, only 19 percent of included students were Latino compared to approximately one-third of students in the sixth and eighth grade. This reflects the fact that students who are in bilingual education are not included under the policy until their fourth year in the program.

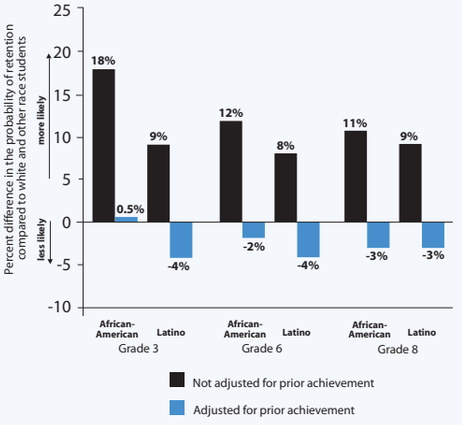
On average, between 1997 and 2002, 69 percent of students subject to the policy in the third grade were African-American, yet 85 percent of retained students were African-American. This means that African-American students were disproportionately retained in the third grade. This discrepancy was smaller in the sixth and eighth grades. On the other hand, Latino students were retained at rates similar to their proportion of students subject to the policy in sixth and eighth grades.

Much of the higher retention rates among African-American students, however, can be explained by the fact that African-American students have lower test scores. Graph 2 shows the difference in the retention rates for African-American students and Latino students versus "white and other" students, adjusted and unadjusted for students' prior reading and math scores. Before adjusting for test scores, it appears that African-American students were much more likely to be retained than other students, particularly in the third grade. But, after adjusting for prior achievement, African-American third graders were not more likely to be retained and were actually less likely to be retained in the sixth and eighth grades. The same is true when adjusting for differences in the prior achievement of Latino students. Thus, African-American students were retained at higher rates largely because these students were the most likely to be at risk under the policy due to low test scores.

In addition to differences by race and ethnicity, previous research has shown boys are much more likely to be retained, particularly in the early grades (Roderick, 1994). But in Chicago, where the retention decision was based on standardized test scores, boys were only slightly more likely to be retained. Thus, gender differences in rates of retention may be different under high-stakes testing than when retention is based on a teacher's decision. Forty-seven percent of all students included in the policy in the third grade were boys, yet boys made up slightly more than half (51 percent) of retained third graders. Boys were still slightly more likely to be retained after controlling for their prior tests scores (see Graph 3).

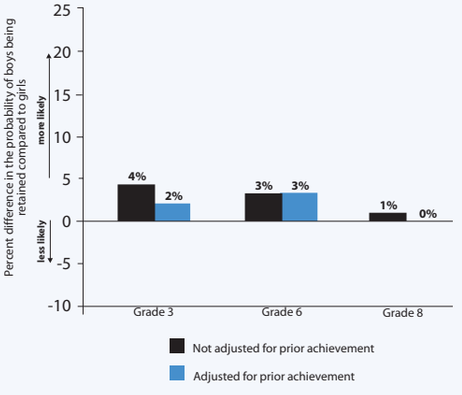
Graph 2

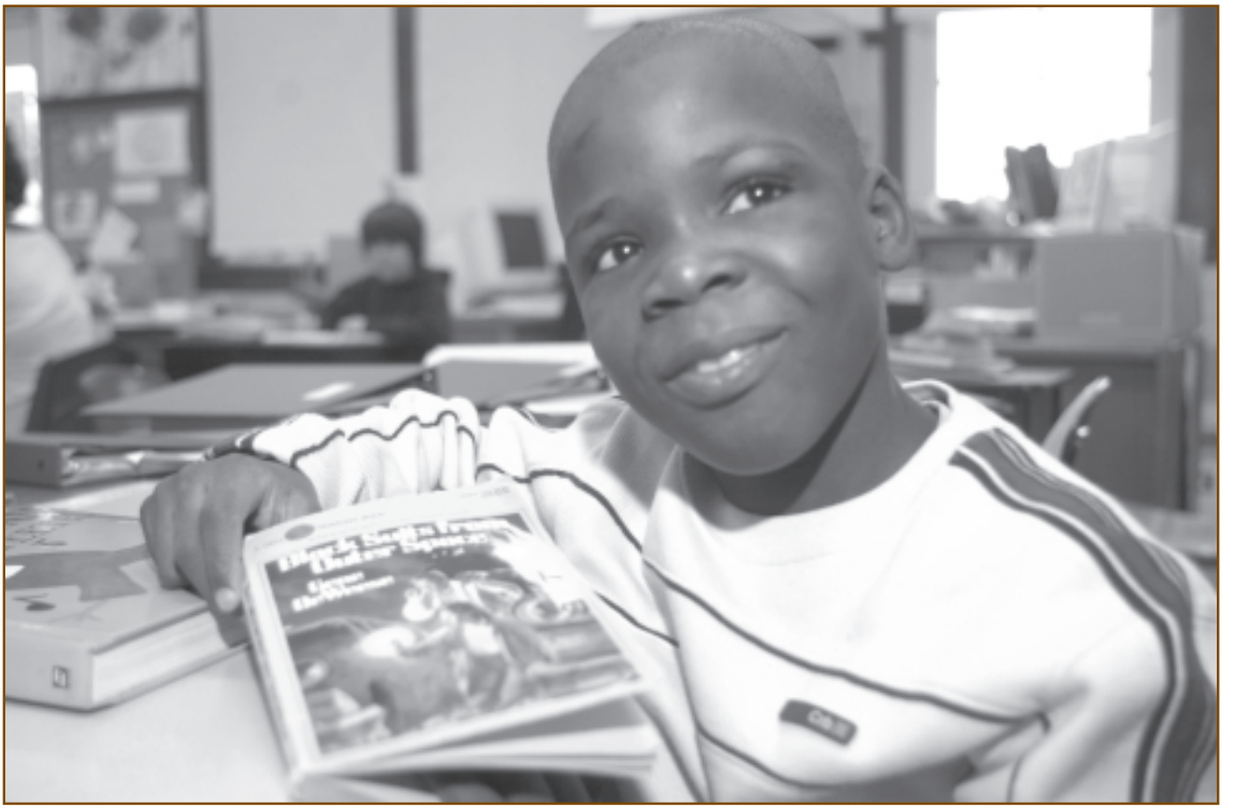
After Adjusting for Prior Reading and Math Test Scores, African-American and Latino Students Were Less Likely to Be Retained than Other Students, 1998-2000



Graph 3

After Adjusting for Prior Reading and Math Test Scores, Boys Were Still Slightly More Likely to be Retained than Girls in Third and Sixth Grades, 1998-2000





John Booz

CHAPTER 4

What Happened to Retained Students?

Main findings:

Retained students struggled their second time through the grade. Even with an extra chance to pass the promotional gate in January, fewer than 60 percent of retained third and sixth graders in 1998 and 1999 were able to raise their test score to the promotional cutoff.

Close to 20 percent of retained third and sixth graders were placed in special education within two years of the initial retention decision. Retained students were placed in special education at three times the rate of low-achieving students prior to the policy, and at significantly higher rates than low-achieving students who were promoted at the end of Summer Bridge.

The January test resulted in many retained students rejoining their age-appropriate classmates. Over one-quarter of third graders and one-third of sixth graders who were retained in 1998 and 1999 rejoined their age-appropriate classmates, having skipped fourth or seventh grade. In 2001, the district discontinued the January test and the rate of double promotions fell.

Tracking the Progress of Third-, Sixth-, and Eighth-Grade Students Retained in 1996-1997 to 2000-2001

The main theory of action behind Chicago's approach to retention was that a second year through the same curriculum combined with multiple chances to pass the standardized test would be effective in helping retained students raise their test scores and be promoted. A first and most basic approach to evaluating the success of retention under high-stakes testing is to ask: How successful were retained students in raising their test scores to the promotional standard during their second time through the grade?

Third and Sixth Graders during Their Retained Year

Table 4-1 tracks the progress of retained third and sixth graders during their second time in the grade. Among the first group of retainees in 1997, only 39 percent of third graders and 42 percent of sixth graders were able to raise their test scores to the promotional cutoffs by the spring of the next school year. Thus, most retained students needed to attend Summer Bridge a second time. By the end of summer 1998, only half of retained third graders and 57 percent of retained sixth graders had raised their test scores to the promotional cutoff.

Students who were retained in 1998 and 1999 had an extra chance to meet the test-score cutoffs in January. This extra chance raised only slightly the proportion of retained students who met the promotional cutoff (see *Passed by August* in Table 4-1). It did, however, change the retention experience because many of these students rejoined their age cohort. For the 1998 and 1999 cohorts, over one-quarter of retained third graders and approximately one-third of retained sixth graders rejoined their age-appropriate classmates within two years of the retention decision.[1]

The fact that only around 60 percent of retained third and sixth graders in the 1998 and 1999 cohorts were able to meet the promotional cutoff, even with an extra chance to pass in January, seems disappointing. In these years, some retained students had taken the same level of the ITBS five times between the end of their first time in the grade and the end of their second time through Summer Bridge.[2] After two years in a grade, and in some cases two summers of Summer Bridge, these students still had not raised their test scores to a year below grade level (2.8 grade equivalents) for third graders and a year and a half below grade level for sixth graders (5.3 or 5.5 grade equiva-

Table 4-1: Retained Students Struggled Their Second Time Through the Promotional Gate
Percent of retained students who passed the promotional cutoff and their status the following fall

	Third Grade				Sixth Grade			
	1997 Cohort	1998 Cohort	1999 Cohort	2000 Cohort	1997 Cohort	1998 Cohort	1999 Cohort	2000 Cohort
Cutoff score	2.8	2.8	2.8	3.0	5.3	5.3	5.5	6.0
Total retained fall	**4,644**	**4,878**	**4,522**	**2,657**	**3,047**	**2,789**	**2,768**	**1,778**
Percent passing promotional cutoff score:								
Passed in January or promoted mid-year	3%	28%	28%	7%	5%	30%	33%	7%
Passed by May the next year	39%	53%	51%	30%	42%	52%	47%	22%
Passed by August the next year	**51%**	**62%**	**59%**	**34%**	**57%**	**63%**	**57%**	**26%**
Status the following fall:								
Promoted total [1]	61%	73%	74%	65%	67%	75%	69%	69%
Retained again	21%	10%	6%	12%	15%	9%	10%	8%
Placed in special education during the retained year	10%	10%	7%	12%	9%	8%	12%	12%

[1] Some students who passed the cutoff score left CPS before being promoted the next fall and so percentages passing and waived do not equal the percentage promoted. Students not falling into promoted, retained, or special education categories left CPS before the fall after their retained year.

lents). Even those who passed seemed to do so only marginally. In 2001, when the cutoff scores were raised (by .2 grade equivalents in third grade and by .5 grade equivalents in the sixth grade), the proportion of retained students who met the promotional cutoff fell from 59 to 34 percent in the third grade and from 57 to 26 percent in the sixth grade.

Part of the reason for these low passing rates was that high proportions of these students were placed in special education during their retained year and were no longer subject to the policy. Approximately 7 to 12 percent of retained third graders and 9 to 12 percent of retained sixth graders were placed in special education during their retained year. In the same years, only 2.5 percent of all third graders and 1.5 percent of all sixth graders were placed in special education. Thus, retained students were placed in special education at nearly five times the rate of the average third grader and nearly seven times the rate of the average sixth grader.

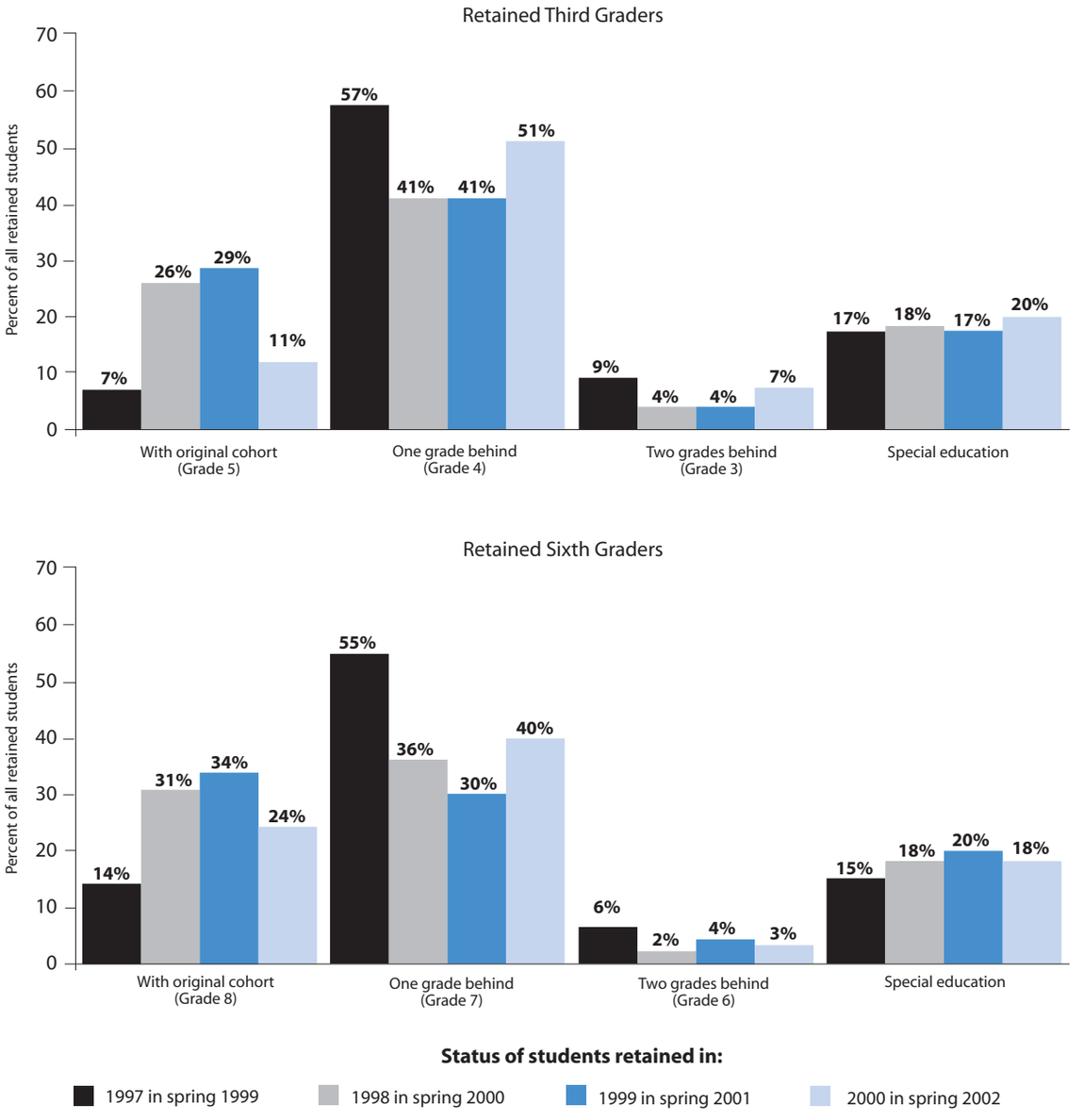
Third and Sixth Graders Two Years after Their Initial Retention

