

Ending Social Promotion: The Response of Teachers and Students

February 2004

Robin Tepper Jacob
Susan Stone
Melissa Roderick

CHARTING REFORM IN CHICAGO SERIES

CONSORTIUM ON CHICAGO SCHOOL RESEARCH © 2004

Ending Social Promotion:

Response of Students and Teachers

February 2004

Robin Tepper Jacob

Susan Stone

Melissa Roderick

TABLE OF CONTENTS

Introduction	1
Chapter 1: The Educators' Perspective on the Impact of Ending Social Promotion in Chicago and on the Practice of Retention.....	9
Chapter 2: The Instructional Response to High-stakes Accountability Trends in Test Preparation and Content Emphasis in Mathematics and Reading 1994-2001	31
Chapter 3: The Student's Perspective: Trends in student reports of academic support from teachers and parents, engagement in school, and perceptions of academic expectations and efficacy	57
Interpretive Summary	79
Appendices	87
Endnotes	99
Works Cited	103



John Booz

INTRODUCTION

In the mid-1990s, Chicago entered a second wave of school reform with the mayoral takeover of the school system. The hallmark of the new administration's approach was strong accountability. In 1996, the Chicago Public Schools (CPS) made national headlines when it embarked on an ambitious accountability agenda by coupling a new school-level accountability program with high-stakes testing for students. Under this new initiative to end social promotion, the city's lowest-performing third, sixth, and eighth graders would repeat a grade at least once if they did not meet minimum reading and math test-score cutoffs on the Iowa Tests of Basic Skills (ITBS), and in subsequent years, additional metrics of student performance. Additionally, low-performing schools were placed on academic probation, receiving significant intervention and pressure to improve.

Since 1996, test scores among the lowest-achieving students have risen dramatically, particularly in the upper grades. The proportion of schools with extremely low performance, moreover, has fallen dramatically.¹ But there is considerable debate about the source of these test-score increases and the degree to which observed learning gains can be sustained or generalized to other contexts. While test-score increases may reflect improved student motivation, parental involvement, or changes in classroom content or pedagogy, many have suggested that the gains may simply be a result of extensive test-preparation activities and "teaching to the test," or greater motivation on the part of students on the day of testing.²

By 1999, there was still no conclusive evidence about the results of the policy's implementation. Nevertheless, the nation had taken notice of Chicago's improved test scores. That year, President Clinton made it evident that Chicago had the right idea, calling for an end to social promotion across the nation.

The central question is: Did high-stakes accountability cause CPS's teachers, parents, and students to change their behavior in ways that would lead to higher achievement, or does the evidence suggest that CPS's initiatives resulted in simply more focus on testing? This report takes an in-depth look at CPS teachers' responses to the high-stakes-testing initiatives and the impact on students' school experiences. It examines teachers and principals' assessments of the policy, tracks changes in instructional practice over time, and examines trends in critical student indices.

We begin by exploring the view of the educators in low-performing schools. In 1999 and 2001, the Consortium on Chicago School Research's surveys included supplemental questions about CPS's efforts to end social promotion. The supplement asked teachers and principals to assess the impact of the policy on stu-

dent learning and behavior, on parental attitudes and involvement, and on their own educational practices. The survey allowed us to examine educators' views of grade retention, and in particular, the extent to which educators felt the policy was consistent with their own pedagogies.

We then look at changes in teachers' reports of their instructional practices both before and after the landmark 1996 reforms. Most prior research finds that teachers do not support accountability policies that rely only on standardized test scores to judge student performance.³ Yet there is also considerable evidence that teachers are highly responsive to accountability programs and often align their curriculum with the content of the test, spending more time on test preparation in response.⁴ Since 1994, the Consortium on Chicago School Research has conducted biannual surveys of all CPS teachers and principals, and all sixth-through 10th-grade CPS students. Each year, the survey asks teachers to report how much time they spend on test-preparation activities and on the content they cover in reading and mathematics. These longitudinal surveys allow us to improve upon other studies of high-

Summer Bridge and Lighthouse Programs

The Summer Bridge summer-school program is a critical component of the Chicago Public Schools' efforts to end social promotion. Third, sixth, and eighth graders who do not meet the promotion test-score cutoffs in the spring are required to participate in the Summer Bridge program and are retested at the end of the summer. In the first several years, more than one-third of third, sixth, and eighth graders failed to meet the promotion test cutoffs by the end of the school year, and over 22,000 students have attended Summer Bridge each year.

The program provides six weeks of instruction for three hours per day for third and sixth graders. Eighth graders attend four hours a day for seven weeks. Summer Bridge provides significant reductions in class size (on average 16 students per class) and a highly prescribed and centrally developed curriculum that is aligned with the Iowa Tests of Basic Skills (ITBS).¹ Teachers are provided with day-to-day lesson plans and all instructional materials. They are expected to keep pace with the lesson plans, and monitors visit classrooms to check if teachers are on pace. 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.²

The Lighthouse after-school program was also intended to provide academic support, but with the goal of helping students to meet promotion requirements during the school year. Lighthouse began as a small program in 1996-1997 in 40 low-performing elementary schools.³ Lighthouse expanded rapidly to 147 schools in 1997-1998. By the 1999 to 2001 school year, more than 356 CPS elementary schools provided Lighthouse academic support to more than 81,000

stakes testing that have relied on data collected only after the introduction of the testing.⁵ In this study, we are able to trace changes in the magnitude of teachers' responses over time in low-performing schools and are also able to explore how instructional content changes differ across these schools.

Finally, one of the purposes of high-stakes accountability is to change the behavior of both educators and parents so that students receive greater attention, support, and raised expectations for achievement. It is designed to compel the lowest-performing schools to exert more pressure on students to acquire and retain knowledge and to organize instruction to promote achievement.⁶ Chicago's ending social promotion initiative was intended to make educators and parents pay closer attention to the lowest-performing students. Parents would be encouraged to become more involved and to support and monitor their children's performance due to the prospect that their child would not move on to the next grade without meeting achievement test scores. The policy sends an equally strong message to teachers that they must pay more attention to those students at risk of retention. In addition,

as will be discussed below, the policy was accompanied by significant resources for after-school and summer programs, which were designed to provide academic assistance to low-performing students (see *Summer Bridge and Lighthouse Programs*). Finally, the policy sent strong messages to students that their achievement matters, presumably to provide motivation to students to increase their engagement and work effort. Together, then, these changes in student, parent, and teacher behavior would lead to increases in academic achievement, particularly for those students most behind academically.

Opponents of high-stakes testing worry, however, that such policies will have precisely the opposite effects. If high-stakes testing leads to an overly narrow focus on basic skills and test preparation, students may see their class work as less engaging and relevant to their lives. Opponents worry that the policy will have the most negative effects on the engagement of those students it is most intended to help. If students with low skills feel that promotion test cutoffs are out of their reach and that they cannot do well on standardized tests based on their past experience, they

students for a total cost of over \$16 million, making it one of the largest after-school programs in the country. Lighthouse, like Summer Bridge, was intended to provide additional instructional time and support for passing the ITBS. Lighthouse met for 20 weeks from mid-October to mid-March. The full model included one hour of academic instruction, a snack, and one hour of recreation time for three to four days a week. Students were provided with reduced class size (approximately 18 students per class) and were taught by a member of the regular teaching staff. Unlike Summer Bridge, schools were given wide flexibility in designing their Lighthouse programs and in deciding which grades and which students to serve. The program was not mandatory for students although schools must offer Lighthouse to their retained students. The district offered a Lighthouse curriculum to schools, but a 1999 to 2001 survey of Lighthouse programs found that in over three-quarters of schools, teachers were not using a specific lesson format, and in less than half of schools the Lighthouse curriculum was closely linked to instruction from the regular school day.⁴ Beginning in the 2001 to 2002 school year, Lighthouse was renamed After School Counts, a program that provided the same or greater levels of funding to schools but provided the opportunity for schools to partner with external agencies and gave schools wider flexibility in designing their programs, encouraging schools to move away from solely preparing students for the ITBS.

¹ Roderick et al. (2002).

² Roderick et al. (2002).

³ Chicago Public Schools (2001); Smith, Degener, and Roderick (2001).

⁴ Chicago Public Schools (2002).

may react to the pressure of high-stakes testing by becoming more, not less, disengaged from school.⁷

There has been little research on whether these purported benefits and negative effects of high-stakes testing actually occur, because few studies have looked at how high-stakes testing affects students' experiences and supports in school.⁸ In the final chapter of this report, we use longitudinal data from Consortium student surveys collected in 1994, 1997, 1999, and 2001 to trace trends in critical indices of students' experiences. We look at trends in sixth- and eighth-graders' reports of academic expectations and support from parents and teachers, participation in after-school programs and homework completion, and student reports of their engagement in and self-efficacy toward their schoolwork. We pay particular attention to differences in students' experiences by their prior levels of achievement. Thus, throughout this report, we examine the distributional consequences of high-stakes testing—that is, how the CPS policy might have differentially affected various groups of students and different groups of schools.

This report is unique in that we examine both the short- and long-term impacts of high-stakes testing on teacher behavior and student experiences. Prior research finds that test scores generally rise immediately after the institution of high-stakes accountability but often level off over time.⁹ Most research has examined teachers' instructional responses immediately after the institution of high-stakes testing. Yet, we know little about whether these effects are sustained. Do high-stakes testing policies produce only one-time impacts on behavior or are there persistent trends? Is there evidence that short-run impacts differ from long-term trends? For example, teachers may initially respond to high-stakes testing by spending more time on test-preparation activities, but they may shift their efforts as long-term instructional approaches are implemented and refined. The difference between short- and long-term effects may be particularly important in the Chicago Public Schools where the district has increased

the level of supports given to schools by expanding the Lighthouse after-school program.

In the 1995-1996 schoolyear, CPS's low-performing schools were placed on academic probation, and eighth graders were subject to the first wave of the ending social promotion initiative. Ending social promotion for third and sixth graders, however, did not begin until the 1996-1997 school year. In this report, we consider 1993-1994 as a prepolicy school year, and 1996-1997 as an additional baseline year, measuring the policy's immediate short-term impact. The 1999 surveys measure the impact of the policy after two years of implementation, and the 2001 surveys provide a long-term assessment, four years after the inception of the first set of policies.

Background on CPS's Accountability Policies

Student Accountability Initiative: Ending Social Promotion

The centerpiece of CPS's high-stakes testing program for students is a set of minimum test-score standards on the reading and mathematics sections of the Iowa Tests of Basic Skills (ITBS) for students in the third, sixth, and eighth grades.¹⁰ Students who do not meet the test-score cutoffs at the end of the school year are required to participate in a special summer program, called Summer Bridge, and retake the test in August. Those who again fail to meet minimum test scores are retained in their grade, or if they are 15 years or older, are sent to alternative schools, initially called Transition Centers, and later dubbed Academic Preparatory Centers. In the first several years under the policy, more than one-third of all third, sixth, and eighth graders failed to meet the promotion test-score cutoffs by the end of the school year.¹¹ In these years, CPS retained 20 percent of eligible third graders and approximately 10 percent of sixth- and eighth-grade students.¹²

The promotion test-score cutoffs were set using the grade-equivalent metric. Students are considered on grade level compared to national norms if, when tak-

ing the test in the eighth month of the school year, they obtain a score equivalent to their current grade plus eight months (e.g., 3.8 for the third grade). Between 1996 and 2000, CPS's promotion test-score cutoff for third graders was set at 2.8, one year below grade level. The sixth-grade cutoff was set at 5.3, 1.5 years below grade level. The eighth-grade cutoff was initially set at 7.0, 1.8 years below grade level.¹³

Each year, CPS increased the promotion grade requirements for eighth graders, and in 2000, increased the cutoff for sixth graders. Beginning in the 2000-2001 school year, moreover, the CPS policy was modified. The district began using a range around the promotion cutoff in order to make promotion decisions. And, teacher and principal recommendations as well as other indicators were incorporated in the promotion decisions. This report primarily focuses on the impact of CPS's policy prior to these 2000-2001 policy changes. Thus we cannot evaluate the extent to which increasing cutoffs and thus establishing even higher standards for students may change our findings.

The CPS policy provides two important supplemental support programs designed to help students meet the new standards—Summer Bridge and the Lighthouse after-school program. The Summer Bridge program provides six weeks of extra instruction in reading and mathematics for those students who fail to meet the test-score cutoffs in the spring (see *Summer Bridge and Lighthouse Programs* on pages 2-3). From the start, the Summer Bridge program was seen as the innovation that set CPS's efforts apart from previous unsuccessful efforts to end social promotion because it ensured that students who initially did not meet the cutoff would get extra help and a second chance.¹⁴ Beginning in 1996, CPS began providing a small number of low-performing schools with funds for an after-school program aimed specifically at such schools. Demand for this program and the perceived success of Summer Bridge led CPS to expand the Lighthouse program considerably so that by 2000, Lighthouse was offered in over 425 elementary schools. As it expanded,

Lighthouse became increasingly linked to CPS's promotion policy and focused on the objective of reducing summer-school attendance and reducing the number of students retained.

School Accountability Initiative: Academic Probation

In 1996, the first year of ending social promotion for eighth graders, CPS also initiated a new accountability program for low-achieving schools, outlined in the 1995 reform law that provided the mayor with control of the school board and top administrator appointments. Under this policy, schools with fewer than 15 percent of their students meeting national norms in reading (i.e., scoring at or above the 50th percentile on the ITBS reading-comprehension test) were placed on academic probation.¹⁵ In the first year of the policy, 71 of CPS's 475 elementary schools and 38 of the 66 high schools were placed on probation.¹⁶ Schools on probation were mandated to develop corrective action plans for improvement under the supervision of a probation manager assigned by the central office and were given extra resources for professional development. CPS contracted with a range of universities and non-profit organizations to provide this technical support.¹⁷ Schools that did not demonstrate improvement could be subject to reconstitution or other forms of operational intervention from the top, including the possible dismissal or reassignment of all school personnel. Between 1996 and 2000, approximately 23 of the more than 70 elementary schools initially placed on probation had raised their test scores to the criteria required for removal from the probation list (i.e., at least 20 percent of students scoring at or above national norms in reading). Reconstitution was never invoked for elementary schools. The district did move to reconstitute seven high schools (also intervening in five high schools in 2000); however, responses from the union led to a redesign of that approach, making the threat of reconstitution less credible. Thus, there was more positive, supportive attention than there were sanctions in the elementary schools. Because of the probation policy, the lowest-performing elementary schools had

even more initiative and interventions to increase reading test scores across all grades.

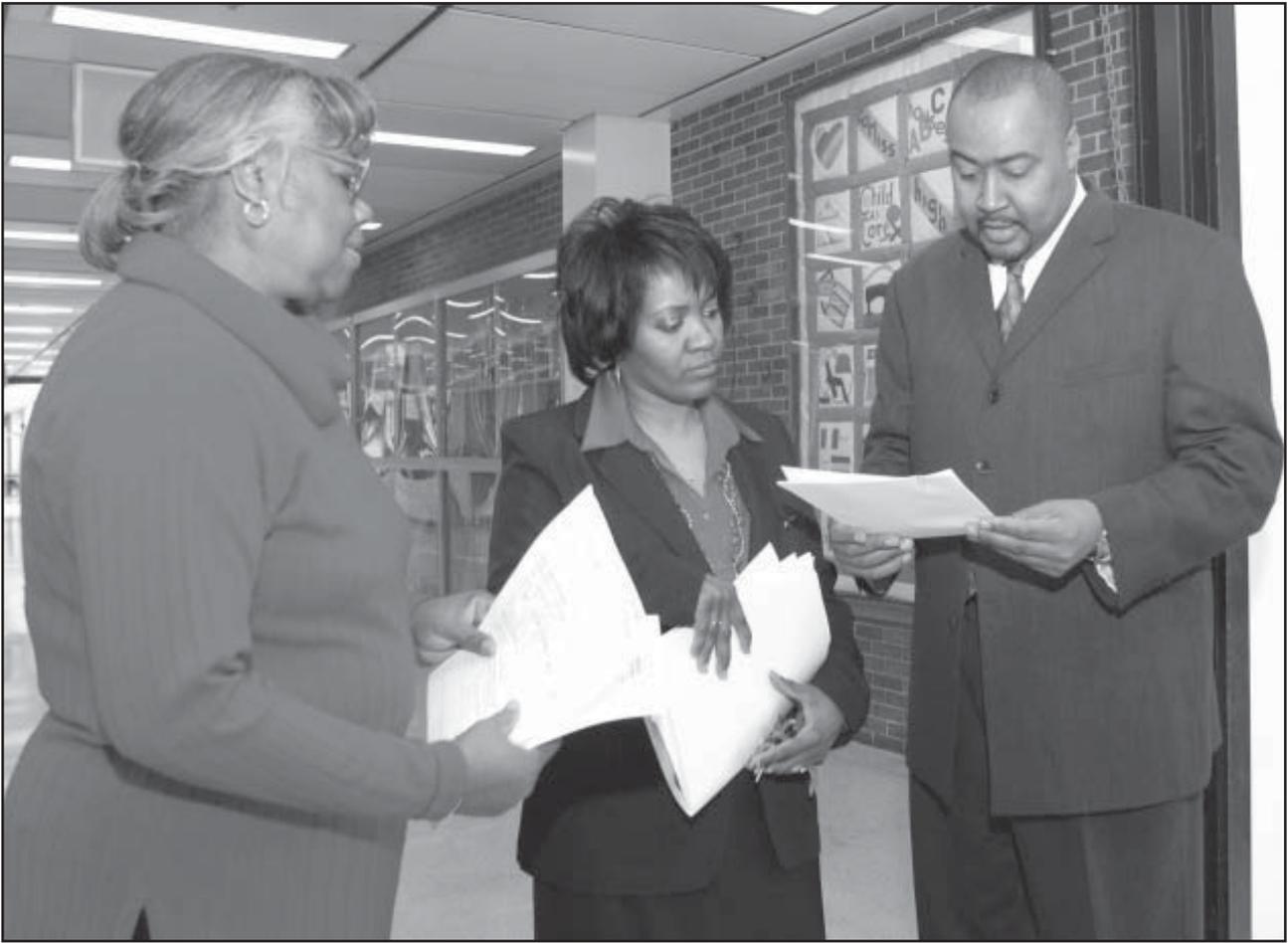
The Focus of This Report

This report is part of a larger multiyear evaluation of the effects of the Chicago Public Schools' efforts to end social promotion. In Chapter 1, we focus exclusively on teachers and principals' views of the ending social promotion initiative and the practice of grade retention. In Chapters 2 and 3, we use survey and interview data to look at changes in instructional practice and in students' experiences over time. Because CPS began both school and student accountability programs simultaneously, it is difficult to disentangle the effects of ending social promotion from the effects of the school-based accountability initiative conclusively, particularly because low-achieving students are concentrated in low-achieving schools. Thus, the results presented in Chapters 2 and 3 reflect the aggregate impact of all these policies on teacher behavior and student experiences. At some points in the report, we begin to examine the potential differential effects of school and student accountability by examining how trends differ across schools of different achievement levels and across students of different achievement levels, regardless of the school they attend.

It is also important in evaluating trends to recognize that high-stakes accountability was not the only change taking place in the Chicago Public Schools in the 1994-2001 period. There is evidence that test scores began rising in the early 1990s and that the first wave of decentralization reform was correlated with an increase in achievement in certain schools that may have laid the basis for continuing improvement after the 1996 reforms.¹⁸ The economy in Chicago was improving significantly throughout the 1990s and an influx of immigrants changed the racial and ethnic composi-

tion of the CPS student body. The 1995 reforms, in which the mayor gained substantial authority over the school system, also brought major investment in school infrastructure; changes in governance; and expansions of programs, including magnet schools, that may have led to a general improvement in test scores in all grades regardless of the highly touted accountability program. In this report, we were careful to focus on indicators that would be affected most by high-stakes accountability as opposed to these other general trends. Where appropriate, however, we estimated trends statistically adjusting for the incoming test scores of students in that grade in an effort to account for larger influences that may have shaped teacher practice and student experiences.

The goal of this report is to provide a careful and rigorous evaluation of the potential positive and negative influences of high-stakes testing on teacher behavior and students' experiences in school. We assemble the best evidence available, and we report on critical indicators of instruction and student supports for achievement that prior related research studies have employed.¹⁹ As a result, this report provides some insights into the variety of influences, both good and bad, that increased accountability can have on teachers and the classroom environment. Researchers and practitioners reading this report may think of other indicators that could be more useful in measuring the impact of high-stakes testing. We hope that this report will serve as an impetus for additional research on high-stakes testing both in Chicago and nationally. In the interpretative summary, we reflect on what we have learned, the generalizability of our findings to other contexts, and what issues our findings raise for future research.



John Booz

The Educators' Perspective on the Impact of Ending Social Promotion in Chicago and on the Practice of Retention

by

Robin Tepper Jacob and Melissa Roderick

The introduction of high-stakes accountability for schools and the end of social promotion for students in 1996, changed the context of teaching in the Chicago Public Schools (CPS). Teachers in low-performing schools faced substantial pressure to raise test scores, and all teachers were faced with the challenge of improving the achievement of their lowest-performing students. While most prior research finds that teachers resent accountability programs that either reward or sanction them for their students' performance, we know much less about how teachers view accountability policies directed primarily at students.¹ On the one hand, teachers could view district efforts like CPS's promotion policy as something that supports their own work in the classroom—helping to motivate students, sending the message that achievement matters, and ensuring that students have the basic skills they need before they advance to the next grade. On the other hand, teachers might believe that high-stakes testing for students is nothing more than high-stakes testing directed squarely at them: limiting autonomy in the classroom, placing excessive pressure on students and teachers, and undermining their professionalism by assuming that teachers' own judgments of students' performance are wrong or inadequate.

In this chapter, we look at how teachers assessed the impact of CPS's efforts to end social promotion and the extent to which they viewed the policy as consistent with their own views for learning. In both 1999 and 2001, the Consortium on Chicago School Research's biannual survey included a special supplement designed to explore how teachers and principals viewed and assessed the impact of the effort to end social promotion. In this chapter, we draw on these surveys and on teacher interviews conducted in low-performing schools in 1999 to examine four central questions:

1. How did teachers and principals assess the impact of ending social promotion on students,

parental involvement, and on their own instructional practice?

2. What do teachers and principals' responses tell us about the most positive aspects of the policy, and what areas raise concern for them?
3. How did teachers' assessments of the policy vary by their characteristics and by the racial and ethnic compositions and achievement levels of the schools in which they worked?
4. To what extent did teachers and principals change their assessments between 1999 and 2001 as the policy matured?

Data Used in This Chapter

This chapter draws on three sources of data: teacher surveys, principal surveys, and personal interviews with a sample of teachers in five low-performing schools. First, it uses survey data collected from teachers in both 1999 and 2001. The Consortium on Chicago School Research regularly surveys Chicago Public School (CPS) teachers about their behaviors and beliefs as classroom teachers. In the 1999 and 2001 surveys, a special supplement to the surveys contained 34 questions that asked teachers about their attitudes toward retention and the new end of social promotion policy as well as their assessments of the impact of those policies.

Of the elementary schools in the system, 80 percent were represented in the 1999 survey and 73 percent were represented in 2001. In 1999, 7,900 teachers, or approximately 47 percent of the 16,895 elementary school teachers in the system, responded to the survey. In 2001, the response rate increased slightly to 8,572 elementary school teachers. We conducted an analysis of the surveys to check whether the characteristics of schools and teachers who responded to the surveys were representative of the school system as a whole.

Among the school-level data, there was no evidence of response bias. The proportion of teachers in low-income, minority, and low-performing schools who responded to the survey was the same as for the school system as a whole. When comparing teacher demographic information reported on teacher surveys to that contained in the CPS personnel files, we found that the 1999 survey had fewer African-American respondents than the system as a whole (33 percent in the survey versus 41 percent systemwide). Survey respondents were also more highly educated than teachers in the system as a whole (54 percent reported a graduate degree or higher on the survey versus 31 percent in the system as a whole). There were no differences between the survey respondents and the system with respect to gender. Results were similar in the 1997 and 2001 surveys.

Principal survey responses constitute the second data source used in this analysis. A similar principal supplement was used in both 1999 and 2001. Approximately 315 of the 450 CPS elementary school principals responded to this survey in 1999, while 375 principals responded in 2001. Again, there was no evidence of response bias among the principals who responded to the survey. The proportion of principals in low-income, minority, and low-performing schools responding to the survey was the same as for the population as a whole.

Examining Teachers and Principals' Assessments of Ending Social Promotion

In both 1999 and 2001, the Consortium on Chicago School Research's biannual survey asked teachers and principals to assess the following aspects of Chicago's promotion policy:

- The efficacy of the programmatic supports (i.e., Lighthouse and Summer Bridge) in serving students with low skills;
- The influence the policy has had on student, parent, and teacher motivation;
- The influence of the policy on teachers' own instruction (see Table 1-1); and

- Whether they believed the policy was consistent with their own instructional philosophies and attitudes toward the practice of grade retention.

Principals were asked also about the degree to which efforts to raise students' test scores had shifted energy and resources away from other school and student needs.

Teachers and Principals' Assessments of the Impact of Programmatic Supports

Teachers and principals were very positive about the impact of Summer Bridge and Lighthouse. As noted in the introduction, Chicago's efforts to end

A third data source provided a supplement to the survey data and a context for teacher survey responses. Personal interviews were conducted with 43 teachers who taught in the promotion-gate grades (third, sixth, and eighth) in one of five K-8 schools. The five schools in which teacher interviews were conducted were located in two of the five Chicago neighborhoods with the highest retention rates. In these neighborhoods, one-third or more of the students subject to the promotion criteria were retained. These neighborhoods served either predominantly African-American or predominantly minority (Latino/African-American) students. Of the two neighborhoods selected for the study, one was predominantly African-American and the other was mixed Latino and African-American. There were a total of 50 third-, sixth-, and eighth-grade teachers in these five schools. The final interview sample consisted of 43 teachers (26 percent male, 35 percent African-American, 28 percent Latino, and 37 percent white), representing 86 percent of teachers in the promotion-gate grades in these five schools.

In the interviews, teachers were asked to talk about their attitudes toward CPS's efforts to end social promotion and to comment on both the negative and positive aspects of the reform initiative. Interviewers also asked teachers about the degree to which the policy had influenced their teaching practices and about the pressure they felt as a result of the promotion policy. Finally, teachers were asked to comment on their experiences working with students who had been retained as a result of the new promotion policy and to reflect on the effectiveness of the program supports that had accompanied the policy, such as the Summer Bridge summer-school program and the Lighthouse after-school program.

The reform environment in which the effort to end social promotion was implemented included a number of initiatives designed to increase educators' accountability. In addition to retaining students who failed to meet minimum Iowa Tests of Basic Skills (ITBS) test-score requirements, CPS began in 1996 to place schools with fewer than 15 percent of their students performing at or above national ITBS norms on an academic warning list (academic probation) and threatened to reconstitute schools that failed to show improvement. While interview questions asked specifically about teachers' beliefs and practices regarding the end of social promotion, in some instances, the teachers' responses reflected the general increase in accountability.

Table 1-1: Survey Questions: Assessing the Impact of Ending Social Promotion 1999 and 2001

The 1999 and 2001 Consortium surveys asked teachers and principals to indicate how much they agreed with the following statements about the CPS ending social promotion policy:

How much do you agree ...	
About the impact of programmatic supports	<p>Overall the Summer Bridge program has had a positive effect on the students who have used it.</p> <p>Overall the Lighthouse program has had a positive effect on the students who have used it.</p> <p>I feel supported in my efforts to help students who are at risk of being retained.</p>
About the impact on parent and student motivation	<p>The threat of retention motivates students to work harder in school.</p> <p>[This policy] has made parents more concerned about their students' progress.</p>
About effects on instruction and their own behavior	<p>[As a result of the new policy] nearly all teachers in this school feel extra responsibility to help students meet standards.</p> <p>[This policy] has made me more sensitive to individual student's needs and problems.</p> <p>[This policy] has focused the school's instructional efforts in positive ways.</p>
About consistency with their own views	<p>[This policy] is consistent with my own views about what is best for student learning.</p> <p>[This policy] limits my attention to higher-order thinking skills.</p> <p>[This policy] places too much emphasis on basic skills.</p> <p>[As a result of the new policy] I spend less time on social studies and science than I used to.</p>

Principals were asked to indicate to what extent they agreed or disagreed with the following statements about the impact of the policy on school resource allocation.

The policy is consistent with the major goals of my School Improvement Plan.

Because we are so concerned with test scores, we are neglecting other important student educational needs.

The promotion policy emphasizes short-term fixes at the expense of carefully planned strategies to address student learning problems.

[The promotion policy] has diverted resources to the third, sixth, and eighth grades at the expense of other grades.

social promotion were accompanied by significant investments in both after-school and summer supports, including the Lighthouse after-school program for low-achieving students and the Summer Bridge summer-school program for students who failed to meet the test-score cutoff at the end of the school year (see *Summer Bridge and Lighthouse Programs* on pages 2-3). Survey results suggest that teachers were quite positive about the influence that these programs had on students (see Figure 1-1). Over 80 percent of teachers and 90 percent of principals in the 1999 surveys agreed or strongly agreed that Summer Bridge and Lighthouse had a positive impact on the students who attended. In interviews with educators in low-performing schools, teachers stressed that Summer Bridge was critical because it gave students a second chance to meet the promotion requirement. Many teachers in these schools felt the policy would be unfair without this second chance for students to improve their test scores.

In the 1999 surveys, teachers reported that they relied heavily on these programs to assist them in working with struggling students. When these surveyed teachers were asked what they were doing to help students who were at risk of being retained or had already been retained, over 90 percent indicated that they referred students to after-school or

other tutoring programs.² Perhaps reflecting their positive assessment of these programs, more than 75 percent of teachers and almost 90 percent of principals agreed in 1999 that they felt supported in their efforts to help students who had been retained and who were at risk of retention.

Teachers and Principals' Assessments of the Impact the Policy Had on Student, Parent, and Teacher Motivation

Teachers and principals were positive about the influence the policy had on student motivation. In the 1999 surveys, 67 percent of teachers and 72 percent of principals agreed or strongly agreed that the threat of retention motivated students to work harder in school (see Figure 1-2).

The majority of the teachers in low-performing schools who were interviewed echoed these sentiments and reported that students had been working harder in response to the threat of retention:

Because the students know, especially in eighth grade, you know we need the score or we're coming to (this school) next year. Especially since I let them know, I'm not the one failing you, I'm not the one holding you back. And so then the students are much more motivated to do their work, especially as the year progresses. (Eighth-grade teacher)

Figure 1-1

Most Teachers and Principals Were Positive about Lighthouse and Summer Bridge (1999)

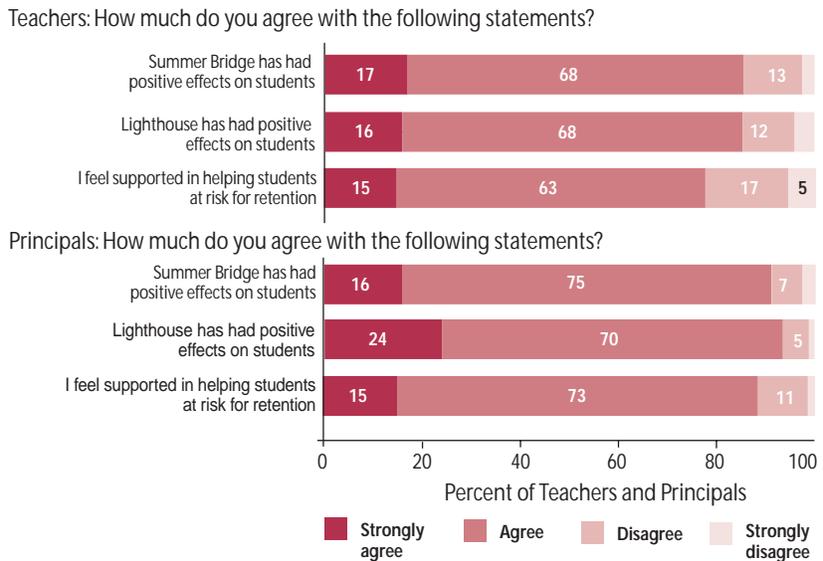
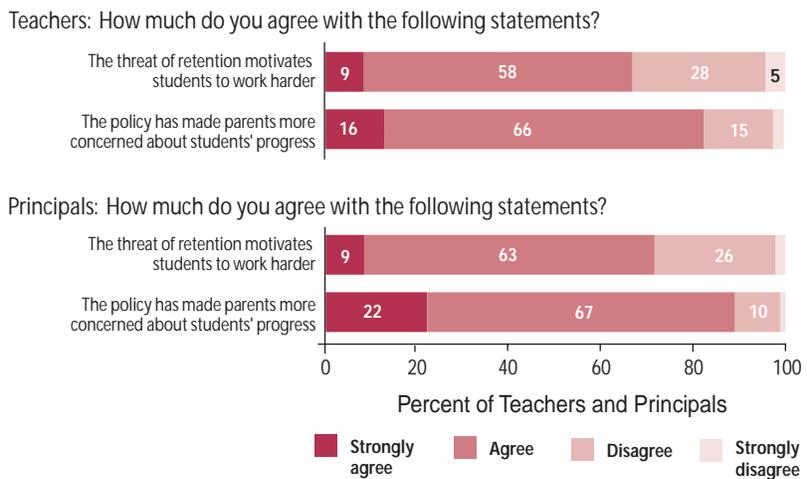


Figure 1-2

Most Teachers and Principals Felt That Students and Parents Were Positively Motivated by Ending Social Promotion (1999)



We have pupils who have become students. That is, they actually do some studying because they have a goal now, whereas before they knew that they were going to be pushed on no matter what they did. (Eighth-grade teacher)

Both survey and interview results suggest that teachers appreciated the policy's strong message to students—that their classroom behavior and performance matter. As one teacher commented, the policy sends students the message that, "I [the student] make decisions and they have consequences, and I have a great deal to do with the outcomes." Many interviewed teachers also reported that they used the policy as a tool to motivate students, by helping students to focus on a goal. One teacher described using a growth chart based on test scores as a way to help students monitor their own progress. Another commented:

As a result of the policy, when kids come to the sixth grade, they feel responsible because they have something; they know they have a set goal that is in front of them. I think all children need that. (Sixth-grade teacher)

Teachers and principals felt that the policy had positive effects on parental involvement. Almost 90 percent of principals and 75 percent of teachers surveyed agreed or strongly agreed that the policy had made parents more concerned about their child's progress. One teacher in our interviews felt that the promotion policy "laid down the law" for parents, changing the expectations that merely sending a child to school automatically earned that child a diploma. Another teacher noted:

I really feel that there are three people involved . . . the teacher, the child, and the parent. All of them. If they're not following with it [working with each other] . . . it makes it difficult for the others . . . accountability is a positive part of this policy. . . . I do like that people are responsible all the time. (Third-grade teacher)

Ending Social Promotion: Teachers and Principals' Assessments of the Policy's Effect on Their Own Behavior and Instruction

Teachers and principals in 1999 felt that the policy made them more sensitive and responsive to students' needs. What teacher would not be happy with more highly motivated students, greater parental involvement in education, and additional program supports to help struggling students? Indeed, some have suggested that ending social promotion in Chicago has transferred responsibility for poor student performance to the students, parents, and after-school and summer programs, in effect, taking the burden off of teachers. Such an interpretation would suggest that teachers favored the policy because they did not have to change their own behavior. However, survey results do not support this hypothesis. In the 1999 surveys, the majority of CPS teachers and principals reported that they believed the policy positively influenced their behavior and their school's instructional efforts.