Retained students, particularly those who did not meet the cutoff at the end of their retained year, continued to face high rates of special education placement. Two years after their initial retention, between 17 and 20 percent of retained third and sixth graders had been placed in special education (see Figures 4-1 and 4-2). Most of these additional special education placements were students who were either double retained or waived after not meeting the cutoff. For example, in 1997, the district double retained more than 900 students. About half of these students were then placed in special education during their second retained year. In 1998 and 1999, the district reduced significantly the number of these double retentions and moved toward waiving students who had not met the promotional standard. However, the percentage of retained students who were placed in special education within

John Booz

Figures 4-1 and 4-2

Almost 20 Percent of Retained Students Were in Special Education Two Years Later
Status of retained students two years after their initial retention



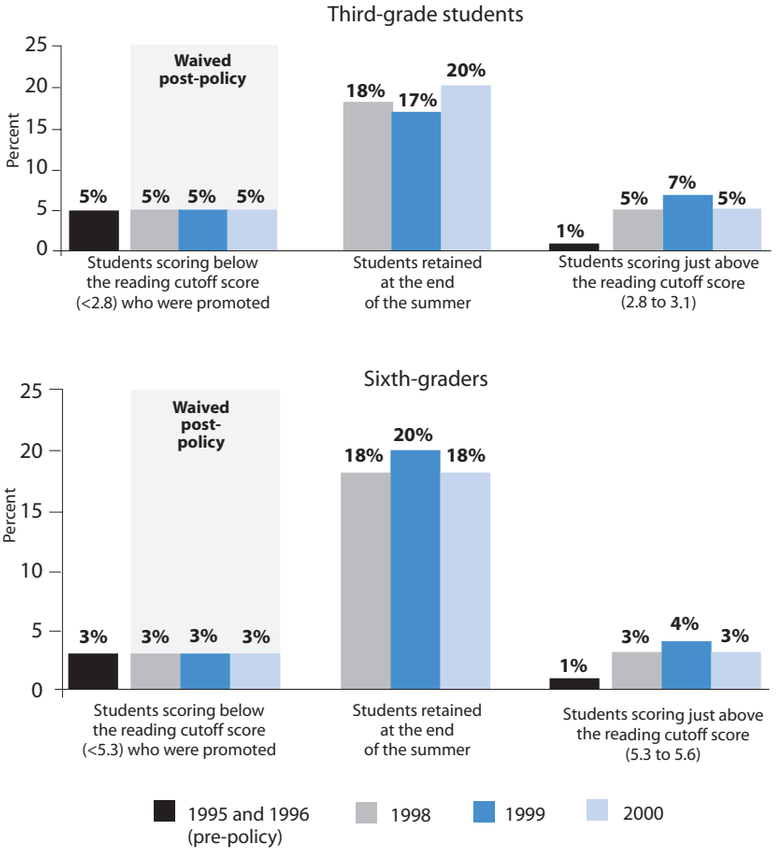
two years of retention increased. On average, 18 percent of third and sixth graders retained in 1998, 1999, and 2000 were placed in special education within two years.

Comparing the rate of special education placement of retained students to the average of their cohort is misleading because low-performing students, in general, face a much higher likelihood of being referred to special education. Figures 4-3 and 4-4 compare the proportion of retained third and sixth graders who were placed in special education to the placement rate of two comparison groups: 1) all students who had test scores below the promotional cutoff both prior to and after the policy, and 2) students who attended Summer Bridge and had test scores in reading just above the promotional cutoff. Students who attended Summer Bridge but had test scores just above the promotional cutoff (within three months above) are a comparison group of low-achieving students who, for the most part, were not retained and were attending Chicago schools under the promotional policy (see next section, Figures 5-1 and 5-2). Thus, this second group allows us to examine whether all low-achieving students were more likely to be placed in special education after the policy was implemented in 1996 versus the extent

Figure 4-3 and 4-4

Retained Third and Sixth Graders Were Placed in Special Education at Much Higher Rates than Other Low-Achieving Students

Percent of students placed in special education within two years after the third-and sixth-grade promotional gate



Note: We do not report the placement rate for students who were retained pre-policy because retention rates were low prior to the policy; in the 1995-1996 school year, only 360 third graders were retained and these students were much more likely to be placed in special education. Approximately 11 percent of third-graders and 10 percent of sixth graders retained prior to the policy were placed in special education two years after retention.

to which this was associated with retention. Prior to the 1997 school year, approximately 5 percent of low-achieving third graders were placed in special education. Among the group of post-policy low-achieving third graders who just passed the cutoff in reading, only about 6 percent were placed in special education within two years of passing the gate, suggesting little change in the overall placement rate of low-achieving students. Thus, within two years of retention, retained students were placed in special education at three times the rate in third grade and six times the rate in sixth grade of other post-policy low-achieving students who had narrowly missed being retained as well as pre-policy low-achieving students.

This initial look at the academic progress of retained students raises questions about the impact of retention. These students continued to struggle their second time through the grade. Clearly, retained students did not experience a "turnaround" in their performance. A large proportion of these students were classified as needing special education services, and many students were promoted despite their failure to meet the test-score cutoff after multiple attempts. It is also clear that the administration of the Chicago policy meant that not all students experienced what is commonly thought of as "retention"—repeating a grade and then progressing on to the next grade. Only 40 percent of third-grade retained students in 1998 and 1999, and only about one-third of sixth-grade retained students in these years experienced a traditional full-year retention. Over one-quarter of third graders and one-third of sixth graders who were initially retained later rejoined their age-appropriate classmates, having skipped fourth or seventh grade. At the same time, two years later, about one-quarter of retained students were either two grades behind or in special education. This analysis raises the question: did retention and the various experiences of retained youth ultimately benefit or harm their academic achievement? To examine this question, we turn to an analysis of achievement effects.

How Should We Interpret Retained Students' High Placement Rate into Special Education?

The fact that so many retained students were placed in special education is an important finding of this report. The question is: How should we interpret this trend? In the interpretative summary, we return to this question and discuss the potential implications of this trend in light of the overall findings of this report. It is worth stating some hypotheses here. First, these special education placements might reflect the fact that after students were retained, teachers identified previously undiagnosed learning disabilities. Second, they might reflect a mislabeling of students as learning disabled because teachers lacked an alternative explanation and strategy for the difficulties students were presenting. And, third, teachers and schools might have referred students to special education out of concern that, without that status and thus exemption from the policy, these students would not be able to progress; in other words, special education could have been used as a means of getting struggling students around the policy and removing them from the accountability system. Most likely the high placement of retained students in special education reflected some combination of the above. Unfortunately, we cannot in this report disentangle the exact causes of this trend. The link between placement in special education and reading difficulties is a national problem and one that every school system needs to address. It is particularly important in Chicago because the school system has seen a dramatic rise in the percentage of students who are placed in special education by the eighth grade, from 15.0 percent in 1996 to 18.8 percent in 2002. Much of this change was driven by a substantial increase in the percentage of late-grade (sixth to eighth grade) special education placement and high rates of placement in special education among students who had been retained.[1] In the next section, we present the post-retention achievement growth of these retained students who were placed in special education.

---

[1] Miller, Shazia Rafiullah and Matthew Gladden. 2002. Changing Special Education Enrollments: Causes and distributions among schools. Chicago: Consortium on Chicago School Research.

CHAPTER 5

Helpful, Harmful, or Harmless?

Estimating the Effects of Retention on Reading Achievement

The central argument for grade retention is that if students have not mastered basic skills, they are served better by repeating a grade and gaining those skills than by moving to the next grade and struggling when faced with more advanced material. What does this imply about their achievement growth? Suppose we had two similar, low-achieving students and decided to retain one student and promote the other. The argument for grade retention is that the retained student would do better because that student would be able to catch up and master the material with which he or she had struggled rather than continue to struggle in the next grade. This argument also suggests that we should continue to see greater achievement growth for the retained students when they move on to the next grade because if they catch up in the retained year, they should do better than students with inadequate skills who were socially promoted to the next grade. Thus, we would expect that the achievement gains of a third-grade retained student in the first and second years after the retention decision would be greater than a similar student who was promoted to fourth grade and then moved on to fifth grade.

The problem in evaluating retention is that students were not randomly retained in Chicago. In order to evaluate the impact of retention versus social promotion, we used a comparison group of low-achieving promoted students to represent what would happen in the absence of retention. In this section, we compare the achievement growth of students whose scores fell just below the promotional cutoff in 1998 and 1999, the majority of whom were retained, to the achievement growth of two comparison groups: students who had reading test scores just above the test-score cutoff in

those years, the majority of whom were promoted; and third graders in 2000 who had similar test scores just below the test-score cutoff, but who were promoted because of changes in the administration of the policy. Before presenting our results, we begin by discussing two central issues that arose in this study and how we resolved them: how we selected retained students and a comparable group with which to estimate the effects of retention on achievement and how we measured their achievement growth.

Why We Did Not Use Pre- and Post-Policy Comparisons

As stated above, in order to estimate the effect of retention, we sought to identify a group of promoted students who looked as much as possible like students who were retained under Chicago's policy. One approach would have been to compare the achievement of low-achieving students prior to the policy's implementation in 1996 to those after 1996. The problem, however, is that test scores in Chicago were rising in the early 1990s and rose rapidly after 1996, particularly in the promotional gate grades. The increase in test scores means there were many more low-achieving students prior to 1996. For example, the proportion of sixth graders with ITBS reading test scores below 5.3 GEs—the promotional cutoff—fell from 37 percent in 1995, the year before the promotional gate took effect, to 14 percent by the end of the summer of 1999.[1] As the number of low-achieving students declines, we expect that the students who continue to have low test scores will most likely represent the most disadvantaged and most persistently low-performing students, and that these students, even in the absence of retention, may have had much lower subsequent performance. Unfortunately, there is no way to identify which students might have remained low-achieving after system-wide test scores rose and after the institution of the promotional requirements.

How about Students in the Third Grade Who Were Promoted When the Policy Changed?

In the previous section, we saw that in summer 2000, CPS decided to promote third graders with test scores above 2.2 GEs rather than the official cutoff of 2.8. This change in the policy provides a unique opportunity to evaluate the effects of retention. In 1998 and 1999, as seen in Figures 5-1, most third graders with test scores just below the cutoff in reading (ITBS reading achievement between 2.4 and 2.7) were retained,

Why Focus on Reading?

In this section, we estimate the effect of retention on reading achievement and define groups in terms of performance in reading. There are two reasons that we focus on reading. First, most students who attended Summer Bridge needed to raise their reading scores but fewer needed to raise their math scores. Over 85 percent of third graders and sixth graders in the 1998-2000 school years attended Summer Bridge because of their reading scores while less than half of students needed to pass mathematics. Thus students' low reading scores accounted for the lion's share of Summer Bridge attendance and the retention decision. Most students in our just-above group attended Summer Bridge for reading and were promoted because they had raised their reading scores at the August testing and had met the cutoff in mathematics. Some students in our just-above group were retained because of their mathematics scores. On the other hand, because so many students had to pass reading, constructing a similar just-above group in mathematics would have resulted in a group with very high retention rates because many students met the cutoff in mathematics but not reading. Thus, a just-above group based on mathematics scores would not have provided a comparison group for retained students. In addition, restricting analysis to only those who needed to pass mathematics would result in extremely low sample sizes and a group that was not representative of most students who were retained and attended Summer Bridge. Specifically, less than 10 percent of students who attended Summer Bridge in the 1998-2000 school years had passed the cutoff in reading but did not meet the cutoff in mathematics.

but in 2000, the majority of students with reading test scores in this range were promoted. We can expect these students to be very similar. Students in each of these three years experienced the same policy, attended Summer Bridge, and performed similarly on the ITBS. The difference in the probability of promotion for these students, then, was solely the year they were born, and thus, the year they attended third grade. But because this policy change occurred only in the third grade, we could not use the same method to evaluate effects of retention in other grades.