In the 1999 surveys, 85 percent of teachers and almost 90 percent of principals agreed that as a result of the policy "nearly all teachers [in this school] feel extra responsibility to help students meet standards." (See Figure 1-3.) More than 80 percent of teachers and principals agreed that the policy had made them more sensitive to individual student's needs and problems. Teachers and principals also agreed that the policy has affected instruction positively. More than 80 percent of teachers and 87 percent of principals surveyed agreed or strongly agreed that the promotion standards had "focused the school's instructional efforts in positive ways."

Our interviews with teachers in low-performing schools confirmed this generally favorable assessment of the impact of the end of social promotion on teacher behavior and classroom instruction. First, interviewed teachers recognized a change in their sensitivity to student learning. One teacher noted that as a result of the policy she will not move on with a lesson until she is sure that 95 percent of the class has mastered the skill. Another described the change in her behavior:

[The policy] . . . made me more accountable. It has . . . kept me on my toes the entire year. There's not one day gone by that I haven't thought about what they need to know and that if they don't pick up these skills they will not pass. And I feel it is my responsibility to get them to pass math.

Many teachers also felt that the promotion policy established clear goals for instruction and, as one teacher explained, helped them “teach smarter.” Over half (56 percent) of the teachers we interviewed said they appreciated that the policy established a fixed, concrete standard for achievement. As one teacher noted:

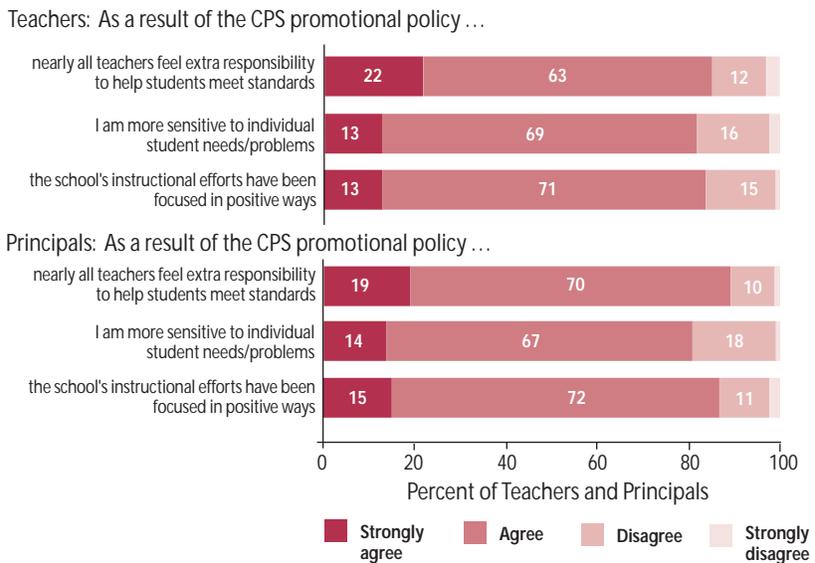
This system I like because there is organization to it. Before, it was like, well, you're starting at A, you need to get to B, but you are kind of on your own to get there. This way there's a . . . structure to work with. And I feel comfortable with that. (Eighth-grade teacher)

Teachers and Principals' Views on the Consistency of the Policy with Their Own Educational Beliefs

In 1999, teachers and principals agreed that the policy was consistent with their beliefs about what is best for student learning.

Figure 1-3

Most Teachers and Principals Reported That the Ending Social Promotion Policy Made Them More Sensitive to Students' Needs (1999)



These survey and interview results suggest that many CPS educators endorsed the basic approach and goals of the policy, including its emphasis on basic reading and mathematics skills. In 1999, three-quarters of teachers (77 percent) agreed that the policy was consistent with their own views about what is best for student learning, while only 23 percent of teachers surveyed agreed that the policy placed too much emphasis on basic skills (see Figure 1-4), and 31 percent agreed that the policy had limited their attention to higher-order thinking skills.

Most interviewed teachers also expressed concern that students were missing the basic skills they needed to succeed and believed that the development of basic skills was the most important task to be accomplished in the classroom. As one teacher explained, kids “can never get enough of the basics; basics are the foundation.” Others commented:

Some of my students don't know multiplication tables, and we're doing integers and fractions. How are they going to comprehend the whole thing about fractions if they don't know how to multiply them? How will you divide if you don't know how to make times tables? It's impossible. So they have to know it. They really have to know. (Sixth-grade teacher)

I think the teachers understand the students are missing basic skills. . . . Our kids are missing so many basic skills . . . basics, period, in everything. (Eighth-grade teacher)

Many teachers also noted that basic skills have always been a top priority: “I don’t think that the promotion policy really dictates [an emphasis on basic skills] because I think we have been doing [that] right along. If you’re doing your job, you are teaching basic skills. . . .”

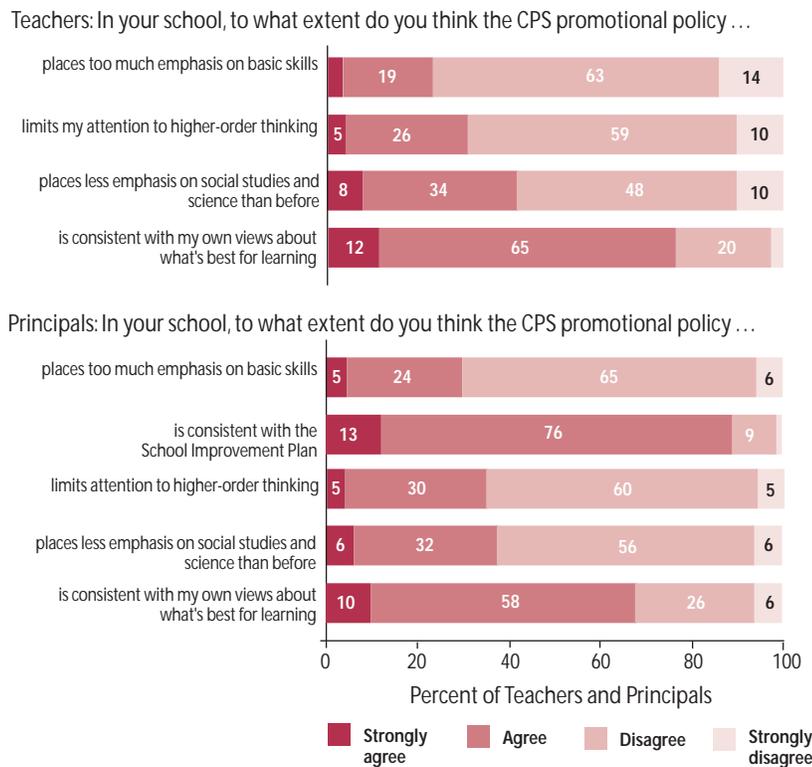
Principals also valued the emphasis the policy placed on basic skills (only 29 percent believed the policy places too much emphasis on the basics) and nearly 90 percent of principals agreed that the policy was consistent with the goals of their school improvement plan. Principals were less likely than teachers (68 percent versus 76 percent) to agree that the policy was consistent with their views of what is best for student learning.

Teachers and principals did report that they spent less time on other subjects. Approximately 40 percent of teachers and principals agreed that, as a result of the promotion policy, they were spending less time on social studies and science than they used to. In-

“[Reading is] the most important thing that they’re going to get out of here. . . . It’s got to be the most important. They can’t function if they can’t read. . . . To me, all the other things should be put aside until they can do this. (Eighth-grade teacher)”

interviews with educators in low-performing schools, teachers expressed varying opinions about whether this shift in emphasis was good. Some teachers believed that reading and mathematics should always be given greater emphasis than social studies and science:

Figure 1-4
Most Teachers and Principals Felt That the Ending Social Promotion Policy Did Not Place Too Much Emphasis on Basic Skills (1999)



To me, if you cannot read then how can you write? How can you do your science? How can you do your social studies? How can you do your math? So, reading skills are most important, then math skills come in second. We have yet to come across a job that does not require any type of math skill. . . . And math is involved even in your social studies. You’ve got time lines and years and latitude, longitude—all that stuff. . . . (Eighth-grade teacher)

[Reading is] the most important thing that they’re going to get out of hereIt’s got to be the most important. They can’t function if they can’t read. . . . To me, all the other things should be put aside until they can do this. (Eighth-grade teacher)

Many teachers reported that even in the absence of the promotion policy, they would place more emphasis on reading and mathematics than on social studies and science: “I think most teachers, regardless of the policy . . . focus more on reading and math,” said one third-grade teacher. Another eighth-grade teacher agreed: “These are the areas that have always been thought of as important anyway, so we’ve always spent more time on reading and math . . . at the expense of writing and social studies, etc. . . .”

However, between one-quarter and one-third of the teachers with whom we spoke expressed concern about the instructional cost of de-emphasizing other content areas.

I think especially at third-grade level when they should really be getting into more social studies, deeper thoughts, higher-level thinking, it’s really hard to spend the time doing it because you’re spending all this time in reading and math and writing . . . you have to have strong reading and math skills . . . because they help you in all the subject areas, but . . . there’s too much emphasis placed on those skills because . . . we just don’t have enough time to do all the other things. I really think we need to spend more time on writing, which we don’t get. (Third-grade teacher)

Some teachers also indicated that standards and testing had constrained their behaviors in the classroom in other ways. Teachers expressed frustration that the basic skills and test emphasis took away from time to develop “soft skills.” Additionally, some teachers felt less able to employ creative instructional methods.

There’s tremendous pressure in our school to do well. . . . There’s tremendous pressure on the teachers to teach [to] the test. And, of course, when that happens you’re going to leave off

more important areas—like learning skills for life. (Sixth-grade teacher)

[The policy] takes away from more—I want to say creative or more general activities. [If not for the policy] I would do things that I enjoy more, academic games and activities like that . . . more creative things. . . . Instead of more cut-and-dried reading, math, drills, practices. (Sixth-grade teacher)

I do feel that . . . too much emphasis is placed on the standardized test. You may just want to plan a whole theme of different things [that] you may not get a chance to do because you want to make sure they have those [basic] skills. . . . A lot of creative things that you want to do you may not. (Third-grade teacher)

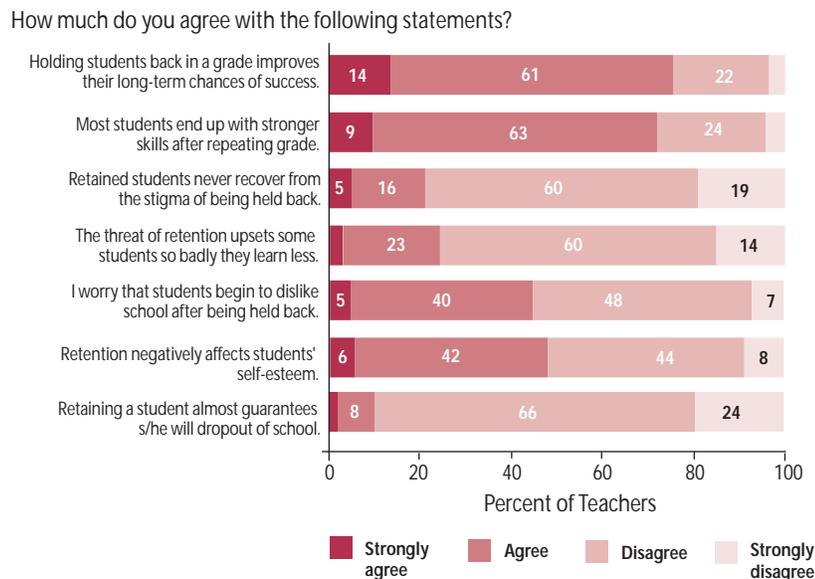
However, despite their reservations, cautions, and concerns, this analysis suggests that the majority of teachers and principals supported and endorsed the general educational goals of the policy. The goals, including an increased emphasis on basic reading and mathematics skills, were consistent with their educational philosophy. Most teachers felt that increased accountability was a necessary and positive step towards improving the educational experience for students, even if it was not the solution to all problems faced in the classroom.



John Booz

Figure 1-5

Most CPS Teachers Surveyed Viewed Retention as Academically Beneficial (1999)



Examining Chicago Teachers and Principals' Attitudes toward the Practice of Retention

Teachers Believe the Practice of Retention Is Educationally Beneficial for Students

Thus far, we have examined CPS teachers and principals' views on ending social promotion, its program supports, and its impacts on their own behavior and on student and parent motivation. The most salient concern expressed by opponents of high-stakes testing is that even if the policy motivates students to work harder in school and helps teachers to focus instruction in positive ways, it does so by using a practice—grade retention—that hurts children.

Whether retention produces short-term effects on achievement is hotly debated in the literature. Some studies conclude that retained students perform significantly more poorly than comparison groups of promoted students when their achievement is compared one year after.³ Others conclude that retention may provide a short-term boost for low-achieving students that will benefit them when they reach postretention grades.⁴ There is, however, little evidence that grade retention leads to long-term improvements in student achievement even among studies that document short-term positive boosts in performance.⁵ Some research suggests that retained students suffer negative academic self-esteem as a result of the retention

experience, although the results are inconclusive.⁶ Most importantly, research consistently finds that retained students are more likely to drop out of school.⁷

Despite these findings, several other studies have found that teachers often do not view retention negatively.⁸ Chicago teachers appear to share this view (see Figure 1-5). The 1999 surveys asked teachers and principals a series of questions about the effects of retention. Consistent with the national studies, three-quarters of Chicago teachers felt that retention improved students' long-term chances of success.

These survey responses also indicate that a majority of teachers believe that retention does not have long-term negative effects on students. However, teachers expressed greater concern about the immediate impact of retention on students' self-concept and engagement in school. Forty-five percent of teachers surveyed in 1999 worried that "students begin to dislike school after being held back," and 48 percent agreed that, "retention negatively affects students' self-esteem."

Principals expressed somewhat greater concern about the practice of retention than did teachers. Only two-thirds of principals surveyed in 1999 agreed that retention improves students' long-term chances of success, and 64 percent of principals (as opposed to 72 percent of teachers) believe that students end up with stronger

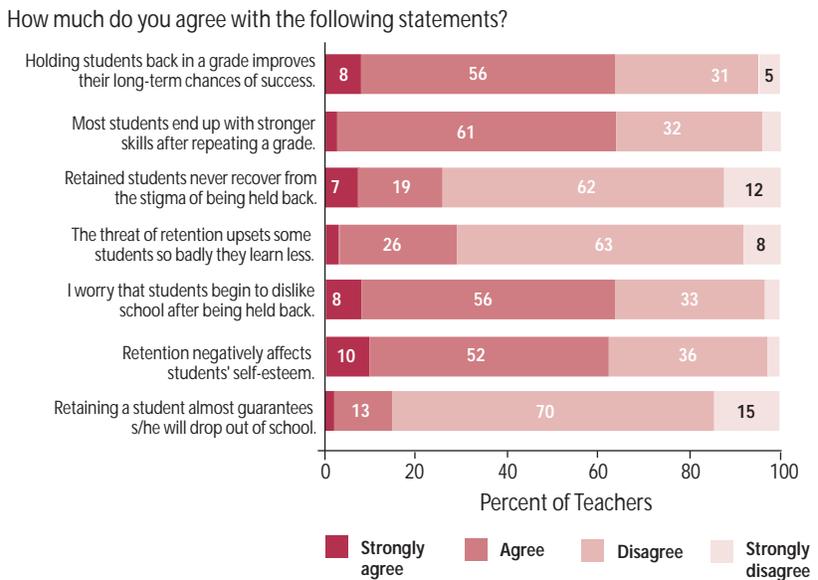
skills after repeating a grade (see Figure 1-6). Principals also expressed greater concern about the immediate impacts of retention on students. Sixty-four percent of principals versus 45 percent of teachers worried that “students begin to dislike school after being held back.” Principals were also more likely to agree that “retention negatively affects students’ self-esteem.”

While survey results suggest that teachers had few concerns about the impact of retention, the teachers in low-performing schools we interviewed were somewhat more ambivalent. Many interviewed teachers argued that the benefits of retaining a child depended heavily on the characteristics of the child and the circumstances surrounding the retention decision. Other teachers, especially those who had experience working with retained students, expressed serious reservations about the impact that grade retention could have on a child’s long-term chances of success. One teacher said she felt that the impact of retention was “devastating across the board.” Another believed that retention generally “makes students give up” and “go to the streets.” Others noted:

I think that kids feel it’s humiliating and embarrassing to be held back if the rest of the kids know that they’re a year older than everybody else. . . . And I

Figure 1-6

Principals Viewed Retention as Somewhat Less Academically Beneficial than Teachers (1999)



don’t think they’re learning that much more the second time around. (Sixth-grade teacher)

When this came in to play, no social promotion, I was kind of like, “oh, good.” But in a way, we have really old sixth graders. They’re at an adolescent level where the girl-boy relations, it’s just incredible. And then they don’t like the books we use. I try to pick out things I think they’ll enjoy . . . but . . . it’s real hard to find a reading level that they’ll like that will also have a story that interests them. (Sixth-grade teacher)

Often in interviews, however, teachers’ concerns about retention were not as great as their concerns about the detrimental effects of social promotion. Several teachers were explicit about this tradeoff:

I remember in the past I was sending students with [very low skills to high school] . . . you knew they were going into high school and they were not going to make it. Now I know they are going there with some more knowledge, and they are more prepared, and stopping them from dropping out because they can’t handle the work and the pressure. (Eighth-grade teacher)

The kids feel bad . . . [but] the way it [retention] makes a difference is that they are able to compete when they go to high school, when

they get out. If you send somebody out like that, with a 3.3, how's he going to compete out there? Something had to be done. You know, sometimes medicine tastes bad, but you gotta do it. (Eighth-grade teacher)

These teachers seemed to understand the costs associated with retaining a child but also saw the practice of socially promoting students without grade-appropriate skills as extremely detrimental. While acknowledging the tensions and trade-offs between retention and social promotion, many teachers expressed the hope that retention would pay off in the long run.

Teachers and Principals' Reservations about High-Stakes Testing

The previous discussion highlighted the general support that teachers and principals gave to the district's efforts to end social promotion. Yet educators also expressed some reservations about the policy. Interviewed teachers in the lowest-performing schools expressed three major concerns about the policy:

- The effort to end social promotion places too much emphasis on a single test score.
- The standards create too much stress and pressure for teachers.
- The policy does not do enough to address the needs of the lowest-performing students.

Teachers believe Chicago's effort to end social promotion places too much emphasis on a single standardized test score. In 1999, when we conducted our teacher interviews, Chicago's promotion policy was based solely on whether students met a given cutoff on the ITBS in both reading and mathematics. As described in the introduction, beginning in the 2000 to 2001 school year, the promotion policy was amended in two ways. First, CPS began using a range around the promotion cutoff in order to make promotion decisions. Second, teacher and principal rec-

ommendations as well as other indicators of student performance were incorporated in the promotion decision. These changes appeared to be supported by Chicago educators. In our 1999 interviews, many Chicago teachers agreed with national experts that the use of a single standardized test score was inappropriate. Teachers expressed a general mistrust of ITBS test scores as applied to the individual student and believed that good scores could often be obtained through "luck" or "guess work." Others felt the test emphasized speed over knowledge. As a result, a majority of the teachers felt strongly that teacher input should be used in the decision to promote or retain students.

Well, I think the whole policy places too much emphasis on taking that particular test on that particular day. . . . A child might be sick that day and not do well, and it doesn't show his true performance. Plus, some of the bright kids get overly nervous. It's so hard to get a balance between telling them that this is important and then going too far and getting them upset about it. (Third-grade teacher)

Teachers differed about whether the increased accountability associated with the policy placed too much stress and pressure on teachers. Many interviewed teachers made reference to the increased stress and pressure that accompanied the implementation of ending social promotion. As noted earlier, over 90 percent of surveyed teachers agreed that "nearly all teachers feel extra responsibility to help students meet standards" as a result of the promotion policy. Teachers, however, appeared split in their assessment of whether this increased sense of accountability was positive. Almost 20 percent of the teachers in low-performing schools who we interviewed experienced this pressure in positive ways. As one teacher noted, pressure is a normal part of professional life: "If you are out in the business world and you were selling or managing, you would have expectations and goals, you know—management by objective. Why should it be any different for a teacher?" Another commented:

Everyone needs to be accountable for something. These children need something. Some people might not be here with children first in mind. And you have to have children first in mind. This is for the children, so you have to be accountable. You have to have some type of standard. What did you do? How can you prove you did something? And right now the standardized test is what we have. (Sixth-grade teacher)

However, one-third of the interviewed teachers indicated that the task of raising the achievement scores of at-risk children, some two, three, or four years behind grade level, became more of a stressor than a motivator. Some teachers were concerned that there would be negative personal consequences for failing to get their students up to the test cutoffs. But frequently teachers' concerns focused more on school accountability than on student accountability, confirming the findings of previous research that teachers resent being judged by single test scores. One teacher noted that she "worries a lot" about how kids are "going to do," because "they publish your scores in the paper. . . . Your success or failure seems dependent on these, just these test scores and nothing else they've learned. . . . And you know the public views these scores as . . . whether the school is good or not." Others worried about losing their jobs, being put on probation, or being publicly reprimanded by the principal.

Teachers do not believe enough is being done to help the lowest-achieving students in the system or to address students' other educational needs. Principals feel that the policy diverts resources away from long-term solutions and other student needs. Interviewed teachers, who came from low-performing schools in the district, expressed concern that not enough was being done to address the needs of the lowest-performing students. While the policy provides considerable resources for students who are at risk of failing to meet the promotion cutoff, in the form of the Lighthouse after-school program

and the Summer Bridge program, there are few additional resources or interventions for students who fail to meet the cutoff at the end of summer school. One teacher noted that the policy would be fairer if there were "more intensive services for the students that do fail." Others commented that "the real issues are not being tackled":

Let's talk about the totality of the child, let's tackle the whole family, social conditions, health problems. . . . Does this child eat? Does this child need eyeglasses? Does this child have capability deficits versus skills deficits? (Sixth-grade teacher)

And the policy, I really agree with it for some children. For others, I think we need to get some experts in to tell us how to handle the others who are not going to respond. (Third-grade teacher)

Surveyed principals expressed similar concerns about the potential tradeoffs of the policy. Sixty-three percent of surveyed principals believed that the policy had caused the school to neglect students' other educational needs (see Figure 1-7). Forty-two percent of principals also believed that the policy emphasized short-term fixes over long-term solutions, and 39 percent agreed that the policy had made them divert resources from other grades to the third, sixth, and eighth grades.⁹ Both teachers and principals appeared to be concerned

Figure 1-7

Principals were Ambivalent about the Promotion Policy's Total Impact on Students' Educational Needs (1999)

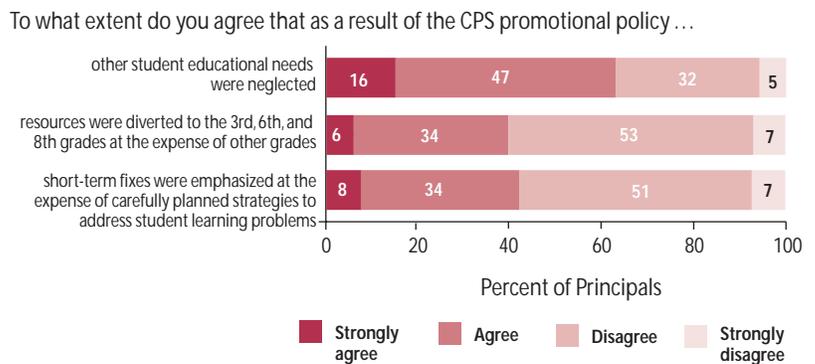


Figure 1-8

Most Teachers Were Positive about the Impact of Ending Social Promotion (1999)

Very positive: Teachers were likely to agree or strongly agree to all questions about the policy.

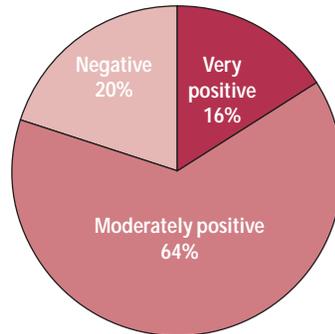
Moderately positive: Teachers were likely to agree that the policy had positive effects on

- program supports;
- student, teacher, and parent motivation; and
- instruction.

Teachers were less likely to agree that the policy

- was consistent with their views and
- supported their efforts to help students.

Negative: Teachers did not agree that there were positive impacts on instruction, program supports, or student and parent motivation.



Note: This measure combines teachers' responses to individual survey items about their views of the policy to end social promotion. The measure explores the overall degree of support each teacher has for the policy. A higher score indicates greater support for the policy. This was measured on a 10-point scale.

that the policies and supports associated with the end of social promotion in Chicago focus more on short-term solutions and fail to address the longer-term problems faced by many students.

Summarizing Teachers and Principals' Views

In 1999, most teachers were moderately positive about the effect of ending social promotion. Our look at individual items suggests that three years after the initial implementation of promotion standards, the majority of educators had positively assessed the impact of the policy on students, parents, and their own behavior (See Figure 1-8). This assessment appears to be based, in part, on the belief that students need a basic-skills education and that the detrimental effects of socially promoting students outweigh the potentially negative effects of retention.

In order to summarize teachers and principals' responses, we created a summary measure that combined teachers and principals' survey answers to a variety of the individual items. The items included in the measure asked teachers to assess:

- The impact of Lighthouse and Summer Bridge and the extent to which they felt supported in their efforts to meet students' needs,
- The impact of the policy on student and parent motivation,
- The effects of the policy on instruction and their own behavior, and
- The degree to which the policy was consistent with their own views.¹⁰

When individual teachers' responses to these items are taken together, approximately 16 percent of CPS elementary school teachers in 1999 could be considered "very positive" about the impact of the ending social promotion policy (see Figure 1-8). The majority, 64 percent, could be characterized as "moderately positive," and approximately 20 percent of teachers' responses suggested that they had "negative" assessments.

Teachers who were characterized as moderately positive endorsed the impact of program supports; believed that the policy had had positive effects on student, teacher, and parent motivation; and believed the policy had focused instruction in positive ways. Those teachers, however, were less likely to agree that the policy was consistent with their own views about what is best for student learning.

Our summary of survey responses suggests that principals were more moderate in their assessment of the impact of ending social promotion than teachers (see Figure 1-9).¹¹ While 81 percent of principals were moderately positive about the impact of ending social promotion, only 8 percent were very positive. At the same time, only 11 percent of principals versus 20 percent of teachers could be classified as having a negative assessment of the policy. This summary measure

confirms what we observed in our exploration of the individual survey items.

In 1999, Chicago principals appeared to have more reservations about the instructional costs associated with the policy. Although principals were equally positive or more positive than teachers about the effects on motivation and program supports, they were, on average, less likely than teachers to feel that the policy was consistent with their own views and believed the policy had instructional tradeoffs in that it limited teachers' attention to higher-order thinking skills. In addition, principals reported that the policy diverted resources away from long-term solutions and more general student problems.

Teachers Differed in Their Assessments of the Impact of Ending Social Promotion and the Effects of Retention

Our summary measure suggests that teachers differed in their overall assessment of the impact of the ending social promotion policy. Did teachers' opinions differ because they held different views on education or because they served students who were differentially impacted by the policy? Teachers in promotion-gate grades (third, sixth, and eighth) were more affected by the policy than teachers whose students were not in grades targeted

Figure 1-9

Most Principals Were Moderately Positive about the Impact of Ending Social Promotion (1999)

Very positive: Principals were likely to agree or strongly agree to all questions about ending social promotion.

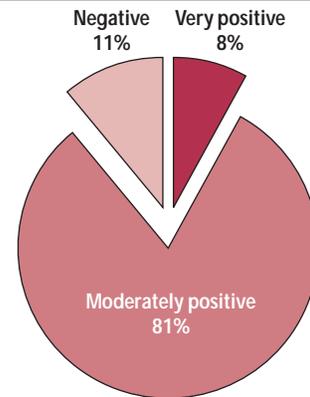
Moderately positive: Principals were likely to agree that the policy had positive effects on

- program supports;
- student, teacher, and parent motivation; and
- instruction.

Principals were less likely to agree that the policy

- was consistent with their views and
- supported their efforts to help students.

Negative: Principals did not agree that ending social promotion impacted instruction or student and parent motivation positively. However, these principals often still agreed that program supports were positive.



for high-stakes testing. Similarly, teachers in low-performing schools faced significant pressure to raise test scores. On the one hand, these teachers may have resented the pressures imposed by high-stakes testing. On the other hand, teachers in these grades and schools may have benefited most from the increased resources and motivational effects associated with the policy.

To explore these possibilities, we investigated how teachers' assessments of the impact of the promotion policy and their assessment of the effects of retention varied by the characteristics of teachers and the schools in which they taught. More specifically, we explored whether teachers' attitudes toward ending social promotion differed by:

- Their own years of experience, education, and demographic characteristics;
- The grade and subjects they taught;
- The achievement level and racial composition of the school in which they taught; and
- The experience of their school under the policy as measured by the proportion of students who were retained.

We conducted our analysis using a multivariate procedure that examined the effect of each of these teacher and school characteristics independently. For example, we were able to examine how experienced teachers felt about the impact of the policy after accounting for the fact

Table 1-2: Which Teachers Were the Most Positive about the Impact of Ending Social Promotion (1999)

	Less Positive	More Positive
Education	Teachers with more than a bachelor's degree	Teachers with a bachelor's degree
Teaching experience	Teachers with less than 15 years of experience	Teachers with more than 15 years of experience
Race/ethnicity	White and other race teachers	Latino and African-American teachers
Grade	No differences by grade	No differences by grade
Subject	<ul style="list-style-type: none"> Teachers in self-contained classrooms Teachers in the content areas (math, science, social studies, and language arts) 	Auxiliary teachers (teachers who do not teach in content areas)

Note: Levels of concern were estimated using a multivariate analysis that calculated the impact of each teacher's response within the framework of other characteristics (e.g., how teachers with bachelor's degrees felt accounting for the fact that older teachers tended to be more likely to have only bachelor's degrees). Additional controls included the race and achievement level of the school.

Table 1-3: Which Teachers Expressed Greater Concern about the Impact of Retention on Students (1999)

	Fewer Concerns	Greater Concerns
Education	Teachers with a bachelor's degree	Teachers with more than a bachelor's degree
Teaching experience	Teachers with more than 15 years of experience	Teachers with less than 15 years of experience
Race/ethnicity	African-American and other race teachers	Latino and white teachers
Grade	Teachers in non-promotional gate grades	Third, sixth, and eighth grade teachers (promotional gate grades)
Subject	No difference by subjects	No difference by subjects

Note: Levels of concern were estimated using a multivariate analysis that calculated the impact of each teacher's response within the framework of other characteristics (e.g., how teachers with bachelor's degrees felt accounting for the fact that older teachers tended to be more likely to have only bachelor's degrees). Additional controls included the race and achievement level of the school.

that experienced teachers teach in different schools, have obtained differing levels of education, and differ in their race and ethnicity (see Appendix A for a full description of the model that was estimated).

In 1999, Experienced Teachers Were More Supportive of Reform, and Highly Educated Teachers Were Less Supportive of Reform

Chicago Public School teachers are, on average, an older teaching force and have obtained a limited number of advanced degrees. In the 1999 teacher survey, for example, over half (54 percent) of teachers reported having more than 15 years of teaching experience, while 21 percent reported having less than five years of classroom experience.

That same year, 42 percent of surveyed teachers reported that they had received only a bachelor's degree.¹² Teachers with only a bachelor's degree were much more likely to be positive about the impact of ending social promotion and expressed fewer concerns about the potential negative impacts of retention (see Tables 1-2 and 1-3). Moreover, the most experienced teachers in the system (teachers with over 15 years of teaching experience) were significantly more positive, on average, about the effects of CPS's efforts to end social promotion and expressed significantly fewer concerns about the impact of retention on students.

Teachers' views on retention and ending social promotion also differed by their race and ethnicity. Thirty percent of teachers in our surveys identified themselves as African-American, and 10 percent identified themselves as Latino. African-American and Latino teachers were significantly more positive about the impact of the promotion policy even after accounting for differences in their years of teaching experience and the achievement level and racial and ethnic composition of their schools. African-American and other race teachers expressed fewer concerns about the impact of retention. Latino teachers and white teachers expressed more concerns about retention even though Latino

teachers had more positive views, on average, about the impact of ending social promotion (see Table 1-2).

Teachers' Assessments of the Impact of the Policy Varied Little by Grade or Subject Taught

Surprisingly, teachers in the promotion-gate grades (third, sixth, and eighth) were no more or less supportive of the end of social promotion than were teachers in other grades. However, these teachers expressed more concerns about the potential negative effects of retention, perhaps reflecting the fact that retained students were concentrated in their classrooms.

Language arts (reading) and mathematics teachers held similar views about the impact of the policy as teachers who taught other subjects (science or social studies). Teachers who taught auxiliary subjects, such as art or physical education, were slightly more supportive (an approximate 4 percent increase in the average level of support) of the reform efforts than were self-contained classroom teachers or mathematics, science, social studies, and language arts teachers. This finding is somewhat surprising because many teachers complained that the pressure of high-stakes testing takes away from emphasis on auxiliary subjects. At the same time, teachers in these subjects may have felt less pressure associated with trying to raise students' test scores.