How about Students Who Barely Passed the Promotional Cutoff at the End of Summer Bridge?

A second approach is to use students who attended Summer Bridge but were promoted because their test scores were just above the promotional cutoff. Because there is wide variation from test to test in students' performance on the ITBS, a difference in one question, right or wrong, could determine whether the promotional cutoff was met. Thus, students within a narrow range around the test-score cutoff should be similar in terms of their average achievement even though there may be variation in their test scores from test to test.

This approach gives us an opportunity to look at both third- and sixth-grade retentions. It compares students who did not meet the test-score cutoff, the majority of whom were retained in 1998 and 1999, to students whose reading test scores allowed them to barely pass at the end of the summer, the majority of whom were promoted. This means that in this section, we do not look at the achievement growth of students who were retained and had very low test scores (ITBS reading test scores below 2.4 GEs, which is about a year and half below grade level in the third grade; and below 4.9 GEs, which is nearly two years below grade level in the sixth grade). These very low-achieving students might have had very different achievement growth even in the absence of retention. In the next section of the paper, we come back to this group to look at descriptive trends in their achievement prior to and after retention or promotion.

John Booz

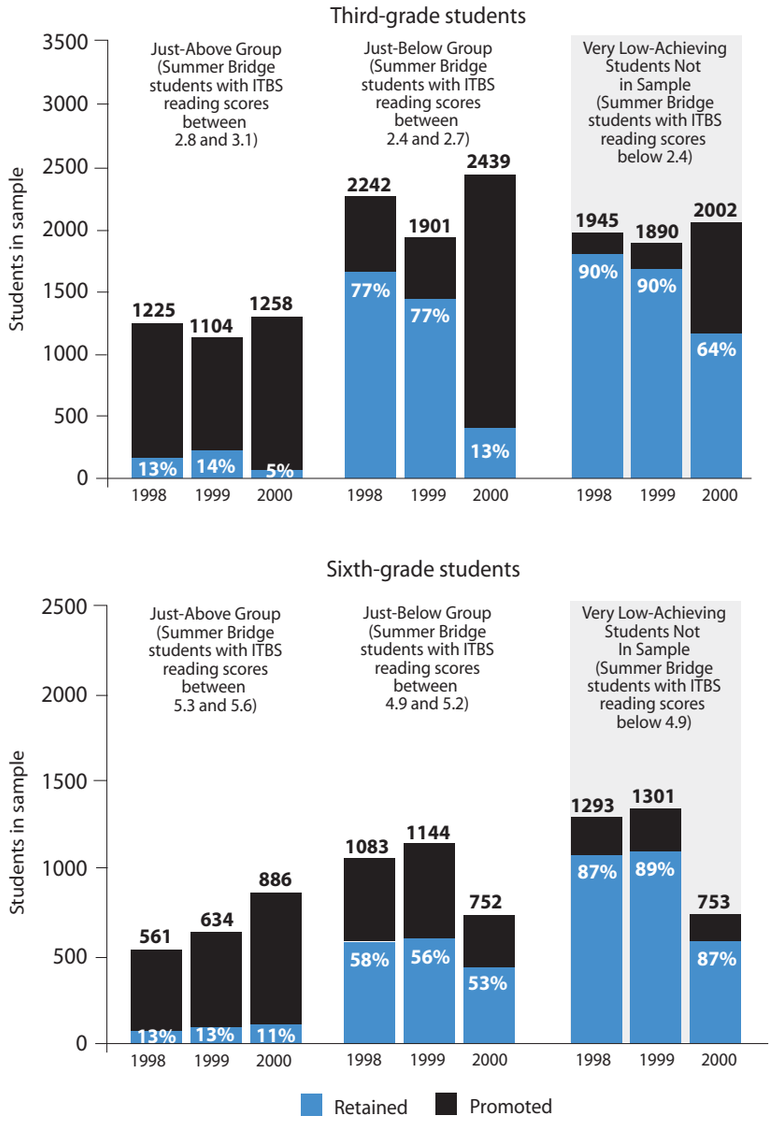
Most students who attended Summer Bridge needed to raise their reading scores but fewer needed to raise their math scores. . . . Thus students' low reading scores accounted for the lion's share of Summer Bridge attendance and the retention decision.

We are not able to do a similar analysis of the effects of eighth-grade retentions. CPS eighth-grade students take the ITBS, while ninth-grade students take the Test of Achievement Proficiency (TAP). Because the two tests are not directly comparable, it is not possible to compare achievement growth for retained and promoted students because retained eighth graders take the ITBS while promoted ninth graders take the TAP.[2]

Figures 5-1 and 5-2 describe our just-below and just-above cutoff samples and the sample of very low-achieving Summer Bridge students who will be excluded from the analysis. We limit our analysis to students who were retained in the 1997-1998 and 1998-1999 school years because waiver rates and the process of waivers were so dramatically different in 1996-1997, the first year of the policy.[3]

Figures 5-1 and 5-2

Using Students Who Scored Just-Above or Just-Below the Test Score Cutoff to Estimate the Effects of Retention



Note: In 1998 and 1999, most third- and sixth-grade students who failed to meet the promotional cutoff were retained. Likewise, most third- and sixth-grade students who passed the promotional cutoff were promoted. In 2000, third graders with test scores between 2.4 and 2.7 were also mostly promoted, providing another comparison group. But because the promotional cutoff for sixth graders was raised to 5.5 in 2000, the number of students in the just-below group who were promoted did not increase.

What Do We Do with Students Who Scored Just Below the Cutoff in Reading but Were Not Retained?

As seen in Figures 5-1 and 5-2, not all students who had reading test scores just below the promotional cutoff in our 1998 and 1999 groups were retained. Twenty-three percent of third graders and many more sixth graders in those years were promoted despite not meeting the reading test-score cutoff. The same is also true for students in the comparison groups. Even though most third graders with test scores between 2.8 GEs and 3.1 GEs were promoted in 1998 and 1999, 14 percent were still retained. This is not a little problem because eliminating these students from the analysis would introduce selection bias. We do not know why some students were promoted when the majority of students with similar test scores were retained, but we suspect that these promoted students may have had more resources or special circumstances that led them to have better achievement potential. The opposite is true for the comparison group of students who had been mostly promoted. Students who were retained in the just-above group had reached the promotional cutoff in reading but either did not pass the cutoff in mathematics or were retained because of poor grades and attendance. Thus, while we might argue that students with reading test scores just-above and just-below the cutoffs are similar, if we remove the best of one group (students with test scores just below the cutoff who were promoted) and remove those most likely to struggle in our comparison group, the two groups would no longer be similar because of the introduction of selection bias.

How do we handle this? In the rest of this section, we use three different comparisons, all of

which provide different ways to approach this problem. In our first comparison, we compare the achievement growth of all students in 1998 and 1999 who scored just below the promotional cutoff in reading, the majority of whom were retained, to all students whose reading test scores allowed them to just pass the promotional cutoff, the majority of whom were promoted. In our second comparison, we simply compare the achievement growth of all students who had test scores just above and just below the cutoff by whether they were promoted at the end of the summer, retained for a full year, passed in January, placed in special education, or retained twice. This second comparison then, has no controls for selection bias. In the third comparison, we explicitly deal with the selection problem at the end of the summer by using a more advanced statistical technique that predicts the probability of retention. By comparing the estimates from these three methods, we can get a better feel for how this selection problem might be affecting our results.

One Year after Retention, Students Who Were Retained Were in Different Grades from Students Who Were Promoted. How Do We Measure Their Achievement Growth Adequately?

The promotional test-score cutoffs were set using the grade-equivalent metric. The grade equivalent is adequate in evaluating student performance against national norms within grade levels. But grade-equivalent scores and growth in grade equivalents are not com-

The Definition of Just-Below and Just-Above Cutoff Groups and the Sample Used in the Analysis

We limit our analysis to students in the 1998, 1999, and 2000 school years who failed to make the promotional cutoff in May, who had to attend Summer Bridge as a result, and whose highest reading test score was between 2.4 and 3.1 for third graders, and between 4.9 and 5.6 for sixth graders. We expect that students with scores within these narrow ranges would have similar underlying achievement.

Our third-grade *just-below cutoff group* consists of students whose highest tests score in reading either at the beginning or end of Summer Bridge was between 2.4 and 2.7. These students failed to meet the promotional cutoff in reading, and in 1998 and 1999, most were retained (see Figure 5-1). In 2000, however, most of these students were promoted.

Our *just-above cutoff* comparison group is defined as third graders whose highest test score in reading in either May or August was between 2.8 and 3.1. These students met the cutoff in reading by the end of the summer and the majority was promoted. Third graders in our just-below cutoff group make up slightly more than half of third graders who failed to make the cutoff in reading at the end of Summer Bridge. In the sixth grade, our *just-below cutoff* group consists of students whose highest test scores were between 4.9 and 5.2 in reading. Our *just-above cutoff* group consists of sixth graders who attended Summer Bridge and who had reading test scores of between 5.3 and 5.6.

The sample consists of students who were in grades three and six for the first time between 1997-1998 and 1999-2000, who attended Summer Bridge, who were included under the policy, and who fell into one of our comparison groups at the end of the summer. Students who are retained and thus are in their second year in that grade are considered part of the retained sample from the prior cohort. We only include students as part of the sample if they had post-year test scores. Thus, students who left the school system at the end of the summer or over the post-test time period were excluded from the sample. Students who were in special education after the gate grade were included if they had test scores.

parable across grades; a grade-equivalent score of 2.8 does not mean the same thing in third grade as it does in fourth grade. This is a particularly acute issue for scores of very low-achieving students. For example, we have found that most very low-achieving students will get a higher test score if they take a higher form of the test (e.g., they are in a higher grade). Since promoted students are in a higher grade, we would find that promoted students do better even if grade retention didn't actually have any negative effects on retained students' achievement. The Consortium on Chicago School Research conducted an extensive equating study that converted ITBS scores to Rasch-scale scores that can be compared across grades and across forms and levels of the test. By using these Rasch-scale scores, then, we can compare test scores and achievement growth of students in the third grade to those in the fourth grade. This means, however, that we present growth in logits (the metric of the Rasch-scale score; see *What is a Rasch-Equated Scale Score and Why is it Important to Measure Achievement in Equated Scale Scores Rather Than in Grade Equivalents*) rather than grade-equivalents. However, logits are somewhat more difficult to interpret.

Finally, throughout this section, we present results as the estimated achievement growth of stu-

What is a Rasch-Equated Scale Score and Why Is It Important to Measure Achievement in Equated Scale Scores Rather Than in Grade Equivalents?

Chicago uses the Iowa Tests of Basic Skills (ITBS), and until 2002, results were reported in the grade-equivalent metric (GE). The GE is adequate in evaluating student performance against national norms within grade levels. The GE metric is not useful in assessing student growth or comparing performance across grades because scores are not directly comparable across test levels. The GE metric was originally intended to make normative scores more easily understood by parents and teachers and was not intended for statistical analysis. Since the GE is not a linear metric, a score of 5.3 GEs at level 12 (grade 6) does not represent the same level of achievement as a score of 5.3 GEs at level 13 (grade 7) especially when different forms are administered year to year. Across-grade-level differences are particularly acute at the bottom of the scale. In general, students at the lower end of the achievement distribution will receive higher scores on the ITBS simply by taking a higher level of the test. Because a low-achieving student taking a fourth-grade ITBS test will probably score higher than if that same student took a third-grade test, estimates of retention will be negative because of the higher test level taken if the student is promoted.

In addition, the GE metric is not linear within test levels because the scale spreads out more at the extremes of the score distribution. One additional correct response at the top or bottom of the scale can translate into a gain of nearly one GE whereas an additional correct answer in the middle of the scale would result in only a fraction of this increase. The lack of comparability of test scores across grades is particularly important when evaluating the effect of retention under Chicago's high-stakes testing program because at two years after the gate year, students who did not meet the cutoff were in several different grades (two grades below their age-appropriate counterparts, one grade below, or on grade level).

To address these problems, the Consortium on Chicago School Research conducted an extensive equating study that converted ITBS test scores to a logit metric using Rasch models that are comparable across time and within and across test levels (Bryk, Thum, et al., 1998). At present, equated Rasch scores are available only through the 2000-2001 school year. Thus, when using our 2000 comparison group, we can examine only the effects of retention one year after the gate grade. In this section, we use equated Rasch-scale scores that have been adjusted for form and level effects. The scale scores, reported in the logit metric, are not in an easily interpretable metric such as the grade equivalent. We can, however, compare the relative size of effects and can express effects as a percentage of the average learning gain.

Using Growth-Curve Modeling to Estimate the Effects of Retention

The simplest method of estimating retention effects is to compare the gate-grade (third or sixth) to post-gate grade achievement growth of students who fell below the cutoff in 1998 and 1999 to our comparison groups. The problem, however, is that whenever students' participation in a program, in this case Summer Bridge, is based on a single test score, some students will be selected for the program because of a bad testing day, and these effects will be most acute for the group with underlying ability close to the cutoff. These effects were most acute for our just-above cutoff group who tended to be selected for Summer Bridge because they had a lower than average test score given their prior trajectory. If we simply compare the achievement growth between the end of third or sixth grade (spring) and their post-gate year, we would overestimate achievement gains among our just-above cutoff group because part of that achievement growth in our just-above group sample would be simply moving from a below average to an average year (as reflected in this summer test score). This is called regression to the mean. A similar pattern was observed in the third grade. Thus, without correcting for the fact that students in our just-above cutoff group can be expected, on average, to follow a worse than expected testing year with more average performance given their prior test-score history, we would most likely find a negative effect of retention.

One approach to addressing regression to the mean is to use available data on students' entire test-score history to estimate each student's growth in achievement both prior to and after the promotional testing year and to compare the post retention or promotion achievement growth of different groups of students. To obtain these estimates, we use a three-level Hierarchical Linear Model (HLM) (see Appendix A).[1] We estimate two models (1) a basic model that compares the achievement growth of students in our just-below group and comparison groups regardless of their retention or promotion status, and(2) an extended model that estimates difference in the average post-gate achievement growth of students by their experience in the post-gate periods (e.g., whether students at the two-year mark had experienced one retention, a double retention, placement in special education, or had rejoined their age-appropriate comparison groups). We further adjusted these estimates for differences in the demographic and school experiences of students that might be associated with post-gate achievement. These characteristics include a student's gender, race and ethnicity, whether the student experienced a prior retention or experienced mobility during the gate grade, and a measure of a student's poverty status. The Appendix and the technical paper that accompanies this report describe these models and calculations in more detail.[2]

---

[1] The HLM framework has several advantages (Bryk and Raudenbush 1992). First, it yields consistent standard errors that take into account the correlation of errors within schools. Second, it provides empirical Bayes residuals that take into account the relative information provided by each individual. Third, for growth curve analysis, the model allows us to include individuals who have missing test score data.

[2] Roderick and Nagaoka (forthcoming).

dents pre-gate (second or fifth grade) to post-gate grade. A central problem that arises when studying the effect of grade retention is that students who are retained are by definition those with lower achievement, and most likely, those who have had a bad academic year. This problem is exacerbated when students are selected for an intervention, such as Summer Bridge or retention, based on a single test score. This was particularly true for our just-above cutoff group. We might expect that students who have a particularly bad school year or bad testing day might follow that year with an average or above-average testing year. This is called "regression to the mean." Our statistical model corrects for these problems. For presentation purposes, we present results as the estimated pre- to post-gate growth in reading achievement.

Main findings:

Comparison 1 compares the estimated learning gains prior to and one and two years after the promotional gate of students who at the end of the Summer Bridge had reading test scores just below the promotional cutoff (the majority of whom were retained) to that of students who scored just above the promotional cutoff in reading (the majority of whom were promoted).

- *In the third grade, there is little evidence that students in our predominantly retained group did better.* Third graders in 1998 and 1999 who just missed meeting the promotional cutoff in reading had slightly higher achievement growth between the pre-gate grade to one year after the promotion or retention decision than students who passed the promotional cutoff with reading test scores within three months above. This was not true two years after the retention or promotion decision.

- *Results for the sixth grade are more negative.* The average estimated growth in reading test scores of sixth graders who just missed meeting the promotional cutoff in reading was more negative both one and two years after the promotion or retention decision than that of sixth graders who passed the promotional cutoff with reading test scores just-above the cutoff score. While negative, the estimated impact was small.

Comparison 1: Third and Sixth Graders in Our Just-Below and Just-Above Cutoff Groups

In this first comparison, we compare the achievement growth of students who failed to meet the test-score cutoff in 1998 and 1999—the majority of whom were retained—to our comparison groups. As discussed above, this is a conservative approach to estimating the effect of retention because we are comparing the achievement growth of all the students who just failed to meet the cutoff (just below) and all those who just met the cutoff (just above) regardless of whether they were retained or promoted at the end of the summer.

Third Graders in the Just-Below and Just-Above Cutoff Groups in 2000 Versus Just-Below Cutoff Groups in 1998 and 1999

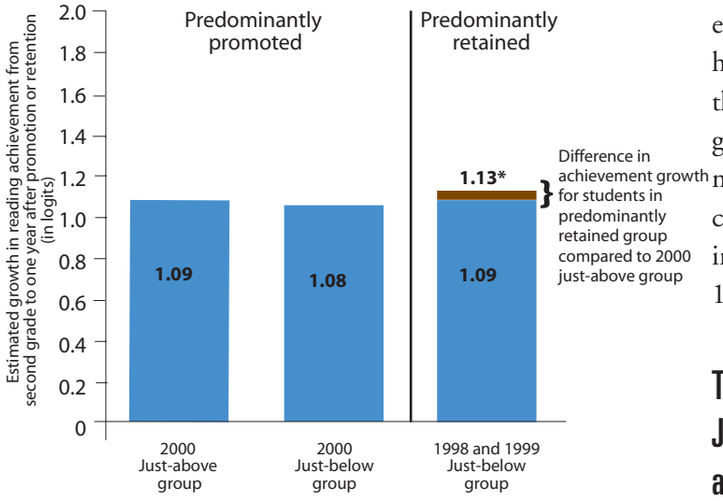
To illustrate our approach, we begin by comparing the achievement growth of third graders in 2000 who scored just below and who scored just above the cutoff. Since most students in these two groups were promoted, we expect their achievement growth would be similar. Figure 5-3 presents their estimated achievement growth between second grade and one year after the initial retention or promotion decision (post-gate 1). These estimates are derived from a statistical model that uses information on a student's entire test-score history as well as those of his or her cohort to estimate each student's growth in achievement both prior to and after the gate-year and to compare the post-gate achievement growth of different groups of students (*see Using Growth-Curve Modeling to Estimate the Effects of Retention* on page 35).

As seen in Figure 5-3, third graders in 2000 who had test scores just above the cutoff gained approximately 1.09 logits between second grade and one year after the initial retention or promotion decision (post-gate 1). Those third graders who just missed the cutoff that year and who were promoted because of the change in the policy also had quite similar achievement growth—approximately 1.08 logits. The similarity in growth of these two groups confirms that in the absence of retention, we might expect students with scores just above and just below the cutoff to be similar in their post-gate achievement.

The majority of third graders in 1998 and 1999 who just missed the promotional gate were retained.

Figure 5-3

Third Graders in the Just-Below Cutoff Group Had Slightly Higher Achievement Gains One Year after the Retention Decision Than the Just-Above Group

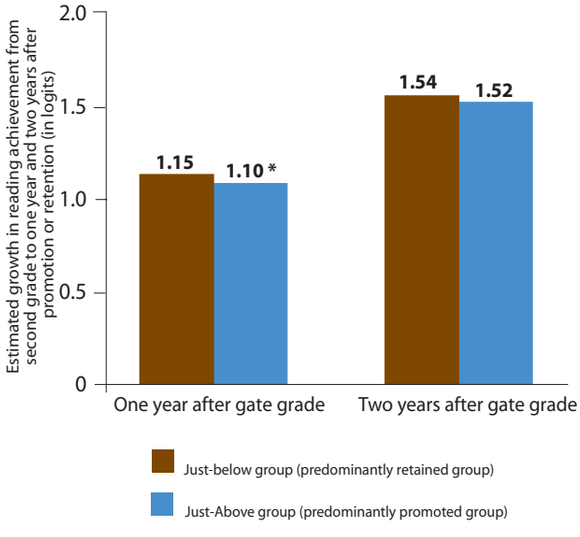


Note: Estimated growth in achievement at second grade to one year after the promotional decision, adjusted for prior achievement and demographic characteristics.

* Equals the estimated difference in achievement growth, which is statistically significant at p<.05 .

Figure 5-4

Third Graders in the Just-Below and Just-Above Cutoff Groups Had Similar Achievement Gains Two Years after the Retention Decision



Note: Estimated achievement growth is adjusted for demographic characteristics and prior achievement.

* Indicates the estimated difference in achievement growth between the predominantly retained and predominantly promoted group is statistically significant at p<.05.

As seen in Figure 5-3, these students had slightly higher estimated average learning gains between second grade and one year after the initial promotion or retention decision than either group of students in 2000 (.04 logits higher). This effect is small, but it suggests that the third graders in the predominantly retained group experienced a slight boost in performance in the post-gate year. This effect occurred after we adjusted for general differences in achievement patterns between the 1998, 1999, and 2000 cohorts.

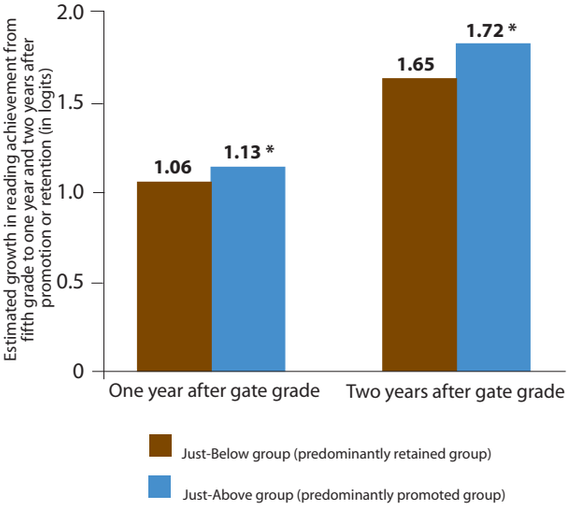
Third and Sixth Graders in the Just-Below and Just-Above Cutoff Groups One and Two Years after the Retention Decision

Figure 5-4 extends our analysis by estimating the reading achievement growth of third-grade students in the predominantly retained and predominantly promoted groups over two time periods: between second grade and one year after the initial promotion or retention decision (post-gate 1), and between second grade and two years after the initial promotion or retention decision (post-gate 2). Here we compare the estimated achievement gains in reading of those students in the 1998 and 1999 predominantly retained group (just-below cutoff group) to students in those same years in the predominantly promoted group (just-above cutoff group). Again, we find that third graders in the predominantly retained group had slightly larger estimated test-score gains between second grade and post-gate 1 than their counterparts who attended Summer Bridge and were promoted with test scores just above the promotional cutoff. However, these gains were short lived. Third graders in our predominantly retained group had similar achievement gains between second grade and post-gate 2 as their counterparts who were predominantly promoted. Thus, viewed as an estimate of the effect of retention, this comparison suggests that retention had little positive or negative effect on reading achievement among third graders.

Results for the sixth grade are more negative. Figure 5-5 compares, for 1998 and 1999, the estimated reading achievement gains of sixth graders who just missed meeting the promotional gate (just-below group) to students who just passed the promotional gate (just-above group). The estimated reading achievement growth between fifth grade and one year after promotion or retention of sixth graders in our predominantly retained group (just-below group) was 1.06 logits. This was statistically significantly lower than the reading achievement growth of the just-above cutoff group, 1.13 logits. This difference in reading achievement growth was also observed two years after the initial promotion or retention decision (post-gate 2). While more negative, these differences are small. If we translate this estimated difference in achievement growth at post-gate 2 into a percentage of the estimated learning gains of the predominantly promoted group, the estimated achievement gains of students in the predominantly retained group was approximately 4 percent lower. Thus, in the sixth grade, our comparison of the just-above and just-below cutoff groups suggests a statistically significant but small negative effect of retention.

Figure 5-5

Sixth Graders in the Just-Below Cutoff Group Had Significantly Lower Achievement Growth Than Students in the Just-Above Cutoff Group



Note: Estimated achievement growth is adjusted for demographic characteristics and prior achievement.

* Indicates the estimated difference in achievement growth between the predominantly retained and predominantly promoted groups is statistically significant at p< .05.

Main findings:

Comparison 2 estimates the pre- to post-gate achievement growth of students in our sample by their status two years after the retention decision.

Again we find little evidence that retention helped boost the achievement growth of third graders. The estimated reading achievement growth from the pre-gate grade to two years after the initial promotion or retention decision was not different between third graders who experienced a full year of retention and those who were promoted at the end of Summer Bridge.

The estimated effect of a full year of retention for sixth graders was much more negative than estimated in Comparison 1. The estimated reading achievement growth for sixth graders who experienced a full year of retention was 25 percent lower than their counterparts who had been promoted at the end of Summer Bridge. This might be an overestimate of the effect of retention because sixth graders who experienced a full year of retention were both those who were not waived at the end of the summer and those who could not raise their test scores to the promotional gate by January, and thus were lower-achieving students.

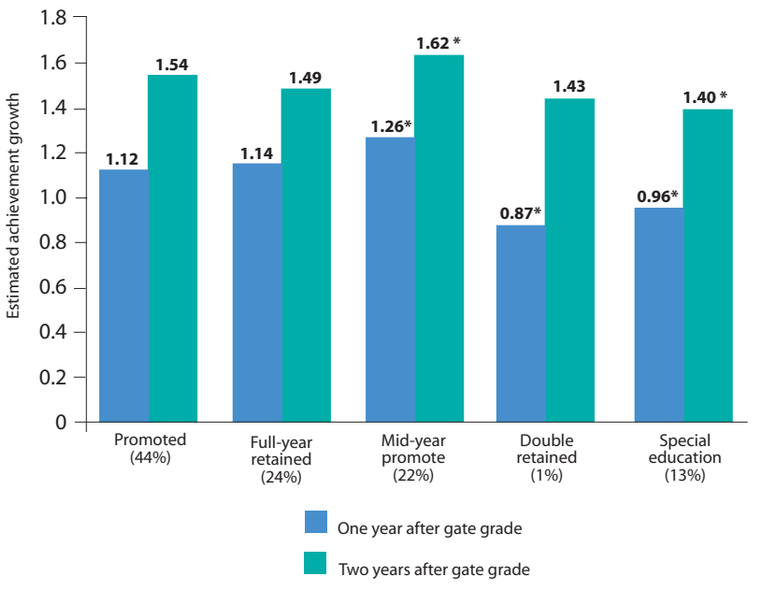
Retained students who were placed in special education experienced a deterioration in their achievement growth relative to that of other low-achieving students and their prior achievement trajectories. Retained students who were placed in special education after retention were, by virtue of that status, struggling in the year after retention. They also continued to struggle two years after the gate grade. These effects were most pronounced among sixth graders.