Teachers' Assessments of the Impact of Ending Social Promotion Did Not Vary by the Achievement and Racial Composition of Their Schools

Teachers in low-performing schools faced a greater task because so many more of their students were at risk of retention. These schools were also faced with the threat of probation and the potential for additional school sanctions if they did not raise their test scores. We looked at how teachers' assessments of the impact of the promotion policy and their views on retention varied by the achievement level and racial and ethnic

composition of the school (e.g., whether the student body was predominantly African-American, predominantly Latino, etc.). The achievement level of the school was assessed by the percentage of students reading at or above national norms in 1994.

Surprisingly, teachers' assessments of the impact of ending social promotion did not differ by the achievement level or racial composition of their school once we accounted for the characteristics of teachers and the number of students in the school who were exempted from the exam.¹³ However, teachers in the lowest-performing schools in the system expressed greater concerns about the impact of retention on their students.

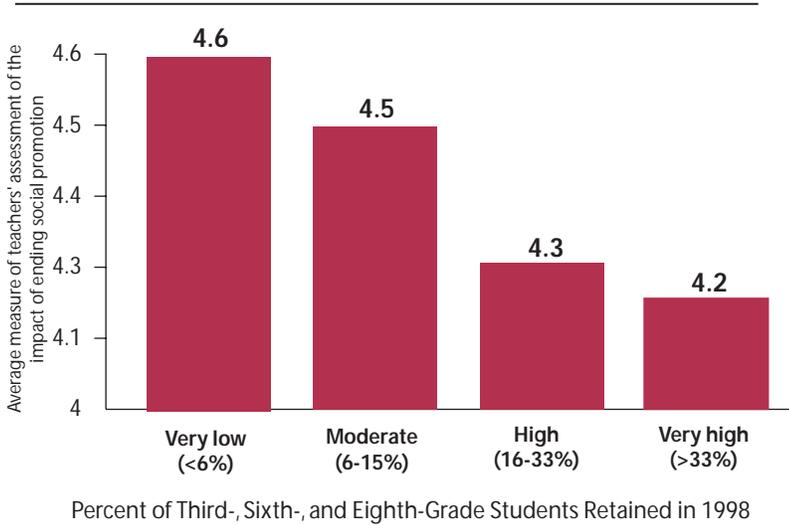
Teachers in Schools with Many Retained Students Were Less Positive about the Impact of Ending Social Promotion and Expressed More Concerns about Retention

While the overall achievement level of the school did not influence teacher assessment of the impact of ending social promotion, their assessments varied significantly by the initial experience of the school under the policy. We examined how teachers' assessments of the impact of the policy in 1999 varied by the retention rate of the school in which they taught in the preceding year, 1998. In 1998, 26 percent of CPS's elementary schools had very low retention rates, e.g., less than 6 percent of third, sixth, and eighth graders in the school were retained.¹⁴

There is a strong relationship between a school's 1998 retention rate and teachers' assessments in 1999 of both the overall impact of the ending social promotion initiative and their views on retention. Specifically, teachers in schools with high and very high retention rates (over 15 percent of third, sixth, and eighth graders retained) responded less positively to the survey items than teachers in schools with lower retention rates (see Figures 1-10 and 1-11).

Figure 1-10

Teachers in Schools with High Retention Rates Were Less Positive about the Impact of Ending Social Promotion (1999)

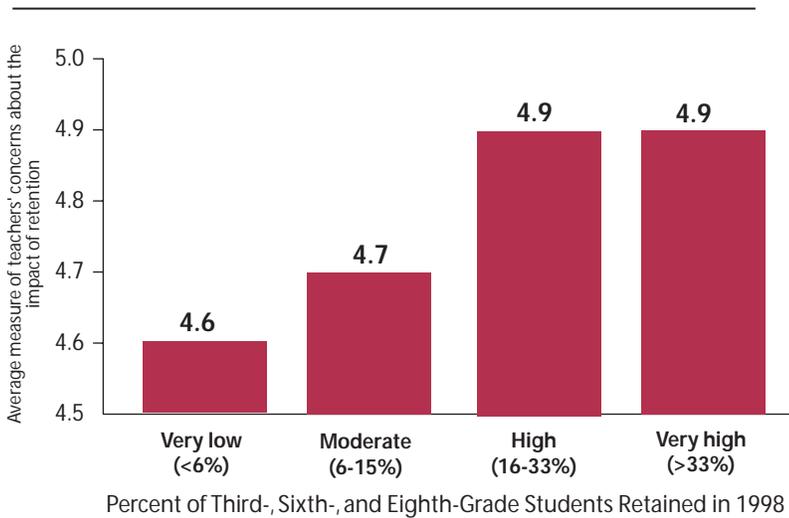


Note: This measure combines teachers' responses to individual survey items about their views of the policy to end social promotion. The measure explores the overall degree of support each teacher has for the policy. A higher score indicates greater support for the policy. This was measured on a 10-point scale.

There are two reasons why teachers in schools with high retention rates may have responded to these survey items more negatively. First, teachers in schools with high retention rates may have been accurately reporting that the policy had had few positive effects in the school. Perhaps the school had not responded positively by organizing and by working to change behavior, and the students had not changed their behavior. Second, teachers in schools with high retention rates may have had a more realistic sense of the costs associated with the policy, having spent a year managing large numbers of retained students.

Figure 1-11

Teachers in Schools with High Retention Rates Were More Concerned about Retention (1999)



Note: This measure combines teachers' responses to individual survey items about their views of the policy to end social promotion. The measure explores the overall degree of support each teacher has for the policy. A higher score indicates greater support for the policy. This was measured on a 10-point scale.

Changes in Teachers and Principals' Assessments of Ending Social Promotion, 1999 to 2001

Up until this point, we have examined teachers and principals' assessments of the short-term impact of the policy, using survey and interview data collected in 1999, two years after the implementation of the policy. We might expect teachers and principals to view the policy differently over time. In the short run, teachers clearly welcomed the program supports provided to help struggling students and were pleased with the increases they observed in parental involvement and stu-

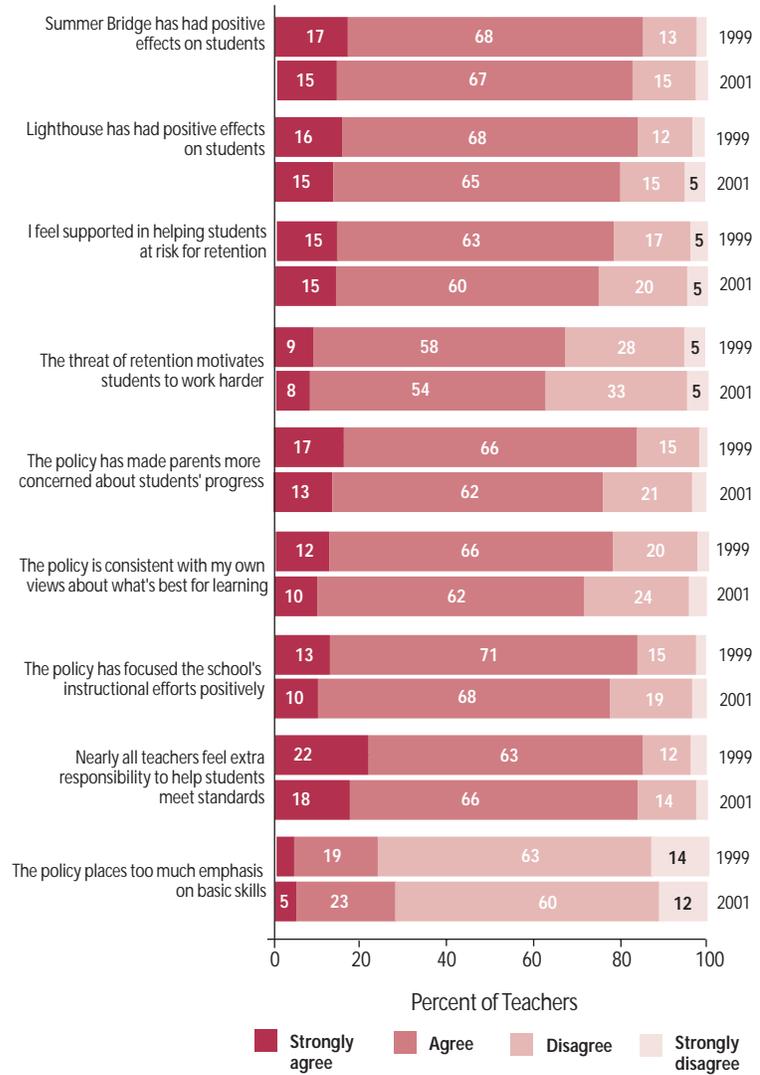
dent motivation. Despite a few concerns, teachers generally felt that their schools were focusing effectively on the needs of their lowest-performing students, and that the policy had met students' needs. Teachers overcame their concerns about the policy and endorsed it. Yet, research on high-stakes testing often finds that test scores rise significantly following the initial impact of high-stakes testing and then level off over time. We might expect to see a similar leveling off of support for such a policy among educators as the initial benefits of the program begin to wane, and as they gain greater exposure to retained students, who often fail to show improvements academically and may suffer emotionally.

In 2001, teachers responded slightly less positively to the same sets of questions about the end of social promotion policy (see Figure 1-12). Teachers were slightly less likely to agree that Summer Bridge and Lighthouse had a positive impact on students and were slightly less likely to agree that they felt supported in their efforts to help at-risk students. To some extent changes occurred in teachers' assessments of the impact of the policy on student and parent motivation. While in 1999, 67 percent of teachers agreed or strongly agreed that the threat of retention motivated students to work harder in school; in 2001, only 62 percent agreed

Figure 1-12

Teachers' Attitudes about the Impact of Ending Social Promotion Changed Little between 1999 and 2001

Comparison Between 1999 and 2001 Teachers' Responses



with this same statement. The proportion of teachers who agreed that parents were more concerned as a result of the policy also fell from 83 percent to 75 percent. Over time it appears that teachers were somewhat less likely to attribute changed behavior on the part of students and parents to the introduction of the promotion-gate test-score cutoffs.

In addition, teachers were also somewhat less likely in 2001 to agree that the policy was consistent with their own instructional views and were somewhat more likely to believe that the policy placed too much emphasis on basic skills. When we combine teachers' responses into a summary measure identical to the one created from teachers' responses to individual survey items in 1999, (see Figures 1-8 and 1-9 on pages 22-23) we find that the proportion of teachers who could be characterized as strongly supportive of the promotion policy declined slightly from 16 percent to 14 percent, while the number who were categorized as moderately supportive remained the same. Yet, counter to expectation, there were no changes in teachers' attitudes toward the practice of grade retention.

In general, it appears that the initial enthusiasm for the policy may have begun to wane slightly by 2001. This may be due to a variety of factors, including changes in the behavior of students and parents after the initial impact of the policy began to fade or to increased expectations on the part of teachers for student behavior and performance. Teachers may also have begun to have greater concerns about the potentially negative aspects of the policy as time passed. Nevertheless, the decline in teacher support for the policy from 1999 to 2001 seems moderate. Even five years after policy implementation, more than three-quarters of the teachers remained either moderately or highly positive about the impact of ending social promotion.

Conclusions

Given that research consistently finds that educators in other states and districts are resistant to high-stakes testing policies and given the pressures that high-stakes testing places on educators, it may be surprising that Chicago Public Schools teachers and principals reacted positively to efforts to end social promotion. However, this analysis suggests that many teachers viewed the policy as something that supported rather than constrained their own work in the classroom. Many educators welcomed the possibilities the policy held for increasing the motivation of students and the involvement of parents. Teachers believed the policy had a positive impact on students by establishing a goal for them to work toward, providing them with motivation to work harder in school, and sending the message that achievement matters. In addition, unlike many high-stakes testing policies, the CPS policy provided teachers with significant resources to help with classroom work. Teachers welcomed these program supports and the availability of extra interventions for students. As a result, it appears that rather than viewing the policy negatively, they viewed it as something designed to help them be more successful in their classrooms. We do find some evidence of weakened support over time, yet even five years after the institution of the policy, most Chicago educators remained positive about both the impact of supports and the impact on students and continued to believe that the policy was consistent with their own instructional foci.

While Chicago educators were, on the whole, positive about the impact of Chicago's initiative to end social promotion, their support was not an overwhelming endorsement. Many felt the policy had room for improvement and believed that both increased teacher input in the promotion decision-making process and

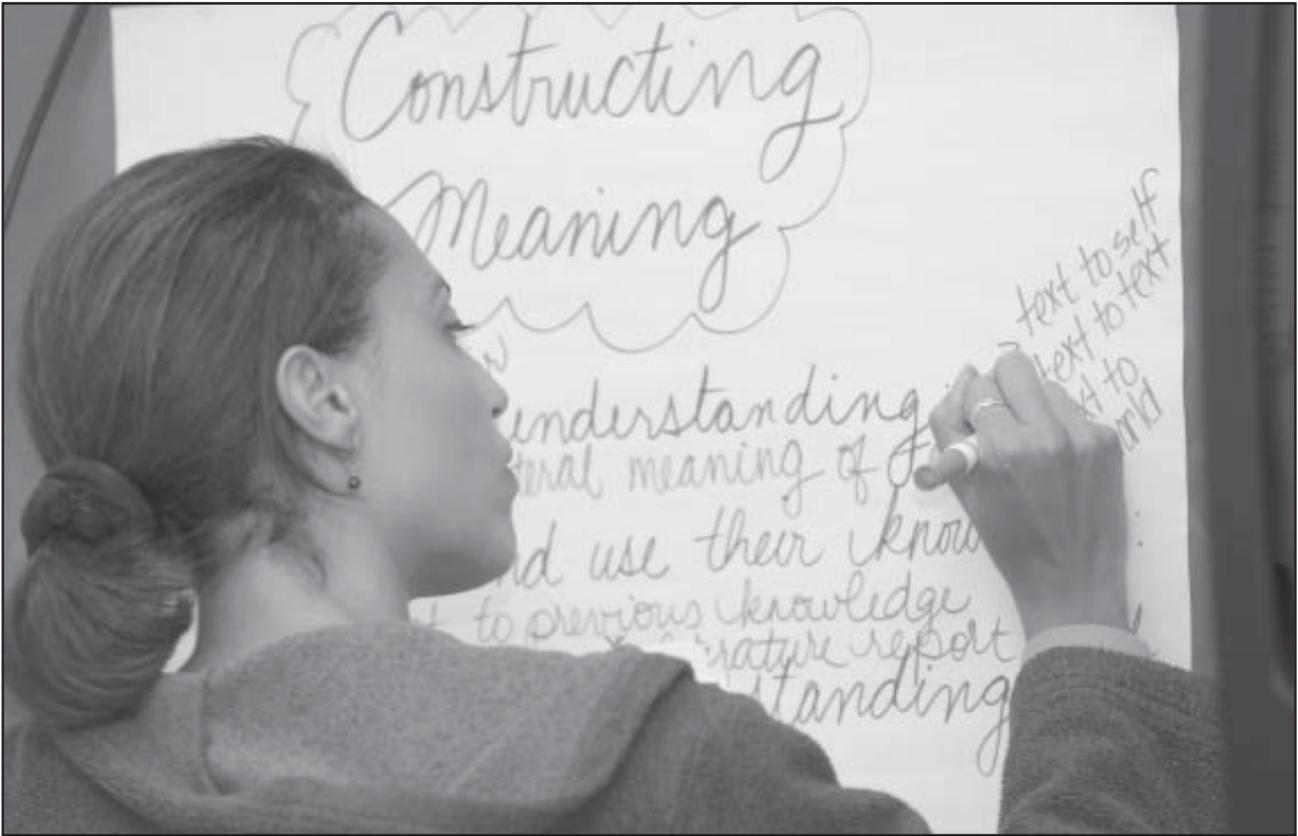
greater attention to a wider range of students' needs were necessary. Interestingly, variance in teachers' assessments of the impact of the policy and of retention was driven largely by differences across teachers in their educational and demographic characteristics rather than the characteristics of their students. Older and more experienced teachers were significantly more positive about the effects of ending social promotion on students and instruction, while younger and more educated teachers viewed the policy and its effects more negatively.

Yet, despite their concerns, the end of social promotion appears to have met some immediate needs of educators, engendering their tentative support. As one teacher noted:

I'd hate to go back to social promotion because the little experience I did have with it, it was still, "It doesn't matter if I do it or not, because I know next year I am moving on". . . . My sister is a school psychologist, and she thinks the idea of having them retained is terrible; it's going to destroy their psyches, and they are going to have no self-image. Well, how much self-image are you going to have if when you get out you can't get a job and people fire you because you didn't follow the instructions or whatever. That cannot be very helpful to your positive self-image, either. So I have done 10 rounds with my sister

on that one, and I don't think the city had any choices. . . . It's [the policy] going to have rough edges to start with. I mean this is . . . a big system, and it takes a while to get things working right. I'm not ready to go back to social promotion, that's for sure. (Sixth-grade teacher)

Opponents of high-stakes testing often base their opposition on research evidence that suggests that retention has few academic benefits. Indeed, research has documented that students who were retained under the policy often struggled academically in the year after they were retained, and teachers in schools with the highest retention rates expressed greater concerns about the policy.¹⁵ In this chapter, we hypothesized that one of the reasons that teachers were so positive about the impact of CPS's effort to end social promotion is that at a minimum they viewed social promotion as more negative than retention, and many Chicago educators viewed retention as having positive educational benefits. At the same time, teachers expressed concern that not enough was being done to assist retained students. Expanding interventions and supports for students who were retained would become crucial to both addressing the critical needs of students with the weakest skills and ensuring that educators' concerns were addressed.