Retained students who were allowed to rejoin their classmates after passing the ITBS in January had achievement growth similar to their low-achieving counterparts who were promoted. Two years after the initial promotion or retention decision, third graders who were retained but then passed in January had slightly higher average learning gains on the ITBS than students who were promoted at the end of the summer. These slight positive effects were not found among sixth graders.

Comparison 2: Estimating the Effect of Retention by Students' Experiences of Retention

So far, we have compared the ITBS reading achievement growth for students who scored just above versus just below the promotional cutoff. As we discussed earlier, not all students who failed the promotional cutoff in reading in 1998 and 1999 were retained at the end of the summer. In addition, we saw in the previous section that many students who were initially retained at the end of the summer later rejoined their classmates after a January promotion and many were placed in special education. Thus, our estimate of the effects of retention in the previous section is the average achievement growth of students who didn't reach the promotional cutoff at the end of the summer and who in fact had very different experiences during the retained year—some were promoted right away, some passed in January, some were retained the whole year, and some were placed in special education. In this section, we disaggregate these different experiences and look at the achievement growth of all students in our sample (those who scored just above or just below the cutoff) by whether they: (1) were promoted at the end of the summer and remained with their age-appropriate classmates two years later (*promoted),* (2) experienced a full year of retention and remained one grade below their age-appropriate counterparts two years later *(retained)*, (3) later rejoined their age-appropriate classmates after initially being retained (*mid-year promotes*), (4) were placed in special education sometime within two years of failing to meet the

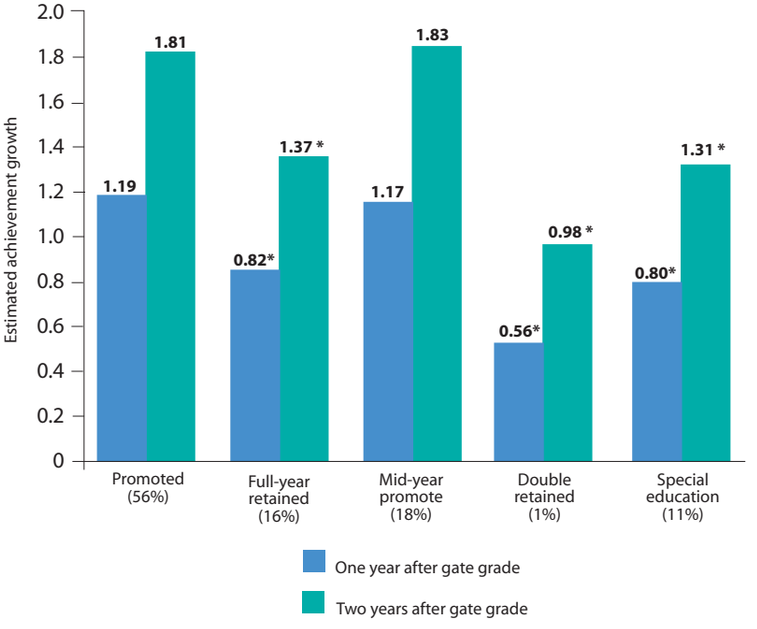
Figure 5-6
Third Graders Who Were Retained One Full Year Had Almost No Estimated Difference in Achievement Growth Versus Promoted Students



Note: Estimated growth in achievement from second grade to one year and two years after the promotional gate, by status, two years after gate grade for students in just-below and just-above cutoff samples, adjusted for cohort and demographic characteristics of students.

* Indicates estimated difference in achievement growth from promoted group is statistically significant at p<.05.

Figure 5-7
Sixth Graders Who Were Retained One Full Year Had Lower Achievement Growth Than Promoted Students



Note: Estimated growth in achievement from fifth grade to one year and two years after the promotional gate by status two years after gate grade for students in just-below and just-above cutoff sample, adjusted for cohort and demographic characteristics of students

* Indicates the estimated difference in achievement growth from promoted groups which is statistically significant at p<.05.

cutoff (*special education*), or (5) experienced a second retention and were two grades behind their age-appropriate classmates two years later (*double retention*). These estimates, again, were obtained using a growth curve model that included statistical controls for student demographic and school characteristics as well as pre-gate differences in their initial achievement.

Figure 5-6 presents our estimates of the achievement growth for third graders in our low-achieving sample by their experience of retention. To restate, our low-achieving sample are those third graders in 1998 and 1999 who were in our previous comparison, that is, students whose highest reading test scores were just below (2.4 GEs to 2.7 GEs) or just above (2.8 GEs to 3.1 GEs) the test-score cutoff.

As in our first comparison, we conclude that third graders who experienced a full year of retention did no better than similar low-achieving students who were promoted. Between second grade and one year after the promotional gate, promoted students in our sample gained 1.12 logits in reading on the ITBS, and students who were retained for a full year gained 1.14 logits in reading. Thus, estimated reading achievement growth of third graders who experienced a full year of retention (full-year retained group) and who remained one grade below

Why Use Multiple Methods to Estimate the Effect of Retention?

In this section, we present three different approaches to estimating the effect of retention on student's achievement growth. In the first two comparisons, we used growth-curve analysis which allowed us to address regression to the mean effects and to examine how students' achievement growth varied by their experience of retention. Comparison 1 attempts to control for selection effects by comparing students who scored just below and just-above the cutoff regardless of promotion or retention, but does not address the fact that the retention experience varied across students (i.e. not all students were retained the full year). Comparison 2 looks also at the achievement growth of students by their experience of retention but may lead to substantial overestimates of the effect of retention.

Finally, our two-stage probit least squares model does address selection concerns but may not adequately address regression to the mean problems because we estimate the effects of retention while controlling for gate-grade test scores which might overestimate the true prior achievement of students in our just above group. In addition, the two-stage probit least squares model does not allow us to examine differences in achievement growth across students by their experience during the retained year. Thus, each method has its strengths and weaknesses, and each method has its own set of assumptions. By using multiple methods, we can gain the insight that each approach provides while ensuring that our conclusions about the effects of retention are not being driven by particular assumptions or problems in any one model.

their age-appropriate classmates was slightly, but not significantly, greater than promoted students at the end of the year after the promotional gate and was not different two years after their initial retention.

Results for sixth graders who experienced one full year of retention (full-year retained group) are much more negative than we estimated in our first comparison. As seen in Figure 5-7, sixth graders in our sample who were promoted at the end of summer had test-score gains of 1.19 logits between fifth grade and one year after the gate-grade. But students in this sample who were retained for a full year gained only .82 logits. Thus, the estimated reading achievement growth between fifth grade and one year after the initial promotion or retention decision for sixth graders in our sample who experienced a full year of retention was .37 logits lower than their low-achieving counterparts who were promoted at the end of summer, and .44 logits lower after two years. This suggests that between fifth grade and two years after failing to meet the cutoff, sixth graders who experienced a full year of retention had estimated reading achievement growth 24 percent lower than their low-achieving promoted counterparts who were promoted at the end of Summer Bridge.

The number of students who were double retained in our sample was small because the district began to grant waivers to retained students who failed to meet the promotional cutoff their second time through the grade. As discussed previously, retained students experienced much higher rates of special education placement one and two years after the retention decision. Eleven percent of third graders and 13 percent of sixth graders in our just-below and just-above cutoff groups were placed in special education within two years after the gate grade. In the third grade, the estimated reading achievement growth between the second grade and one and two years after the gate grade for students who were placed in special education was significantly lower than other promoted low-achieving students. Retained sixth graders who were later placed in special education had substantially lower achievement growth, nearly 30 percent lower, than their counterparts who were promoted at the end of the summer.

Thus, even after accounting for differences in students' prior achievement, students who were placed in special education experienced deterioration in their achievement growth relative to other low-achieving students who were promoted and relative to their prior test-score trajectories.

Finally, a unique and controversial aspect of the Chicago policy was the decision to allow students to rejoin their classmates after passing the promotional standard in January and after completing an extra summer school session. In the third grade, these students had small but significantly greater achievement growth than promoted students both one year and two years after the promotional gate grade. In the sixth grade, the achievement growth of the mid-year promoted group was no different than promoted students in our sample. Even if we assume that these students represent the most resilient of the originally retained students, there is no evidence that the short-term experience of retention had a negative impact on their achievement growth. There is, however, no evidence that it benefited their achievement growth in basic skills in reading.

John Booz

Main findings:

Comparison 3 directly addresses problems of selection into promoted and retained groups at the end of the summer by estimating the probability of retention using variation in the administration of the policy across regions and years.

Not accounting for selection effects leads to substantial overestimates of the effect of retention, particularly in the sixth grade. The estimated effect of retention in the sixth grade falls by two-thirds when accounting for the fact that many sixth graders with test scores below the cutoff were promoted (waived) at the end of the summer and that these students might have done better than those who were retained.

Even after controlling for selection effects, retention was associated with a negative effect on achievement growth in the sixth grade. In the third grade, estimates based on the predicted probability of retention suggest that retention is again associated with no difference in the post-gate achievement growth of students. In the sixth grade, students with high probabilities of retention experienced significantly lower achievement growth in the year after the initial promotion or retention decision.

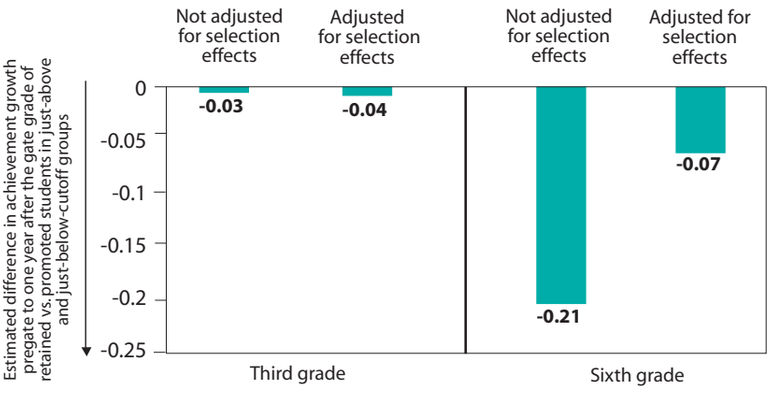
Comparison 3: Estimating the Achievement Effects of Retention by Using a Predicted Probability of Retention

Our analysis in the previous two sections presents a significant problem. In the third grade, we drew similar conclusions about the effects of retention when we compared the achievement growth of students who scored just below versus just above the promotional cutoff and when we compared students in that same sample who were promoted versus those that experienced a full retained year. In the sixth grade, our estimates of the effect of retention were quite different, with the second comparison showing a large, negative effect of retention for those who were retained for the full year, whereas the first comparison showed a smaller negative effect. This disparity may be driven by the fact that so many more sixth graders than third graders who just missed the promotional cutoff (just-below cutoff group) were promoted. In 1998 and 1999, 74 percent of third graders in the just-below cutoff group were retained compared to only half of sixth graders in the just-below cutoff group (see Figures 5-1 and 2-2). One hypothesis is that the much higher waiver rates in our just-below cutoff sixth-grade group diluted the effect of retention, leading us to underestimate the achievement effect. Another hypothesis is that the much higher waiver rates at the end of the summer in the sixth grade meant that when we compare the full-year retained students to promoted students, we are looking only at the most problematic sixth graders who might have done significantly worse regardless of whether they had been retained.

Our final approach is to use a more advanced statistical method to directly estimate the probability of retention and to use the predicted probability of retention as a base for our estimate of achievement effects rather than students' actual retention status. In this approach, we estimate the probability of retention using variables that predict the retention decision, but we do not expect to be associated with unmeasured characteristics that will shape their later achievement growth such as student motivation or parental involvement. We can then use these variables to predict a student's probability of retention and estimate post-gate achievement on the basis of this new predicted variable. This technique, called the two-stage probit least squares model, is described in more detail

Figure 5-8

After Controlling for Selection Effects, the Negative Effect of Retention is Reduced but is Still Significant for Sixth Graders



Note: Selection effects results from being promoted or retained at the end of summer for reasons other than test scores. Estimates are adjusted for cohort and students' demographic characteristics. This figure shows the effect of retention on third- and sixth-grade students' achievement growth. It compares the results of an OLS regression to that of a two-stage probit least squares analysis.

in the accompanying technical paper to this report.[4]

We were able to take advantage of differences in the administration of the policy to find variables that fit these criteria. As noted earlier, there was wide variation in Chicago's six region offices in their willingness to grant waivers. Both this variation across regions and changes in the administration of the policy over time meant that students' probability of retention at the end of the summer varied by the year students were born and the region in which students went to school. These three characteristics of the retention decision allow us to predict the probability of retention at the end of the summer regardless of students' own motivation and achievement characteristics. Once again, we predict differences in the achievement growth of students after we have accounted for students' prior test scores, demographic, and school characteristics and restrict the analysis to students who fell into our just-above and just-below cutoff groups. In this comparison, we wanted to use data on the 2000 cohort and because we only have equated Rasch-scale scores up to 2001, we are able only to look at effects one year after the gate grade (see *What is a Rasch-Equated Scale Score and Why is it Important to Measure Achievement in Equated Scale Scores Rather Than in Grade Equivalents?* on page 34).

Figure 5-8 compares estimates of the effect of retention if we use only information on whether students were retained or promoted (no selection controls) versus using a predicted probability of retention at the end of the summer. A comparison of these estimated effects confirms our suspicion that not accounting for selection effects in the retention decision leads to overestimates of the negative effects of retention even after controlling for students' demographic characteristics and pregate achievement. The effects of selection are significantly greater in the sixth grade. If we simply compare the estimated achievement growth of all sixth graders in our sample who were promoted to those that were retained, we find a large and statistically significant (-.21 logits) negative effect of retention. But, when we use the two-stage probit least squares model that controls for selection effects, our estimates of the effect of retention on achievement growth in reading falls to -.07 logits. Thus, the estimated difference in reading achievement growth between promoted and retained sixth graders one year after the gate year falls by two-thirds when controlling for selection effects (-.21 to -.07). This suggests that our Comparison 2 estimates of the effect of retention may be substantially overestimated.

At the same time, our estimates of the effect of retention on sixth graders' gains in achievement one year after the promotional gate are more negative than we observe in our first comparison. In the sixth grade, our predicted retention results suggest that retention was associated with significantly lower

achievement growth in the year after the gate grade. In the third grade, estimates based on the predicted probability of retention suggest that retention is associated with no differences in students' post-gate achievement growth, as opposed to the slight, short-term boost in performance we estimated in Comparison 1.[5]

Discussion

In this section, we presented three alternative methods of estimating the effects of being retained on student achievement, both one and two years after the retention decision. While there are some differences, the results are generally consistent. First, in the third grade, we find little evidence that students who were retained did better than their low-achieving counterparts who were promoted. In the first two comparisons, we found that students who were in our just-below cutoff (predominantly retained) group and those who experienced a full year of retention and remained a grade behind their age-appropriate classmates had a small boost in performance the year after their initial retention with no substantial positive effects two years after. Our estimates using the predicted probability of retention differ slightly, not showing the small short-term positive effect we saw in the other two comparisons. These differences, however, are moderate. In all three cases, we reach similar conclusions, that retention did not proffer any academic benefits to third graders who were retained, nor did it have any substantial negative effects.

. . . in the third grade, we find little evidence that students who were retained did better than their low-achieving counterparts who were promoted. . . . In the sixth grade, the question is how much did retention hurt?

In the sixth grade, the question is how much did retention hurt? In all three comparisons, we find that retention (either estimated by our just-below cutoff group, our full-year retained group, or by a predicted probability of retention), was associated with a negative growth in achievement, with that effect being larger two years after the gate grade. The most conservative estimate, Comparison 1, suggests that the pre- to post-gate grade learning gains of sixth graders in our predominantly retained group was approximately 3 percent lower at the first year after the gate grade than students in our predominantly promoted group, a difference that was not statistically significant. The two-stage probit least squares results suggest that the pre- to post-gate grade learning gains of students with high probabilities of retention at the end of the summer were 6 percent lower.

Students who were double retained and those who were placed in special education after retention by virtue of that status were struggling in the year after retention. They also continued to struggle two years after the gate grade. These effects were most pronounced among sixth graders. Even if we assume that the significant decline in these students' performance was not solely the result of retention, it is clear that neither placement in special education nor a third year in the same grade were effective educational strategies. There are many reasons why the practice of double retention makes little sense and the district did not double retain students after the first years of the policy. But, since the inception of the policy, we have observed very high rates of special education placements. Consistent with other research, there is little evidence that these students benefited academically from being placed in special education.[6]

In the 1998 and 1999 cohorts, the district allowed many students who were retained to rejoin their age-

appropriate classmates. This largely untested policy was controversial. Would allowing low-achieving students to essentially skip fourth or seventh grade set these students up for failure later on? Or would it provide an appropriate mix of remediation and acceleration that would allow them to avoid the potentially negative effects of retention while providing them extra support? Third graders who were initially retained and then rejoined their classmates had slightly higher learning gains between second and fifth grade than students in our sample who were promoted at the end of the summer and thus attended fourth grade. There were no differences in the pre- to post-gate (fifth to eighth grade) achievement growth of sixth graders who were promoted to seventh grade and retained sixth graders who then skipped seventh grade. While our results should be interpreted with caution, there is no evidence that these students' tested achievement in basic skills was harmed by their short-term experience of retention and their resultant lack of instructional time in the subsequent grade. There is also no substantial evidence of any benefit. We do not know if these students missed substantive content that would have been covered in these grades that may later affect their school performance since the ITBS is a basic skills test in reading and mathematics and is not intended to align to standards about what students should know and be able to do in the content areas.

CHAPTER 6

Placing Results In Context

Main findings:

This section examines the achievement gap (difference from the average test score of their cohort) from first grade to two years after promotion or retention for (1) students who had very low test scores at the end of Summer Bridge (very low-achieving), (2) students who were just-below the reading test-score cutoff (up to .4 grade equivalents below), and (3) students who were just-above (up to .4 grade equivalents above) the test-score cutoff.

*Low-achieving students in all groups started school substantially behind their classmates and had fallen further behind before the promotional gate***.** Neither promotion nor retention led to a significant closing of this achievement gap for those who had low but not the lowest test scores (those who had test scores just above or just below the promotional test score cutoffs.)

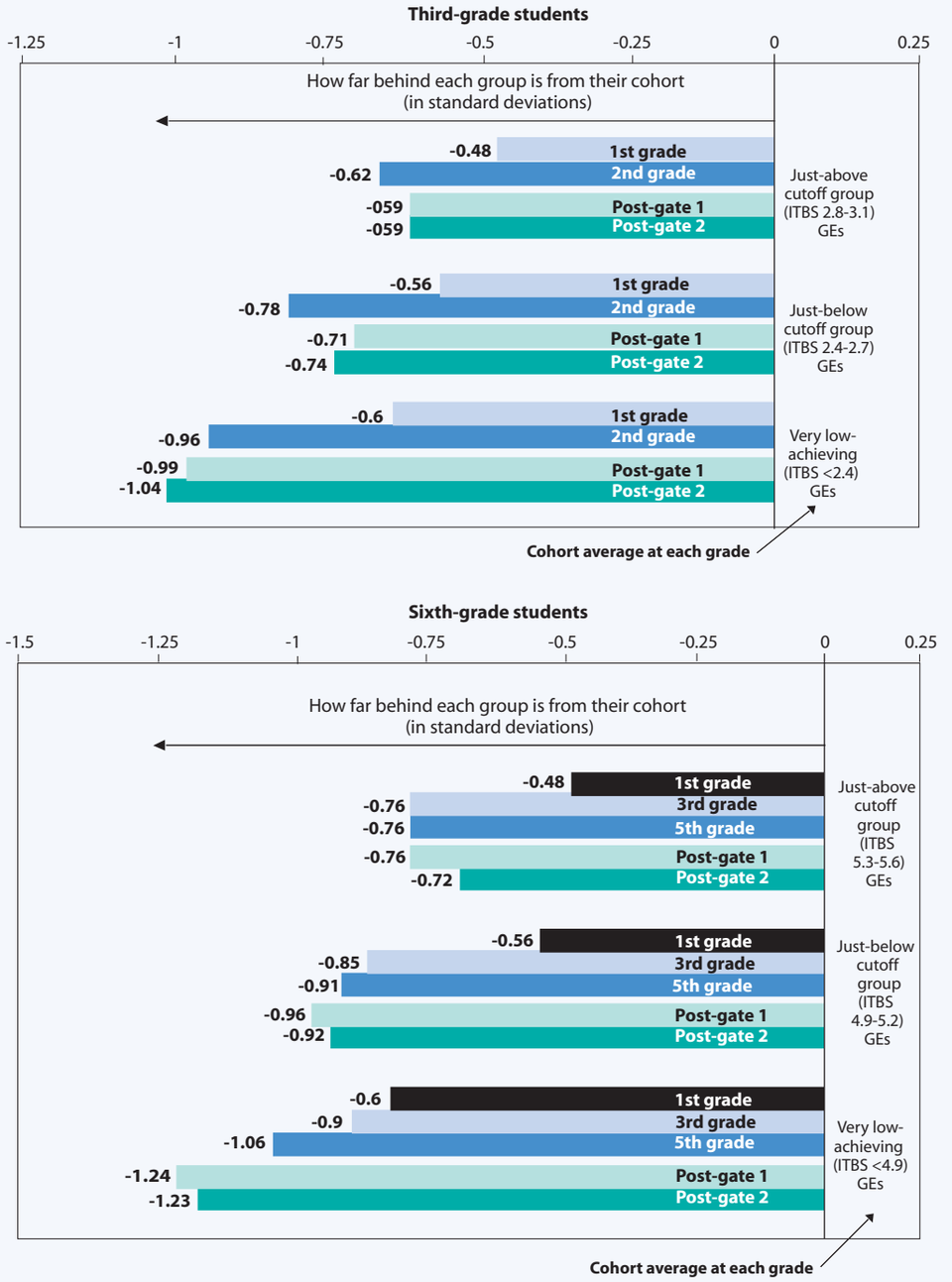
The lowest-achieving students in the school system (the majority of whom were retained) experienced a deterioration in their relative performance after retention. Two years after retention, the average achievement of very low-achieving sixth graders was nearly one and one-quarter standard deviations below the average reading achievement of their cohort. Without a comparison group, we cannot assess how much of their post-gate performance represents a negative effect of retention. But there is no evidence that retention stemmed the significant deterioration in their learning growth.

The Achievement Gap of Low-Achieving Summer Bridge Students Who Did and Did Not Meet the Promotional Test Cutoff

The previous section focused on estimating the effect of retention by comparing groups of students who were as alike as possible. Comparing students within a narrow band allowed us to develop an estimate of the effect of retention on a student's achievement growth. Figures 6-1 and 6-2 place these effects in context by showing trends in the achievement gap of our just-above and just-below-cutoff groups from the average achievement of

Figures 6-1 and 6-2

The Very Low-Achieving Students Started Behind before the Gate Grade and Continued to Fall Behind after the Gate Grade



These graphs compare third and sixth graders at risk of retention to the average achievement of students in their cohort both early in their school careers and after these students were retained or promoted. We look at three groups:

- The very lowest achieving students: Third and sixth graders with **VERY LOW** reading test scores at the end of Summer Bridge who were excluded from our previous analysis of the effect of retention on achievement.
- Low-achieving students with reading test scores **JUST BELOW** the promotional test-score cutoff at the end of the summer, the majority of whom were retained.
- Low-achieving students with reading test scores **JUST ABOVE** the promotional test score in reading (just-above cutoff), the majority of whom were promoted.

See Figures 5-1 and 5-2 for a description of these groups.

their cohort. We also present the average achievement gap of those low-achieving Summer Bridge students who we excluded from the previous achievement analysis—students whose highest reading test score was below 2.4 grade equivalents (GEs) in third grade and below 4.9 GEs in sixth grade. We call this third group our "very low-achieving" group. We defined a cohort as those students who were in the gate grade in the year the students faced the promotional cutoff and who were included in the policy. We measure the achievement gap as the difference from the average reading Rasch ITBS score for that cohort in that year measured in standard deviation units regardless of the grade students in that cohort attended. For example, the average reading test score measured in Rasch-scale scores for sixth graders in our just-above cutoff group was -3.15 logits in the first-grade versus -2.57 logits for their overall cohort. The standard deviation in first-grade reading achievement for that cohort was 1.2 logits. Thus, in the first grade, students in our just-above cutoff group had average reading test scores that were -.58 logits below the average students in their cohort, placing them .48 (-.58/1.2) standard deviations behind.

Students in our just-below and just-above cutoff groups began school substantially behind their higher-achieving counterparts and were already falling farther behind before the promotional gate. For both our third and sixth grade cohorts, the average first-grade reading ITBS achievement of the just-above and just-below cutoff groups was approximately one-half of a standard deviation (.48 logits and .56 logits respectively) below the average achievement of their cohort in the first grade. For all of the low-achieving groups, the achievement gap widened before the gate grade with most of that effect happening between first and third grade. The average reading achievement of sixth graders in our just-below and just-above group was .48 standard deviations below the average of their cohort in the first grade and was fully .91 standard deviations below the average of their cohort by the time they reached fifth grade, the year before the promotional gate. We see a similar trend among our third-grade cohort, with the achievement gap for all our low-achieving groups widening significantly between first and second grade.

After the sixth-grade promotional test, the average achievement of sixth graders in our above-cutoff group did move closer to the average of their cohort. Among this group, their relative achievement gap in reading declined from .79 standard deviations in fifth grade to .72 standard deviations two years after the gate grade (post-gate 2). The negative effect of retention among sixth graders, however, had the effect of maintaining the gap in our sixth-grade just-below group.

Students who were the lowest achievers in the school system (those in our very low-achieving group) were already substantially behind the average student in their cohort on entry into school and were falling further behind in their early grades. And, the lowest achieving students, the majority of whom were retained at the end of the summer of third and sixth grade, experienced deterioration in their relative performance after retention. Indeed, two years after retention, the average achievement of sixth graders in our very low-achieving group was nearly one and one-quarter standard deviations below the average reading achievement of their cohort. Because we do not have a comparison group for these students, we cannot definitely assess how much of their post-gate performance represents a negative effect of retention. Yet it is clear that there was no boost in performance after retention for these students, and that at least in the sixth grade, the gap between their performance and the average student in their cohort widened significantly. In essence, there is no evidence that retention stemmed the significant deterioration in their learning growth.

John Booz

INTERPRETIVE SUMMARY

In the era of No Child Left Behind, virtually every major school system in the United States is wrestling with the question of how to motivate students to achieve while at the same time addressing the needs of students who persistently struggle. The Chicago example illustrates vividly the magnitude of this problem. Perhaps the most compelling data presented in this report are how far and persistently behind the low-achieving students (both promoted and retained) who formed the focus of this analysis were.