John Booz

CHAPTER

2

The Instructional Response to High-stakes Accountability Trends in Test Preparation and Content Emphasis in Mathematics and Reading 1994-2001

by

Robin Tepper Jacob

As we found in the previous chapter, the 1996 reforms placed significant pressure on Chicago Public Schools (CPS) teachers to raise reading and mathematics test scores. But how did teachers change instruction as a result of that pressure? Did they focus more on test preparation, or did they change how they taught or what they taught? In the previous chapter, we found that the majority of teachers and principals believed that the ending social promotion initiative had a positive impact on their own behavior and the focus of instructional efforts. These responses suggest that teachers changed their instructional approach in response to high-stakes testing. But there is significant debate about whether high-stakes testing positively or negatively changes what and how teachers teach. Critics of high-stakes testing argue that the majority of teachers will respond by limiting their classroom content to what is on the test and by spending a great deal of time teaching students to take tests. Teachers could also react to the pressures of high-stakes testing by changing their instructional practices to ensure that all students are exposed to grade-level material and skills and are academically successful.

In this chapter, we take a close look at trends in teachers' reports of the time they spent on test preparation and on the content they emphasized in mathematics and language arts between the 1993 and 1994, and 2000 and 2001, school years. We used survey data collected by the Consortium on Chicago School Research in the spring of 1994, 1997, 1999, and 2001, to trace trends in these areas. Additionally, we drew on teacher interview data to look at how teachers described their efforts to improve test scores (see *Data Used in this Chapter*). We looked specifically at how subject matter emphases and time devoted to test preparation differed across grades, and at whether teachers in low-performing schools and schools that served high proportions of African-American and Latino students exhibited greater changes in instructional practice than teachers in other schools.

How Might Teachers Change What They Teach in the Classroom? Test Preparation and Content Alignment

The most obvious way a teacher faced with pressure to increase test scores could respond is to spend more time teaching students how to take a particular test, a practice often referred to as test preparation. While

educators and policy makers talk about test preparation as if it were a clearly defined activity, there are, in fact, a variety of ways in which teachers may teach students how to take a test. In this report, we identify four different approaches teachers may employ to prepare students for an important exam: 1) teaching test-taking strategies, 2) engaging in test simulations, 3) aligning assessments, and 4) aligning content.¹

Frequently, teachers will prepare students to take exams by teaching them test-taking strategies. These include talking with students about the type and number of questions they can expect in each section of the test, familiarizing students with the vocabulary used in directions, and helping students to identify appropriate strategies for completing the exam in the allotted time. Examples of test-taking strategies include telling students to “mark C” if they don't know the answer to a particular question or instructing students to “fill in all the blanks” if they are running out of time at the end of the test.

Teachers can also help prepare students for an important exam through test simulation, which involves administering practice exams and having students practice taking exams under time constraints. In a high-stakes testing environment, many teachers will give a practice exam to students every Friday or will spend a

Data Used in this Chapter

This chapter draws on much of the same data used in Chapter 1. However, while Chapter 1 uses teacher survey and interview data collected in 1999, this chapter also relies on teacher survey data collected in 1994 (prior to the implementation of the policy) as well as data collected in 1997, 1999, and 2001 (after the implementation of the policy) to explore changes in teachers' reports of their own behavior since the ending social promotion policy began. The 1994 surveys are generally used as the baseline in these analyses. However, in several instances there were no comparable items or measures available in the 1994 surveys. When 1994 data were not available, changes from 1997 to 2001 were explored. There is evidence that the full effect of the policy did not take place until several years after high-stakes testing was implemented, so exploring changes using 1997 surveys as a baseline is not unwarranted; however, these findings should be interpreted with caution.

Surveys were administered to all elementary school teachers (K-8) in the Chicago Public Schools. Responses were received from 6,200 teachers in 1994; 10,300 teachers in 1997; 7,900 teachers in 1999; and 8,572 teachers in 2001. Of the elementary schools in the system, 56 percent were represented in the 1994 survey responses, 88 percent were represented in 1997, 80 percent in 1999, and 73 percent in 2001.



John Booz

week before the test administration having students take repeated timed exams.

Similarly, teachers often align assessments so that students become familiar with providing answers in the form that a particular test requires (e.g., using multiple-choice exams in class or asking questions on class assignments using the same wording and format as the test). Teachers may know, for example, that a test requires students to demonstrate a particular skill (e.g., identifying the main idea of a paragraph in a multiple-choice format). In response, teachers could

alter their classroom exercises and assessments so that students are frequently asked to identify the main idea of a passage, and teachers could use multiple-choice assignments to assess students' knowledge.

Finally, teachers frequently prepare students for important exams by teaching the specific content that will be covered on those exams, an approach we refer to as aligning content. If a teacher knows, for example, that the Iowa Tests of Basic Skills (ITBS) emphasizes fractions and decimals, she could alter her curriculum so that it emphasizes the acquisition of these skills.

In this report, we make a distinction between aligning content (i.e., covering tested subjects and skills in class) and those test-preparation activities that are simply designed to help students get better at taking standardized tests such as the ITBS. Therefore, we define test-preparation activities as those activities that involve teaching test-taking strategies to students, simulating the testing environment or aligning assessments, but not those activities that include content alignment. While teaching test-taking strategies or providing practice exams may enable students to better demonstrate their knowledge, such activities do not

As noted in Chapter 1, almost half of the 16,895 elementary school teachers in the system responded to the 1999 survey. The response rate was slightly higher in 1997 and slightly lower in 1994. In 1999, the survey sample was 16 percent male, 33 percent African-American, 46 percent white, and 13 percent Latino. Of the teachers responding, 54 percent had completed at least a master's degree, and 46 percent had been teaching for more than fifteen years, while 25 percent had been teaching for fewer than six years. Chi-square analysis indicates that the survey had fewer African-American respondents than the system as a whole (33 percent in the survey versus 41 percent systemwide). Survey respondents were also more highly educated than the system as a whole (54 percent reporting a graduate degree or higher in the survey versus 31 percent in the system as a whole). Results were similar in other survey years.

Survey responses provide a rich source of information for comparing teacher practice before and after the implementation of these policies. In addition, this chapter draws on the personal interviews conducted with 43 teachers in the promotion-gate grades (third, sixth, and eighth) who taught in one of five K-8 schools. These data are described in more detail in Chapter 1. Teacher interview responses are used to corroborate survey findings.

Even prior to the 1996 accountability reforms, a substantial proportion of Chicago teachers spent more than 20 hours each year preparing students for standardized exams.

increase students' underlying skills. But if teachers alter the content of their instruction, we might expect students' opportunities to learn to change, potentially leading to increases in achievement that could be generalized to other contexts. For example, we would expect students who learn to understand and manipulate fractions in a variety of contexts, because their teachers emphasized fractions in class, to have a different benefit than students who spend time learning how to answer the fractions questions in the particular way in which they are asked on the ITBS. In the next section we explore evidence for changes in the time teachers spent on direct test-preparation activities—activities designed to help students improve their test-taking skills. We then investigate the changes that teachers made in the content they covered in their classrooms.

Trends in Teacher Reports of Time on Test Preparation (1994-2001)

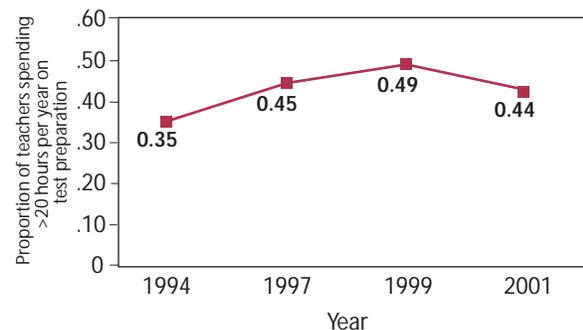
In 1994, 1997, 1999, and 2001, the Consortium's surveys asked teachers to indicate how many hours they had spent that year preparing students for standardized tests such as the ITBS and Illinois Goals Assessment Program (IGAP)/Illinois Standards Achievement Test (ISAT). Teachers were asked to report whether they had spent less than four hours on test preparation that year, between four and 12 hours, between 13 and 20 hours, or more than 20 hours a year. In this section, we examine changes in the proportion of teachers reporting that they spent more than

20 hours a year on test preparation.² To ensure that these data reflect pedagogical trends rather than changes in the composition of the teaching force or changes over time in the teachers who answered the surveys, we estimated the proportion of teachers across surveys who spent more than 20 hours a year on test preparation, accounting for differences in racial composition, gender, years of education, years of teaching experience, and grades and subjects taught by the teachers answering the survey. Therefore, results can be interpreted as the changes in test preparation that would have been observed if the composition of the teaching force remained similar between 1994 and 2001. (Note: All models were estimated using Hierarchical Linear Models (HLM). See Appendix B for a complete description of the models estimated.)

Time on Test Preparation Increased Significantly from 1994 to 2001

Even prior to the 1996 accountability reforms, a substantial proportion of Chicago teachers spent more than 20 hours each year preparing students for standardized exams. Our model estimates suggest that in 1994 over one-third of all Chicago teachers were spending more than 20 hours each year on test-preparation activities (see Figure 2-1). However, the proportion of teachers reporting that they spent more than 20 hours

Figure 2-1
A Substantial Proportion of Teachers Spent More Than 20 Hours a Year on Test Preparation



Note: All figures are scaled to show half a standard deviation above and below the 1994 mean. Estimates shown are from a hierarchical linear model that controls for the demographic characteristics of the teachers responding to the survey. See Appendix B for a complete description of the model.

a year on test preparation increased significantly after the implementation of high-stakes testing in 1996, and continued to rise through 1999. By 1997, the estimated proportion had increased to 45 percent, and by 1999, to nearly 50 percent. In 2001, the estimated proportion of teachers who reported spending more than 20 hours a year on test preparation declined somewhat, from nearly 50 percent to 44 percent.

Increases in Time for Test Preparation Were Largest in the Promotion-Gate Grades

The 1996 policies placed more emphasis on testing in all grades since schools were placed on probation based on the average percent of students reading at or above national norms in the entire school. However, teachers and students in the third, sixth, and eighth grades—the grades where students faced test-score cutoffs for promotion—faced the greatest pressure.

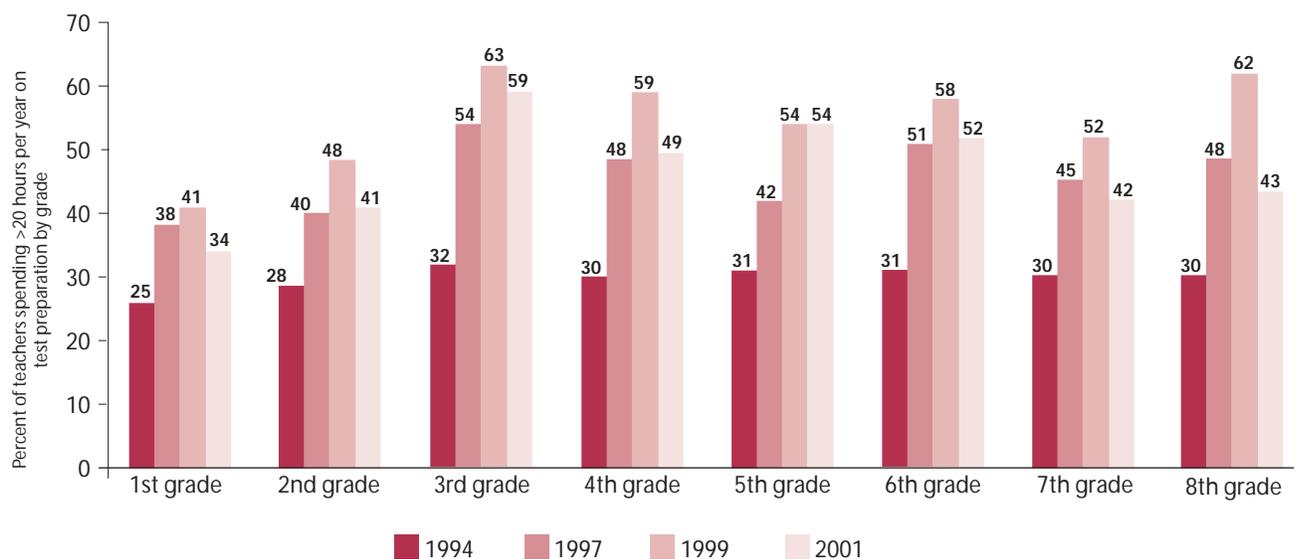
Not surprisingly, time on test preparation increased substantially in these grades. Prior to the implementation of high-stakes testing, an equal proportion of

teachers in each grade (about 30 percent) reported spending 20 hours or more on test preparation each year. Between 1994 and 1999, our estimates suggest that the increase in the proportion of third- and eighth-grade teachers spending more than 20 hours per year on test preparation outpaced the increases observed in other grades. By 1999, almost two-thirds of all teachers in the third and eighth grades reported spending more than 20 hours per year on test preparation. In no other grades were such high levels of test preparation reported (see Figure 2-2).

The proportion of teachers who reported high levels of test preparation increased substantially in the sixth grade, although the increase was not as marked as in the third and eighth grades. We estimated that the proportion of sixth-grade teachers spending more than 20 hours per year on test preparation increased by 26 percentage points between 1994 and 1999 (see Figure 2-3). In addition to these three promotion-gate grades, similarly large increases occurred in the fourth grade.

Figure 2-2

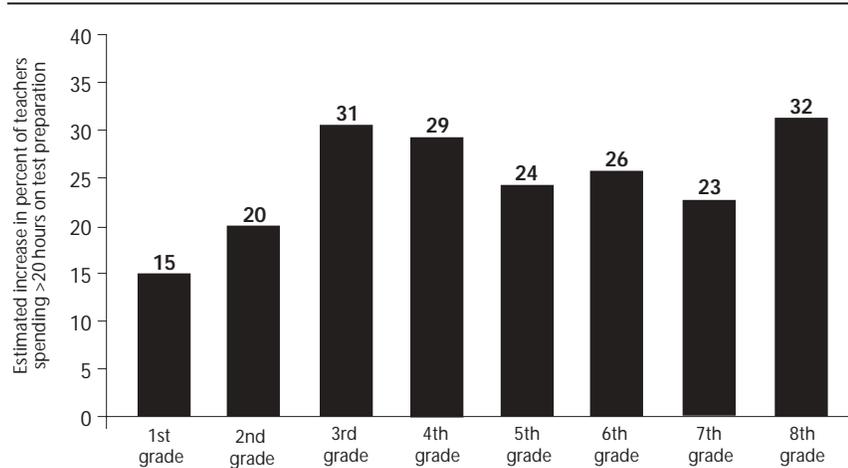
Teachers Reported Spending Higher Levels of Time on Test Preparation since 1994



Note: See Appendix B for a complete description of the model. A separate model was estimated for each grade.

Figure 2-3

The Percent of Teachers Spending High Levels of Time on Test Preparation Increased Most in the Third and Eighth Grades



Note: The change is calculated by subtracting the 1999 estimate from the 1994 estimate (see Figure 2-2). See Appendix B.

Time on Test Preparation Increased Most in the Lowest-Performing Schools

CPS's accountability policy provided the strongest incentives for the lowest-performing schools. These schools faced significant pressure to change behavior in order to get off of, or avoid being placed on, academic probation. We looked at changes in test preparation between 1994 and 2000 by both the achievement level of the students and the racial composition of the school.³ Specifically, we estimated the proportion of teachers reporting that they spent more than 20 hours during the year on test preparation, taking into account teachers' race, gender, education, years of teaching experience, grade level, and subject matter taught, as well as the racial composition of the schools in which they worked. Thus, we estimated the change in test preparation that occurred in schools of different achievement levels if all schools had had similar teaching staffs and had students with similar racial backgrounds.

Between 1994 and 1999, teachers in the lowest-performing schools reported dramatically higher levels of test-preparation activity. The estimated proportion of teachers who reported spending more than 20 hours per year on test preparation increased by an average of 23 percent in schools that had initially been placed on academic probation (schools with fewer than 15 percent of students reading at or above grade level on national norms). (See Figure 2-4.)

The estimated proportion of teachers who reported spending more than 20 hours a year on test preparation also increased significantly in schools that faced the threat of academic probation (schools in which 15 to 24

percent of the students scored at or above national norms in reading). In contrast, increases in test preparation were relatively modest in schools with better performance. The estimated proportion of teachers who reported more than 20 hours a year of test preparation increased by 11 percent, on average, in schools with 25 percent or more of their students reading at or above grade level on national norms. Indeed, using a statistical model to impute time on test preparation, (see *Another Look at Time Spent on Test Preparation* on page 38) we estimated that in the lowest-performing schools, the average time teachers devoted to test preparation doubled between 1994 and 1999.

While school performance was a strong predictor of time devoted to test preparation, the racial composition of the school in which a teacher worked was not. Schools with higher proportions of Latino and bilingual students showed slightly greater increases in test preparation between 1994 and 1999 than did predominately African-American schools. However, we need to use caution interpreting these results. In 1994, teachers in African-American schools were already spending a great deal of time on test preparation. In 1994, over 45 percent of teachers in moderate-performing African-American schools reported that they spent more than 20 hours each year on test preparation. We would,

therefore, expect to see a smaller increase in the proportion of teachers reporting that they spent more than 20 hours per year on test preparation since such a large proportion of the teachers in these schools were already responding in this category prior to the policy’s implementation. Yet, results from our simulation suggest that the amount of time spent on test preparation in African-American schools doubled between 1994 and 1999.