Chicago's effort to end social promotion intended to address persistently low performance in two ways. First, the initiative aimed to decrease the numbers of very low-achieving students prior to the retention decision by using a combination of incentives and resources—incentives for students to work harder and for teachers and parents to direct attention to students at risk and resources for low-achieving students through after school programs and Summer Bridge. Although it is debatable whether it is a result of high-stakes testing, achievement test scores rose significantly in the period after 1996, particularly in the sixth and eighth grades, and the proportion of students with very low test scores fell.[1] But many students, over one-quarter of third graders and approximately 15 percent of sixth and eighth graders, did not raise their test scores to the promotional gate requirements. For these students, Chicago's approach was retention—a second dose of the material they had struggled with and continued incentives to work hard through additional chances to pass the promotional requirements.

Did retaining these low-achieving students help? The answer to this question is definitely no.

In this report, we focused on the question: Did retaining these low-achieving students help? The answer to this question is definitely no. In the third grade, there is no evidence that retention led to greater achievement growth two years after the promotional gate, and in the sixth grade, we find significant evidence that retention was associated with lower achievement growth. There is also evidence that retaining students under CPS's promotional policy significantly increased the likelihood of placement in special education.

Retaining a student under policies such as Chicago's presents teachers with an extremely difficult problem: What should teachers do with a student who is struggling, has been consistently behind, and needs to make substantial progress in raising his or her test scores to the promotional requirement in a short period of time? The Chicago administration gave little guidance or support to teachers in diagnosing that problem, in designing effective strategies, or in providing extra resources and training. The basic theory of action was that a second dose of the same material with continued incentives to pass the test would be enough. The conclusion of this report is that there is very little support for that approach to remediating poor reading skills. When a second dose wasn't enough, as evidenced by the many students who failed to raise their test scores to the promotional standard in the next year, teachers were faced with the question of how to create an alternative intervention. For the school system, the alternative intervention was ultimately to waive students to the next grade and wait for the next promotional gate.

Teachers and schools, however, increasingly turned to diagnosing the problem as a need for special education placement. These special education placements might reflect the fact that after students were retained, teachers did identify undiagnosed learning disabilities. They might reflect a mislabeling of students as facing learning disabilities because teachers lacked an alternative explanation and strategy for the difficulties students were presenting. Or, teachers and schools might have referred students to special education out of concern that, without that status and thus exemption from the policy, these students would not be able to progress—thus using special education as a means of getting struggling students around the policy. Most likely, the high placement of retained students in special education was driven by some combination of the above. This is not only a high-stakes testing problem. A National Research Council study (2002) on minority students in special and gifted education recently concluded that reading difficulties are one of the most frequent reasons that teachers refer students to special education. Retention under high-stakes testing without an alternative intervention, however, clearly highlighted those reading difficulties and increased the likelihood that teachers would look to special education as an answer. The problem, however, is that there is little research support for the idea that placement in special education leads to a remediation of students' reading problems, particularly in the upper grades.[2] At its best, special education can provide early intervention to prevent or ameliorate reading problems and give students access to services and more appropriate teaching.

The lack of additional resources and clear directives about how to approach the learning needs of retained students was certainly a problem for all schools. But, this problem was further exacerbated by the fact that, in Chicago, the teachers and schools that were most affected by the policy often had the fewest resources to draw on in mobilizing alternative approaches. Indeed, one of the central critiques of retention under high-stakes testing is that such approaches assume that low performance is an indicator of an individual student's problem rather than an indicator of poor teacher and school performance. Put crudely, this cri-

tique argues that it is not surprising that students who were retained struggled their second time through the policy because Chicago's approach to retention relied on those teachers and schools that failed the students in the first place to address the same students' learning needs the second time around.

There is some support for this argument. In 1998, there were 416 schools in Chicago that had more than 20 third-grade students subject to the policy. Fifty percent of retained third graders, however, were retained in the 100 Chicago public schools with the highest retention rates, and nearly two-thirds were retained in 150 high-retention schools. In the 100 schools with the highest retention rates, 42 percent of included third graders were retained, or 11.4 students per class of 27. In the sixth grade, one-half of all retained students were in 100 schools. Eighth-grade was even more concentrated, with 61 percent of retained students in 100 schools. While there were some efforts in the first year to provide some resources to schools hardest hit by retention, the Chicago administration never differentiated its approach toward retained students by whether a student was one of many in a class, one of many in a school, or one of only a few students to be retained.

If a second year in the same grade or placement in special education are not effective strategies for the remediation of very poor reading performance, then what is an alternative approach? Surely, social promotion alone is also ineffective. There is no evidence that low-achieving students performed significantly better when promoted to the next grade. We hope that the evidence presented in this report will spur debate both in Chicago and nationally over alternatives to social promotion and retention, as well as to more broadly identifying and managing the needs of low-achieving students who are persistently struggling. Other city and state school systems, such as Boston and North Carolina, have adopted approaches where students who are identified under high-stakes testing as not progressing receive more focused intervention as an alternative to retention. Such approaches attempt to adopt a middle ground by trying to garner some of the sorting and information function of high-stakes testing with alternative interventions. Unfortunately, there has been little research on whether these alternatives to retention, when implemented under high-stakes testing, provide an effective approach, and whether these school systems have actually implemented strong instructional and treatment strategies and monitored the content and implementation of the follow-up remediation plans.

One approach that is supported by the evidence presented in this report is to focus on earlier identification of learning problems. The average student in both our just-above and just-below cutoff groups, as well as those in our very low-achieving group, started substantially behind the average CPS student in first grade, and the achievement gap for these groups widened most significantly between first and third grade, before CPS's promotional policy took effect. Waiting until third or sixth grade to identify these students and intervene seems a nonjudicious use of resources. This does not mean that high-stakes testing should occur in the first grade. It does mean, however, that school systems must invest in developing effective early assessment, instruction, and intervention approaches that identify students who are not moving forward and provide appropriate supports.

All alternative approaches will require additional investment. In the end, the practice of retention is monetarily and academically costly. It involves investing in an extra year of schooling. It makes students overage for grade, and as a result, increases the risk of school dropout, an outcome with a substantial set of social costs. Instructionally, high-stakes testing leads to substantial costs in time on test preparation, and it directs resources away from early intervention.[3] If an expensive policy is simply not working, as concluded in this report, it would make little sense to invest more money in it rather than to redirect that money toward alternatives.

The problem with all alternatives, however, is that retention's real and social costs are not in the budget of an urban school system. As a result, ending retention does not mean that there are then available resources to redirect to alternative programs or

approaches. Not retaining a student does not mean that the school district can then use the money toward hiring a reading specialist. For example, using a crude calculation, if CPS receives from the state $5,000 for every student enrolled, retaining 7,000 students means that over time there are more students in CPS in a given year and a cost to the state over those students' school career of $35 million. This is certainly a high estimate because retention ultimately may make it more likely that some of those students will drop out. The alternative, lowering first-grade class size or providing intensive support in the early grades for students who are struggling, means trying to do more with $5,000 and does not bring any additional monies. Reducing rates of special education placement, particularly late placements will, however, reduce the expenses of a school system, and thus adopting this strategy not only has academic benefits, as this report suggests, but also provides a means of supporting more effective alternatives. Just as we need to understand whether alternatives to retention provide effective approaches, finance reform at both the state and federal levels must take seriously the magnitude of the task that urban school districts face and the importance of providing resources to adequately address the needs of those students who require effective alternatives.

APPENDIX

A Hierarchical Linear Model for Estimating Effects of Retention: The third- and sixth-grade models

In Chapter 5, we present results based on a growth-curve model that estimates the pre- and post-gate reading achievement growth of students in our just-above and just-below cutoff groups adjusted for overall differences in the achievement characteristics of each cohort as well as the students' demographic and school characteristics. To obtain these estimates, we use a three-level Hierarchical Linear Model (HLM) with test scores from the ITBS as the outcome (Roderick, Jacob, and Bryk 2003). We estimate two models (1) a basic model that compares the achievement growth of students in our just-below cutoff and comparison just-above cutoff groups regardless of retention or promotion, and (2) an extended model that estimates difference in the average post-gate achievement growth of students by their experience in the post-gate periods (e.g., whether students two years after the gate grade had experienced one retention, a double retention, placement in special education, or had rejoined their age-appropriate comparison groups). For the purposes of explication, the basic model is presented in full for the sixth grade.

Level 1— Measurement Model

$$Y_{ijk} = \phi_{0jk} + \phi_{1jk}(Test\ Year) + \phi_{2jk}(Sixth) + \phi_{3jk}(Post\ Gate\ 1) + \phi_{4jk}(Post\ Gate\ 2) + e_{ijk} \quad (1)$$

Level 1 is a repeated measures model in which Y_{ijk} is the achievement at Year i, for student j, in school k.[1] The *Test Year* variable is centered on third grade so that it takes on the value of zero if the test score is from when a student was in third grade (i.e., -1 for the year before third grade and 1 for the year after third grade). *Sixth* is a dummy variable that equals one in the year the student was in the sixth-grade promotional gate and zero otherwise. *Post-gate 1* is a dummy variable that equals one in the first post-gate year regardless of the grade in which a student is in enrolled. Similarly, *Post -gate 2* is a dummy variable that equals one in the second post-gate year. The coefficients on these dummy variables represent the extent to which the student's test score in that year deviated from the score that would have been expected based on the student's initial status and growth trajectory prior to the gate grade. We call these terms the achievement value-added for that grade.

Level 2— Students and Cohorts Within Schools

At Level 2 we model the baseline (third grade) intercept, the sixth-grade value-added term, and the two post-gate value-added terms as a function of cohort differences in achievement (1998 or 1999) and prior differences in test scores between our just-above and just-below cutoff groups (or in the case of comparison 1 whether third graders fell below the cutoff in the years 1998 and 1999 versus 2000). The baseline (third grade) and achievement growth (*test year* slope) of sixth graders are further adjusted for differences in the demographic characteristics of groups that may have affected their prior growth. This model assumes that the pre-gate achievement growth of students in our just-below and just-above-cutoff groups were comparable, an assumption consistent with our observation of the data and the theory behind using the just-above cutoff groups as a comparison for estimating the counterfactual.

$$\phi_{0jk} = \eta_{00k} + \sum_{q=1}^{6} \eta_{0qk}(Demographics_q)_{jk} + \eta_{07k}(Below)_{jk} + r_{0jk}$$

$$\phi_{1jk} = \eta_{10k} + \sum_{q=1}^{6} \eta_{1qk}(Demographics_q)_{jk} + r_{1jk}$$

$$\phi_{2jk} = \eta_{20k} + \eta_{21k}(Below)_{jk}$$

$$\phi_{3jk} = \eta_{30k} + \eta_{31k}(Below)_{jk}$$

$$\phi_{4jk} = \eta_{40k} + \eta_{41k}(Below)_{jk}$$

Demographics is a vector of student demographic characteristics that allows us to adjust for demographic differences in the groups that might shape their post-gate performance. These demographic characteristics include dummy variables for a student's gender (*male*), ethnicity (*African-American*, *Latino,* and *white and other races*), whether the student experienced a retention prior to the gate year (*prior retention*), and whether the student moved schools in the year of the promotional gate (*mobile*). We also include a measure of student poverty status *(neighborhood poverty)* derived from a geo-coding of students' addresses to their census block in 1990, and that is included as part of the Consortium on Chicago School Research's core data set.

Level 3 – School Level

At the third level, we allow the intercept and linear growth rate to vary randomly across schools.

$$\eta_{00k} = \nu_{000} + u_{00k}$$

$$\eta_{0qk} = \nu_{0q0}$$

$$\eta_{07k} = \nu_{070}$$

$$\eta_{10k} = \nu_{100} + u_{10k}$$

$$\eta_{1qk} = \nu_{1q0}$$

$$\eta_{20k} = \nu_{200}$$

$$\eta_{21k} = \nu_{210}$$

$$\eta_{30k} = \nu_{300}$$

$$\eta_{31k} = \nu_{310}$$

$$\eta_{40k} = \nu_{400}$$

$$\eta_{41k} = \nu_{410}$$

ENDNOTES

Chapter 1

[1] House (1998); Roderick (1994); Rose, Medway, Cantrall, and Marus (1983); and Shepard and Smith (1989).

[2] Roderick, Jacob, and Bryk (2003); Jacob, Stone, and Roderick, (2004).

[3] Byrnes (1989); Jacob et al. (2004); Tompchin and Impara (1992).

[4] Alexander, Entwistle, and Dauber, (1994).

[5] Alexander, et al. (1994); Holmes (2000); Peterson, DeGacie, and Ayabe (1987).

[6] Allensworth (2004); Grissom and Shepard (1989); House (1998); and Roderick (1994)

[7] House (1998).

Chapter 2

[1] In 1997, the eighth-grade cutoff was set at 7.0, 1.8 years below grade level, and was then raised each subsequent year to phase in higher standards for older students. In 1998, the cutoff was raised to 7.2; then in 1999 to 7.4; and finally to 7.7 in the spring of 2000. In that year, the sixth-grade cutoff was also raised to 5.5.

[2] The promotional decisions for special education students are based on criteria outlined in their Individual Education Plans (IEP) and may include test scores as one of the criteria. We call students *included* if their promotional decision was strictly based on the test-score cutoff. From 1996 to 1998, students were excluded if they had been enrolled in a bilingual education program for fewer than three complete years as of the prior September 30. In 1999, the criterion was changed to four years from three years.

[3] Most Chicago students remain in elementary school until eighth grade. It is Chicago's policy that students are not allowed to remain in elementary schools past age 15. As part of the ending social promotion initiative, CPS created Academic Preparatory Centers (APC), separate schools for overage students who did not meet the promotional cutoff. These APC's were supposed to provide support for students to raise their test scores and prepare for the transition to high school (Miller 2003). In the 2003-2004 school year, CPS substantially revamped the APC's, moving them into high schools with a revised curriculum and program.

[4] Some students who passed in January were promoted mid-year, particularly in the APC's, though schools were given wide flexibility in how to administer the policy. In the fall of 2000, this policy was discontinued after complaints from schools, particularly high schools, over the difficulties of dealing with students promoted mid-year. Students in APC's continued to take a January test but were not allowed to move to the high school mid-year. The system also provided an additional catch-up program, Making the Grade. The Making the Grade program was never large. In 1998, approximately 400 students were double promoted through this program. However, as seen in Table 4-1, many students rejoined their classmates after passing the test in January and then participating in the promotional summer school that was intended to "fill in" the material they would miss when they skipped a grade.

[5] Summer Bridge provides significant reductions in class size, on average 16, a highly prescribed and centrally developed curriculum that is aligned with the ITBS. Teachers are provided with daily lesson plans and all instruction materials. A multiyear evaluation of the program concluded that students in Summer Bridge, particularly in the sixth and eighth grades, experienced significant increases in their test scores over the summer and that the program was effective in raising the proportion of students who met minimum standards for promotion (Roderick, Engel, and Nagaoka, 2003).

[6] In the 1997-1998 school year, the district provided extra teachers to the 65 schools most affected by the promotional standards to reduce class size and hired retired teachers for extra supports.

[7] Stone and Engel (forthcoming).

Chapter 4

[1] Tracking the grade progression of retained students who passed in January is not an easy task. Students who passed in January had the opportunity to rejoin their classmates by being promoted mid-year. In addition, some retained students were given the opportunity to be double promoted after participating in a double-dose summer school. Schools were given flexibility in deciding whether to immediately move these students who had met the standards mid-year or to move them ahead at the end of the school year after a summer school. Most of the students were ultimately moved ahead to rejoin their age-appropriate classmates. We present the number of students who we know were enrolled in their age appropriate grade though we do not know when the transition occurred (during or after the retained year). This data was also checked by looking at the level of the ITBS the student was tested two years after the retention decision; e.g. whether students were taking a level 13 (seventh grade) or level 14 (eighth grade).

[2] During this period, the Chicago Public Schools relied largely on three forms of the ITBS. Thus, those students who did not meet the criteria had taken the same form of the ITBS twice.

Chapter 5

[1] Roderick et al. 2000.

[2] The non-comparability of ITBS and TAP tests are just one of many problems in estimating achievement effects among retained eighth graders in Chicago. One way of getting around this problem would have been to estimate a covariate model examining whether the relationship between ITBS scores in eighth grade and TAP scores in ninth grade (similar to our post-retention grade growth model) was different for retained and promoted youths. However, high dropout rates in the Chicago Public Schools, and, in particular, very high dropout rates among retained eighth graders could lead to a misestimate of the effects of retention because better-performing students would be more likely to remain in school and have test scores.

[3]We might expect that both the characteristics of retained and waived students and the experience of retention were different in the first year of the policy. As we saw, waiver rates were very high in the first year reflecting significant problems in implementation and high waiver rates among Latino students. And, the experience of retained students might have been quite different in the first year as schools had not had time to formalize their decisions and approach to students who did not meet the cutoff. Our analysis of the demographic characteristics of promoted versus retained students in the below-cutoff group in 1998 and 1999 showed no systematic demographic differences between the two groups.

[4] Roderick and Nagaoka, (2004).

[5] The problem with the two-stage least squares probit model is that we do include a student's gate-year test score as a control variable. At the beginning of this section, we argued that because students in our above-cutoff group were selected into Summer Bridge because of a below average test score, we might expect their post-gate achievement growth to be overestimated because of regression to the mean, leading to more negative estimated effects of retention. This is why we estimated a growth-curve model in Comparison 1 and Comparison 2. The two-stage least squares probit estimates are correct for selection effects and three years of prior achievement but do not completely adjust estimates for regression to the mean. Therefore, it is not surprising that results from the two-stage least squares probit model are slightly more negative than we estimated in Comparison 1.

[6] Lyon et al.(2001); Reynolds and Wolfe (1999).

Interpretive Summary

[1]Bryk (2002); Roderick, et al. (2002).

[2] Lyon, et al. (2001); National Research Council (2002); Reynolds and Wolfe, (1999).

[3] Jacob, et al. (2003).

Appendix

[1] The third- and sixth-grade model is estimated as a linear growth curve. For low-achieving students, the relationship between achievement and grade is relatively linear in the Rasch metric. This is not true for the larger cohort where growth in the Rasch metric tends to decelerate in the upper grades. Roderick, Jacob, and Bryk (2003) used a nonlinear quadratic model to fit growth in estimating the achievement effects of high-stakes testing for the entire cohort of sixth and eighth graders. The Level 1 model we used is slightly different than the one used in the Roderick et al. paper in that grade rather than year was the repeated measure.

WORKS CITED

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 grades*. New York: Cambridge University Press.

Allensworth, Elaine. 2004. *Ending social promotion in Chicago: Dropout rates in Chicago after implementation of the eighth-grade gate.* Chicago: Consortium on Chicago School Research.

Allensworth, Elaine. 2004. Consequences of the Implementation of the Eighth-Grade Promotional Gate on Dropout Rates in Chicago: Disentangling the influence of grade retention from that of rising achievement.

American Educational Research Association. 2000. AERA position statement concerning high stakes testing in preK-12 education. Washington DC.

Bryk, Anthony S. 2002. No child left behind: Chicago style. What has really been accomplished? Paper presented at the Taking Account of Accountability conference, Chicago.

Bryk, Anthony S., Yeow M. Thum, John Q. Easton, and Stuart Luppescu. 1998. *Academic productivity of Chicago Public Schools: A technical report*. Chicago: Consortium on Chicago School Research.

Bryk, Anthony S. and Stephen W. Raudenbush. 1992. *Hierarchical linear models.* Thousand Oaks, CA: Sage Publications.

Byrnes, Deborah A. 1989. Attitudes of students, parents and educators toward repeating a grade. In *Flunking grades: Research and policies on retention*, eds. L. A. Shepard and M. L. Smith. London: Falmer Press, 108-31.

Finlayson, Harry J. 1977. Nonpromotion and self-concept development. *Phi Delta Kappan,* 59(3): 205-6.

Grissom, James B. and Lorrie A. Shepard. 1989. Repeating and dropping out of school. In *Flunking grades: Research and policies on retention,* eds. L. A. Shepard and M. L. Smith. London: Falmer Press, 34-63.

Holmes, Charles T. 2000. Promoting the end of retention. *Journal of Curriculum and Supervision*, 15 (4):300-14.

Holmes, Charles T. 1989. Grade level retention effects: A meta-analysis of research studies. In *Flunking grades: Research and policies on retention,* eds. L. A. Shepard and M. L. Smith. London: Falmer Press, 16-33.

Holmes, Charles T. and Kenneth Matthews. 1984. The effects of non-promotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research,* 54(2):225-36.

House, Ernest R. 1998. *The predictable failure of Chicago's student retention program.* Boulder: University of Colorado School of Education.

Jacob, Brian A. and Lars Lefgren. 2002. *Remediation education and student achievement: A regression-discontinuity analysis.* NBER Working Paper No. 8918. Cambridge, MA: National Bureau of Economic Research.

Jacob, Robin Tepper, Susan Stone, and Melissa Roderick. 2004. *Ending social promotion in Chicago: The response of teachers and students.* Chicago: Consortium on Chicago School Research.

Keshk, Omar M. G. 2003. CDSIMEQ: A program to implement two-stage probit least squares. *The Stata Journal,* 3(2):1-11.

Lyon, G. Reid, Jack M. Fletcher, Sally E. Shaywitz, Bennett A. Shaywitz, Joseph K. Torgesen, Frank B. Wood, Ann Schulte, and Richard Olson. 2001. Rethinking learning disabilities. In *Rethinking special education for a new century,* eds. C. E. Finn, A. J. Rotherham and C. R. Hokanson. Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute, 259-87.

National Association of School Psychologists. 2002. Position statement on student grade retention and social promotion. Bethesda, MD.

National Research Council. 2002. *Minority students in special and gifted education.* Washington DC: National Academies Press.

Peterson, Sarah E., James S. DeGracie, and Carole Ayabe. 1987. A longitudinal study of the effects of retention/promotion on academic achievement. *American Educational Research Journal,* 24(1):107-18.

Pierson, Louisa and James P. Connell. 1992. Effect of grade retention on self-system processes, school engagement and academic performance. *Journal of Educational Psychology*, 84(3):300-7.

Reynolds, Arthur J. and Barbara Wolfe. 1999. Special education and school achievement: An exploratory analysis with a central-city sample. *Educational Evaluation and Policy Analysis,* 21(3):249-69.

Reynolds, Arthur J. 1992. Grade retention and school adjustment: An exploratory analysis. *Educational Evaluation and Policy Analysis,* 14(2):101-21.

Roderick, Melissa and Jenny Nagaoka 2004. *Retention under Chicago's High-Stakes Testing Program: Helpful, harmful, or harmless?* Chicago: Consortium on Chicago School Research.

Roderick, Melissa, Mimi Engel, and Jenny Nagaoka. 2003. *Ending social promotion in Chicago: Results from Summer Bridge.* Chicago: Consortium on Chicago School Research.

Roderick, Melissa, Brian A. Jacob, and Anthony S. Bryk. 2003. High stakes testing in Chicago: Effects on achievement in promotional gate grades. *Educational Evaluation and Policy Analysis*, 24:333-58.

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, Jenny Nagaoka, Jen Bacon, and John Q. Easton. 2000. *Update: Ending social promotion in Chicago: Passing, retention, and achievement trends among promoted and retained students.* Chicago: Consortium on Chicago School Research.

Roderick, Melissa, Anthony S. Bryk, Brian Jacob, John Q. Easton, and Elaine Allensworth. 1999. *Ending social promotion in Chicago: Results from the first two years.* Chicago: Consortium on Chicago School Research.

Roderick, Melissa. 1994. Grade retention and school dropout: Investigating the association. *American Educational Research Journal*, 31(4):729-59.

Rose, Janet S., Frederick J. Medway, V. L. Cantrell, and Susan H. Marus. 1983. A fresh look at the retention-promotion controversy. *Journal of School Psychology,* 21:201-11.

Shepard, Lorrie A., Mary Lee Smith, and Scott F. Marion. 1996. Failed evidence on grade retention. *Psychology in the Schools,* 33(3):251-406.

Shepard, Lorrie A. and Mary Lee Smith, eds. 1989. Introduction and overview. In *Flunking grades: Research and policies on retention*. London: Falmer Press, 1-15.

Stone, Susan and Mimi Engel (forthcoming). Same old, Same old? A qualitative look at retention under Chicago's ending social promotion policy.

Tomchin, Ellen M. and James C. Impara. 1992. Unraveling teachers' beliefs about grade retention. *American Educational Research Journal,* 29:199-223.

This report was produced by the Consortium's Publications and Communications Department:
Sandra Jennings, *Associate Director of Publications and Communications*
Melissa Dean, *Publications Editor*
Kumail Nanjiani, *Webmaster*
Lura Forcum, *Copy Editor*
Patricia Collins, *Publications Project Assistant*

ABOUT THE AUTHORS

Jenny Nagaoka is the Project Director of the Consortium's research on the post-secondary outcomes of Chicago graduates. Prior to this current project, she was the Project Director of the Chicago Public School's Student Development Planning Initiative, a joint project with the University of Chicago and the Chapin Hall Center for Children. Previously, she was a Research Analyst at the Consortium on Chicago School Research, where her work included research on quality of classroom instruction and an evaluation of Chicago's summer program. She received her BA from Macalester College in St. Paul, Minnesota and her MA from the Harris School of Public Policy at the University of Chicago.

Melissa Roderick is an Associate Professor at the School of Social Service Administration at the University of Chicago. She is an expert in urban minority adolescent development and urban school reform. At the Consortium, Professor Roderick has conducted research on student performance in the transition to high school and leads the Consortium's multiyear evaluation of Chicago's efforts to end social promotion. This work included an intensive investigation of the impact of Chicago's Summer Bridge program, analysis of teachers' response to the policy, and quantitative and qualitative analysis of the impact of retention on students' progress and achievement. Her work brings together rich longitudinal investigation of student experiences with large-scale statistical analyses of student and school performance. Professor Roderick received a master's degree in public policy from the Kennedy School of Government at Harvard University, and a PhD in public policy from Harvard's Graduate School of Arts and Sciences.

Consortium on Chicago School Research

Mission

The Consortium on Chicago School Research aims to conduct research of high technical quality that can inform and assess policy and practice in the Chicago Public Schools. By broadly engaging local leadership in our work, and presenting our findings to diverse audiences, we seek to expand communication between researchers, policy makers, and practitioners. The Consortium encourages the use of research in policy action, but does not argue for particular policies or programs. Rather, we believe that good policy is most likely to result from a genuine competition of ideas informed by the best evidence that can be obtained.

Founded in 1990, the Consortium is located at the University of Chicago.

Consortium on Chicago School Research
1313 East 60th Street, Chicago, IL 60637
773-702-3364 fax -773-702-2010
www.consortium-chicago.org

ISBN 0-9726035-4-9
90000>
9 780972 603546