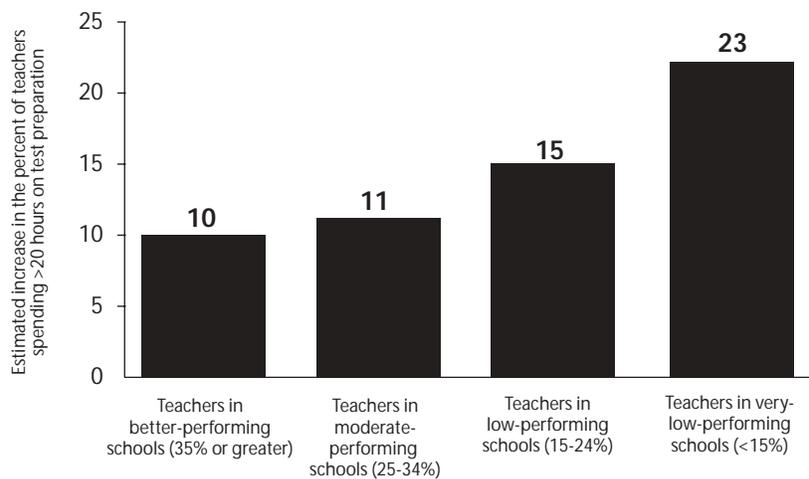
Interviewed Teachers Emphasized That Test Preparation Was Critical to Increasing Test Scores

Teachers in the low-performing schools where we conducted our interviews confirmed that they had placed new emphasis on test preparation in response to the 1996 accountability programs. When asked to talk about the strategies they had found for successfully raising student test scores, most teachers indicated that they emphasized test preparation using a wide variety of approaches. First, many teachers indicated that test simulation and familiarizing students with the layout and format of the test was, in their experience, the most effective strategy for raising test scores.

When they saw the actual form . . . to see how the questions were worded, it worked better than any-

Increases in the Percent of Teachers Spending High Levels of Time on Test Preparation Was Greatest in Low-Performing Schools

Figure 2-4



Note: See Appendix B. In 1994, 24 percent of the city’s 456 elementary schools were **very-low-performing** (<15 percent of students reading at grade level on national norms). Thirty-six percent of schools were **low performing** (15-24 percent of students reading at national norms). Sixteen percent of schools were **moderate performing** (between 25-35 percent of their students reading at or above national norms in reading). Twenty-four percent of schools were **better performing** (more than 35 percent of their students were reading at or above national norms on the Iowa Tests of Basic Skills).

thing I had ever used. When I explained to them, “You know how to do this.” . . . If I would go to the board and do it, they’ll say, “Oh, okay.” (Sixth-grade teacher)

Timed lessons. Bubbling in . . . the actual answer sheet. That was similar to the ITBS. Whereas, you know, just using notebook paper and actually write the problem out. They had to use scratch paper and then they had to transfer the answer onto the bubbling sheet. (Sixth-grade teacher)

Also, teachers talked about teaching test-taking strategies and about explaining to students how to understand the questions they were likely to see on the exam. A number of teachers observed that while their students knew the skills being tested, they often did not recognize what they were being asked to do in a testing situation. One third-grade teacher explained how she prepared students for the test:

[I taught them] what “how many in total” means . . . because many times they know the material; they know the operation that they have to do, but they don’t understand the vocabulary.

Another Look at Time Spent on Test Preparation

Looking at the proportion of teachers who reported spending more than 20 hours per year on test preparation gives us an estimate of how many teachers placed a great deal of emphasis on test-preparation activities. It does not tell us how much time teachers were actually spending. A teacher who reported more than 20 hours could be spending 25 or 75 hours. In addition, we don't know if those teachers who were already spending 20 hours or more on test preparation in 1994 also increased their time. For example, even before the new promotion policy was implemented, 47 percent of all third-grade teachers reported spending more than 20 hours per year on test preparation. If many of these teachers were spending a little over 20 hours per year in 1994 but had increased the number of hours to a little over 50 by 1999, we would not be able to detect this change only by looking at the changes in the number of teachers who reported spending more than 20 hours per year.

Statisticians have developed approaches to dealing with the problem of interval data where respondents have an upper limit on their response categories (e.g., 20 hours or more). Put simply, if we assume that time spent on test preparation is normally distributed, we can use information on both the percentage of teachers who responded to each interval (each response category) and knowledge of the normal distribution to estimate the average time spent on test preparation each year as well as how different subgroups might have responded. Specifically, we used an application of regression analysis called Tobit analysis for interval regression to estimate the mean time spent on test preparation each year as well as changes for particular subgroups (African-American versus Latino schools).

Using this technique, we estimated that in 1994 teachers were spending, on average, approximately 10.5 hours per year preparing students for standardized tests. By 1999, that number increased to 21 hours. Among predominantly African-American schools where teachers were spending a great deal of time on test preparation prior to the policy's implementation, we estimated that test preparation increased from almost 16 hours per year in 1994 to almost 31 hours in 1999. Among the very-low-performing schools, the estimated time increased from an average of 14 hours to an average of 32 hours. Among third-grade teachers, the estimate increased from 14 hours in 1994 to 31 hours in 1999. Thus, our estimates suggest that the average time spent on test preparation doubled between 1994 and 1999.

Another described her strategies:

I made a list of vocabulary. I gave them a list of the words and told them "This means this. Whenever you see this word here, that means you're going to have to do one of these." That helped them. From the beginning of the year, I gave them vocabulary and tried to do word problems. (Seventh-/Eighth-grade teacher)

Finally, teachers talked about teaching students to problem solve in relation to the test. They gave students practice exams and taught skills such as process of elimination and estimation:

They had forty-five minutes to do the assignment . . . to read it and do it. And then we went over and discussed the answers, the process of elimination, and why they chose the answers. [We would go] back in the paragraph, identifying where they found that answer and why they

chose that answer, to help them understand why the answer was right or wrong. (Eighth-grade teacher)

I teach them strategies to deal with things they find unfamiliar or don't know. I teach them about cause and effect. I also focus on problem-solving capabilities; things that help them find a way out of a box. (CPS teacher)

My classes were behind in both reading and math. When I went through the ITBS book, they could not do any of the problems....I drilled them for two weeks. I showed them how to do the problems, explained to them, had them go to the board, and then tell me how they did the problem, how they got the answer. [They had to] use the proper vocabulary/math terminology when explaining the steps. . . . I did not want them to fail because I did not want them

to be another statistic. I did word attack and decoding skills and phonics skills with them. [We practiced] context clues, what we call higher-order thinking skills, and I drilled them all year on that because they were missing the very basic, very, very basic. They could not identify main idea, author's purpose, prediction. (Eighth-grade teacher)

One problem with using survey data to track changes in test preparation is that it is difficult to know how to interpret teachers' survey responses. In this report, we have made a distinction between test-preparation activities designed to help students improve their ability to take tests and changes in content coverage designed to improve test scores. However, when teachers indicated on surveys how much time they spent on test-preparation activities, they may have been reporting on the time they spent on specific test-taking strategies or they may simply have been referring to any activity that they believed would lead to higher test scores, including teaching topics they knew would be on the test (i.e., aligning content).

However, an analysis of interviewed teachers' descriptions of their test-preparation activities suggests that they described direct test-preparation activities—practicing test simulation, familiarizing students with the vocabulary and kinds of questions asked on the ITBS, and working on direct test-taking strategies. Although many interviewed teachers indicated they had aligned their curriculum with the ITBS content, few talked about curriculum alignment as a specific test-preparation strategy. In summary, these interview results (see *Defining Test Preparation* on page 40) suggest that when teachers reported increased time on test preparation on sur-

veys, they were most likely reporting time on test-taking strategies rather than on content alignment.

How Much Time Are Teachers Really Spending?

From survey data analysis, we know that many CPS teachers spent more than 20 hours each year on test preparation and that that proportion increased significantly between 1994 and 2001. Survey data do not, however, tell us the actual amount of class time teachers devoted to these activities. A teacher who reported spending more than 20 hours a year may have spent as few as 21 hours each year or as many as 250 hours. In our 1999 interviews with teachers in low-performing schools, we asked teachers to give estimates of the total number of hours they spent on test preparation throughout the year. We also asked teachers to indicate how much time they spent on test preparation in the month prior to the ITBS. For example, one teacher indicated that in the “three weeks before the test, we spent at least one period a day specifically on test-taking strategies and practice tests, but did nothing before then.” Assuming that one period is approximately an hour long and that preparation took place five days a week for three weeks, this teacher's time on test preparation would be approximately 15 hours. Another teacher indicated that he devoted at least 20 minutes a



Defining Test Preparation

As shown in Figure 2-1, many more teachers reported spending over 20 hours a year on test preparation since the end of social promotion, but it is unclear how we should interpret these reports. The survey question used in the analysis asked teachers to indicate how much time they spent during the school year “preparing students for standardized tests such as the ITBS or IGAP/ISAT.” To what activities were teachers referring when they answered this question? Were they indicating how much time they spent teaching test-taking strategies, administering practice tests, and using aligned assessment tools? Or were they referring to the time they spent covering the topics they knew would be on the ITBS (curriculum alignment), the effects of which are more ambiguous?

Teachers’ interviews provide some insight. In interviews as well as surveys, teachers were asked to indicate how much time they spent preparing students for the Iowa Tests of Basic Skills (ITBS); however, in interviews they were then asked to talk about the kinds of activities they engaged in during these test-preparation periods. Of the teachers we interviewed, 80 percent (28 teachers of 35) talked specifically about teaching test-taking strategies and familiarizing students with the form and format of the test during this time. These teachers talked about spending time exposing students to the “layout and format of the test.” They described using short, timed exams each day to help familiarize students with taking timed exams, using exams where students have to fill in the bubbles, and giving students timed ITBS practice tests. They also reported teaching student test-taking skills. An additional five teachers (14 percent) talked about using some combination of curriculum alignment and test preparation. While over half of the teachers we talked with (22 of 43 teachers) mentioned that they had aligned the content of their curriculum with the content of the ITBS, only 4 percent talked about covering ITBS-specific content when asked to describe their test-preparation activities. The following is a list of activities teachers indicated they engaged in during test-preparation time:

- Give practice tests
- Teach test-taking strategies
- Teach strategies to deal with things students find unfamiliar or don’t know
- Expose students to the layout and format of the test
- Give short, timed problems at the beginning of each day
- Use a bubble format for in-class exams
- Teach students to use the process of elimination in answering multiple-choice exams
- Continually give timed exams
- Use test-preparation booklets

day, five days a week for specific test preparation all year long. Thus, we estimated that this teacher spent 60 hours.⁴

Among all 35 interviewed teachers, we estimated that the average teacher spent 53 hours on test preparation in 1999. Two teachers reported spending substantially more of their instructional time on test preparation (240 hours and 180 hours), while three teachers reported spending no time at all.

However, 53 hours may overestimate the time spent on test preparation in the system as a whole since the interviewed teachers all worked in high-risk neighbor-

hoods, in schools that were on or faced the possibility of academic probation, and with students in the promotion-gate grades where the pressure to increase test scores was greatest. Indeed, we analyzed the survey data using a statistical method that allowed us to impute the mean hours of time spent on test preparation from survey responses (see *Another Look at Time Spent on Test Preparation* on page 38). From this analysis we estimated that in 1994, teachers were spending an approximate average of 10.5 hours per year, and by 1999, that average increased to 21 hours, substantially below the 53-hours-per-year estimate. However, among

teachers in the lowest-performing schools, the amount of time spent on test preparation increased from 14 hours in 1994 to 32 hours in 1999, still below the estimate of 53 hours, but nonetheless a substantial amount of time.

While these analyses provide estimates of the total time devoted to test preparation throughout the year, the impact of test preparation on the instructional environment differed depending both on how concentrated the time spent on test preparation was, and on when the test preparation occurred during the school year. Suppose that teachers do spend, as our interview sample suggests, an average of 53 hours each year on test preparation. If there are 180 days in the school year, this is, on average, about an hour and a half each week—the equivalent of a block of reading or mathematics time. Based on previous Consortium work, this represents approximately 8 to 9 percent of actual instructional time during the school year.⁵

Yet, this 8 to 9 percent would have had a different impact on instruction if it had been spread out over the entire school year than if the time was concentrated in the weeks immediately preceding testing. Indeed, if most of the time on test preparation was concentrated in the weeks prior to testing, 53 hours of test preparation is more than two weeks of instructional time. If we combined the time spent on test preparation with the instructional time lost to giving tests, this might mean that a month of instructional time is lost to testing each spring.

About half of the teachers interviewed indicated that a majority of the time they spent on test preparation took place in the month prior to testing, although most of these teachers indicated that they spent an hour a day or several hours each week on test preparation and did not stop regular instruction altogether. Five interviewed teachers indicated that they spent all of their instructional time in the weeks prior to the test on test-preparation activities.

. . . among teachers in the lowest-performing schools, the amount of time spent on test preparation increased from 14 hours in 1994 to 32 hours in 1999.

Trends in Teacher Reports of Content Emphasis (1994-2001)

Many Chicago teachers reacted to the pressures of high-stakes testing by increasing the amount of instructional time devoted to test preparation. Previous research has found that teachers will often react to high-stakes testing by changing what they teach to ensure that students are being exposed to the topics that are covered on the test.⁶ CPS placed schools on probation solely on the basis of their ITBS reading scores. The promotion decisions at third, sixth, and eighth grades were made on the basis of students' ITBS reading and mathematics scores. This reading and mathematics emphasis provided an incentive to shift instructional attention away from subjects such as science, art, or social studies. In addition, within reading and mathematics, teachers had an incentive to align their curriculum with the specific topics covered on the ITBS. The Iowa Tests of Basic Skills, as its name implies, is a basic skills test. The reading portion of the ITBS focuses on testing reading comprehension skills in three areas: (1) factual meaning, (2) inferential and interpretative meaning (e.g., drawing conclusions or inferring traits or feelings of a character), and (3) evaluative meaning (e.g., identifying the main idea of a paragraph). The mathematics portion of the ITBS tests skills in: (1)

Changes in Subject Emphasis Versus Changes in Instructional Foci within Subjects

Previous research has argued that teachers in other states and districts who are faced with pressure to increase test scores in particular subject areas tend to devote less time to subjects that are not covered on the exam.¹ In this chapter, we evaluate changes in content emphasis within particular subjects. Unfortunately, we do not have data to evaluate changes in the time teachers spent on reading or mathematics versus other kinds of activities. For example, while there is evidence that teachers spent more time on grade-level mathematics topics, we don't know whether the total instructional time devoted to mathematics changed. Similarly, while we have seen evidence that teachers gave significantly more emphasis to reading comprehension when they taught reading, we don't know whether the total time that teachers spent on reading instruction increased.

Evidence from surveys suggests that Chicago Public School teachers were split in their opinions about whether they placed less emphasis on science and social studies than they used to as a result of the accountability system. In both 1999 and 2001, Consortium on Chicago School Research surveys asked teachers whether they felt that as a result of the ending social promotion initiative, they “place less emphasis on social studies and science” than they used to. In 1999, 57 percent of the teachers responding disagreed with this statement. Results were similar in 2001.

One interpretation of teachers' survey responses is that the majority of teachers did not change their allocation of instructional time across subject areas, but simply the topics they covered within content areas. They might also have changed how they taught in content areas. For example, when teaching social studies, teachers may have placed greater emphasis on reading instruction within content areas. Another interpretation is that teachers were already spending limited time on social studies and science prior to the accountability program. As we saw in the previous chapter, many interviewed teachers stressed that reading and basic skills have always been their highest priority. Finally, teachers may have been reluctant to admit that they had shifted their subject-matter emphasis because they believed it would reflect badly on them. Findings from teacher interviews (see Chapter 1: Data Used in This Chapter) suggest that in fact many teachers have included more reading and mathematics instructional time. However, because we have not explored changes over time in subject-matter emphasis on Consortium surveys, we are limited in our ability to interpret these findings or to comment on the extent to which our interview findings might be generalized beyond teachers in low-achieving schools.

¹ See, for example, Smith (1991).

problem solving, (2) data interpretation, and (3) mathematical computation. In this section, we look at whether teachers in Chicago shifted their instructional foci in reading and mathematics to emphasize these topic areas. Unfortunately, we do not have the ability to evaluate whether teachers shifted the allocation of their total instructional time towards reading and mathematics versus other subject areas because we lack longitudinal data that asked teachers the percentage of time devoted to different subject areas (see *Changes in Subject Emphasis Versus Changes in Instructional Foci within Subjects*).

Mathematics Content Emphasis: Measuring Whether Teachers Focus on Grade-Level Subject Matter

In an influential Consortium report, Julia Smith and her colleagues conducted a careful analysis of Chicago Public Schools students' opportunities to learn mathematics prior to the implementation of the high-stakes testing initiatives in 1996.⁷ Using the 1994 Consortium survey data, Smith and her colleagues identified the content that teachers would need to cover in order to get their students to score at grade level on the ITBS and compared that to teachers' reports on Consortium surveys of the time they spent on each of these skills in

their classrooms. For example, for eighth graders, grade-level material included statistics and probability, the ability to graph equations, simplify algebraic expression, and solve the equation for a line. Students would need to be exposed to these topics if they were to have the opportunity to score at grade level on the eighth-grade ITBS. Sixth-grade material involved long division, number properties, and rounding.

The main finding of Smith and her colleagues' analysis was that pacing in mathematics instruction slowed down dramatically in the upper elementary grades. Students in the seventh and eighth grades were rarely exposed to grade-level material such as algebra or probability, and teaching in the upper elementary grades continued to focus on lower-level material. By eighth grade, three-quarters of teachers surveyed reported content emphasis below that expected for students to be deemed at grade level on the ITBS. This problem was worst in high-poverty schools, in majority African-American schools, and in predominantly minority schools.⁸

If teachers reacted to high-stakes testing by aligning their curriculum with the content of the test, we might have expected that teachers would have spent more time on grade-appropriate mathematics material. The battery of mathematics questions was asked in 1994, 1997, 1999, and 2001, allowing us to track over time whether teachers reported greater emphasis on grade-appropriate mathematics skills. We drew on the Smith, et al. analysis of the ITBS and compared how pacing had changed over time and across grades throughout this period (see *How We Determined Whether Teachers Taught "Grade-Level" Mathematics Content* on page 48).

Because CPS mathematics scores rose throughout this period, we estimated changes in teachers' emphasis on grade-level-appropriate mathematics content—or the proportion of time teachers devoted to grade-level or near-grade-level material—using a multivariate procedure that adjusted for incoming test scores in each grade. Thus, our mathematics content exposure estimates reflect changes in teachers' reports after accounting for the fact that in each grade, teach-

ers were more likely to have students with higher mathematics test scores coming into their classrooms in 2001 than in 1994.

Teachers Spent More Time on Grade-Level Mathematics Material

Even after controlling for rising levels of achievement among students, teachers' survey responses suggest that students' exposure to grade-appropriate mathematics content improved in the eighth grade (see Figure 2-5). There was less improvement in other grades.⁹

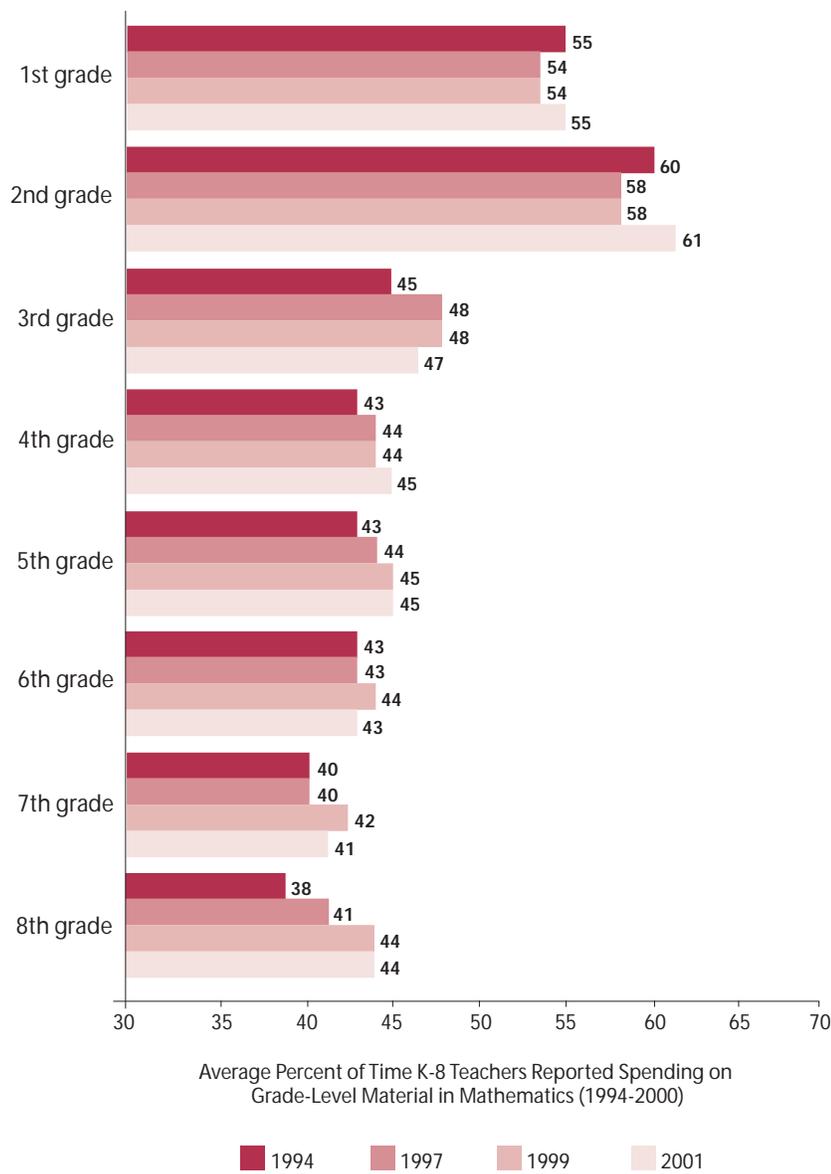
Consistent with the findings of Smith and her colleagues, first- and second-grade teachers were much more likely to report spending time teaching grade-appropriate mathematics topics than upper-grade teachers. However, seventh- and eighth-grade mathematics content exposure improved from 1994 to 2001, with the most dramatic increases occurring among eighth-grade teachers.

In 1994, the average eighth-grade teacher reported spending only 38 percent of her time on topics that could be considered at grade level on the ITBS (see Figure 2-6). By 2001, the average eighth-grade teacher reported spending 44 percent of her time on grade-level material. Similarly, the average proportion of time that eighth-grade teachers reported spending on the most basic mathematics concepts—those associated with first- through third-grade material (such as simple addition, reading a clock, and multiplication facts) declined from nearly one-quarter of their mathematics instructional time in 1994 to 18 percent in 2001. While an increase of 6 percentage points may seem relatively small, it represents an increase of well over one-half of a standard deviation above the 1994 levels.

In some ways these findings are promising. Eighth-grade teachers were more likely in 2001 than in 1994 to report that they exposed students to material that would be considered at grade level on the ITBS. But these results also suggest that mathematics pacing remained a problem. While in the early grades, teachers reported spending the majority of their time on grade-level mathematics material, there was a dramatic decline

Figure 2-5

Eighth Grade Teachers Increased Students' Exposure to Grade-Level Mathematics Content from 1994 to 2001—Other Grades Showed Less Improvement



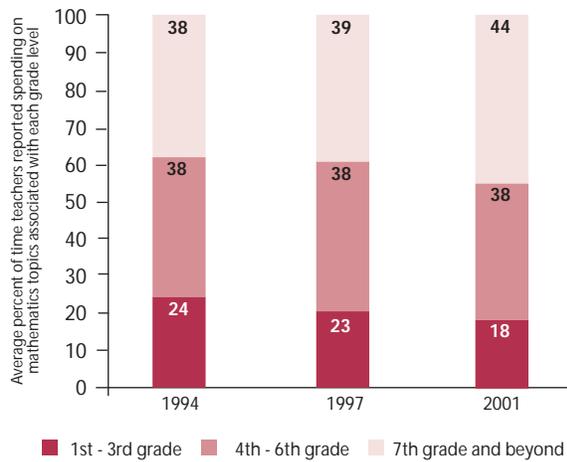
Note: Grade-level material includes content considered appropriate for students in that grade or the grade immediately above or below it. Estimates shown are from a hierarchical linear model that controls for the demographic characteristics of the teachers responding to the survey. For grades four through eight, where valid test score data are available, the models also control for the incoming achievement level of the students in that grade. See Appendix B.

in the introduction of new topics in third grade, and that continued to slow throughout the elementary school years (see Figure 2-5).

Between 1997 and 2001, Teachers Increased Instruction Time on Reading Comprehension and the Reading Skills Tested on the ITBS

Beginning in 1997, the Consortium on Chicago School Research added to its surveys a detailed set of questions, parallel to the mathematics questions, which asked reading and language arts teachers to report on the amount of classroom time they devoted to a variety of reading and language arts activities. Because these survey items were not asked in 1994, we do not have a prepolicy versus postpolicy comparison, however, as we saw with test preparation and with exposure to grade-appropriate mathematics content, it does not appear that the full effect of the policy took place immediately after it was implemented. Thus, by exploring changes between 1997 and 2001, we can understand the general trends that resulted from the policy's implementation (see *How We Measured Reading Alignment and Emphasis on Reading Comprehension* on page 49).

Figure 2-6
In 2001, Eighth-Grade Teachers Spent More Time on Grade-Level Mathematics and Less Time on Primary Grade Mathematics



Note: See Appendix B.

We assessed reading content alignment in two ways. First, we analyzed the proportion of time teachers reported spending on reading versus other language arts activities, such as writing, going to the library, practicing public speaking, and teaching study skills. We call this measure, “time spent on reading comprehension.” Second, we explored the degree to which teachers emphasized ITBS-assessed reading comprehension skills. This measure combines into a single scale (measure reliability=.83) language arts teachers’ reports of how much class time they spent on reading comprehension skills, such as analyzing and interpreting literature, differentiating fact from opinion, or drawing inferences. The measure is placed on a 10-point scale, with 10 indicating a high degree of content alignment with the ITBS. We call this measure “ITBS reading alignment.”

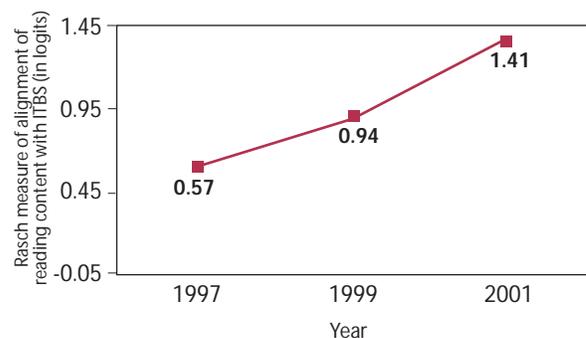
Both our measures of reading alignment and of time spent on reading versus other language arts activities suggest that Chicago Public Schools teachers placed more emphasis on reading comprehension, focusing specifically on the reading skills that the ITBS measured. The measure of reading alignment rose signifi-

cantly between 1997 and 1999, and then again in 2001, for a total change of more than one-half of a standard deviation (see Figure 2-7). We also saw a small increase in the proportion of language arts time that teachers reported devoting to teaching reading comprehension versus other language arts activities. In almost all grades except fifth, we saw significant increases in the proportion of language arts time teachers reported spending on teaching reading (see Figure 2-8). As with mathematics pacing, the largest increases occurred among seventh- and eighth-grade teachers.

We estimated trends in our measure of reading alignments with the ITBS (i.e., the extent to which teachers reported emphasizing the specific reading skills tested on the ITBS) once we adjusted for the entering reading test scores of students in that grade. Because the first- and second-grade ITBS test measures different reading skills than the third-grade reading test, we restricted this analysis to fourth through eighth grade (thus controlling for students’ third- through seventh-grade reading ITBS test scores). Even after accounting for the fact that students in 2001 were entering the upper grades with significantly higher ITBS reading scores, we found that seventh- and eighth-grade teachers, in particular, were much more likely

Figure 2-7

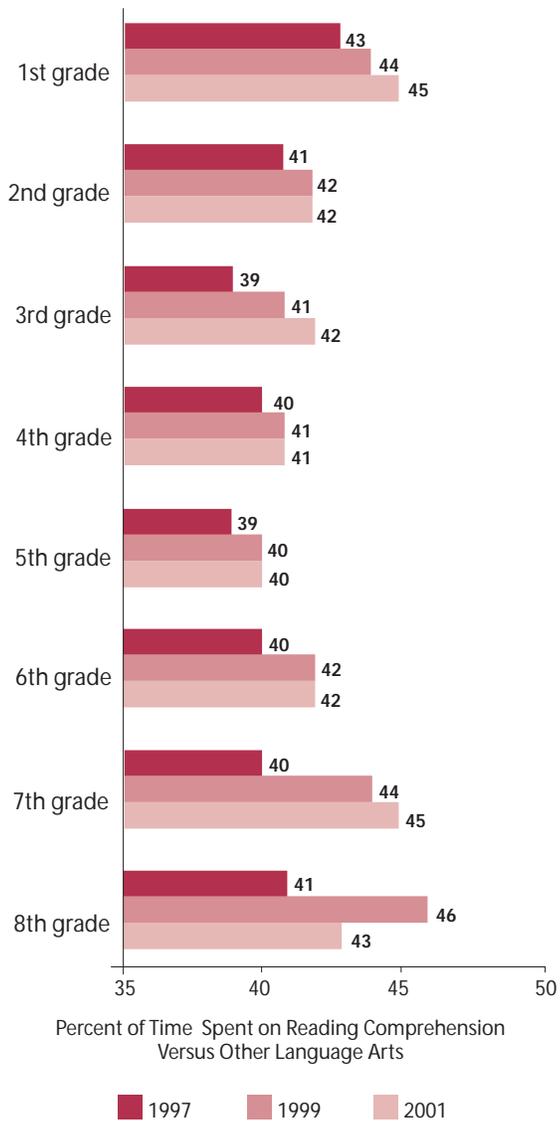
Teachers Reported Spending Much More Time Teaching ITBS-Tested Reading Skills 1997-2001



Note: Figure is scaled to show half a standard deviation above and below the 1997 mean. See Appendix B.

Figure 2-8

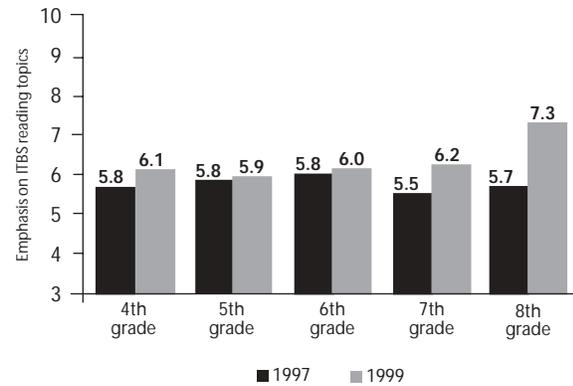
Seventh- and Eighth-Grade Teachers Reported the Largest Increases in the Time They Spent on Teaching Reading Comprehension 1997-2001



Note: Reading comprehension is the percent of time a teacher devotes to reading-comprehension topics in her classroom as opposed to writing or other language-arts activities, such as going to the library. For grades four through eight, where valid test-score data are available, the models also control for the incoming achievement level of the students in that grade. Estimates shown are from Hierarchical Linear Models (HLM) that control for the demographic characteristics of teachers responding to the survey (see Appendix B).

Figure 2-9

Seventh- and Eighth-Grade Teachers Significantly Increased Their Emphasis on ITBS-Covered Reading Topics



Note: Estimates shown are from a hierarchical linear model that controls for the demographic characteristics of the teachers and adjusts for the incoming reading test scores of students (thus results are shown only for fourth through eighth grades). Results are shown on a 10-point scale. See Appendix B.

to report aligning the content of their reading instruction with the ITBS (see Figure 2-9). We observed small and insignificant changes in the lower grades. We need to be careful in interpreting these results, however, because we do not have a prepolicy comparison. It might be that teachers in all grades shifted to greater emphasis on those skills measured by the ITBS between 1994 and 1997. But clearly, this shift continued in the seventh and eighth grades after 1997.

Differences across Schools in Mathematics Pacing and Reading Comprehension

In the previous section, we found that test preparation increased significantly in the lowest-performing schools in the system. We conducted a similar analysis to examine how increases in the time spent on grade-level mathematics material and on reading versus other language arts activities varied by the school's achievement level. This analysis also accounted for changes in the race, gender, education, years of teaching experience, grade level, and subject matter taught of the teachers responding to the survey over time, as well as

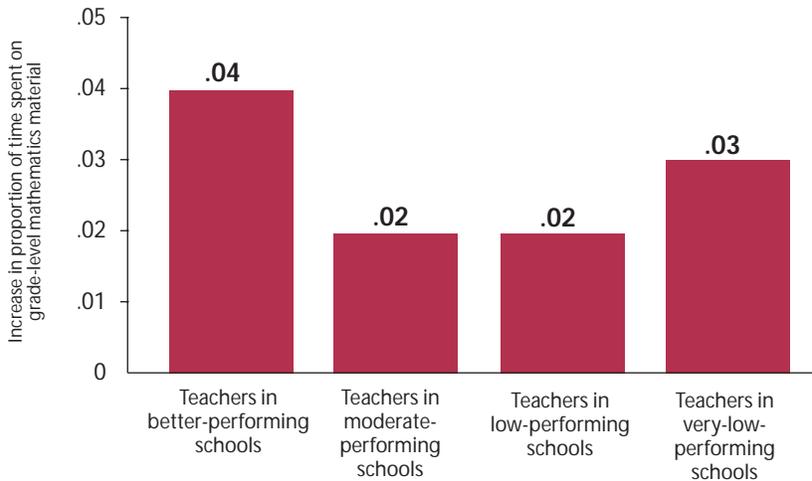
the racial composition of the school in which they worked. Thus, we estimated the changes in mathematics pacing and emphasis on reading comprehension that occurred in schools of different achievement levels if all schools had had similar teaching staffs and had students with similar racial backgrounds.

To restate, our measure of mathematics content emphasis represents the extent to which teachers in each grade reported emphasizing topics that were identified as grade appropriate. Mathematics emphasis increased across all schools, but the largest increases were observed among the lowest-performing schools (those with <15 percent of students reading at national norms) and among the highest-performing schools (see Figure 2-10). The 3 percentage-point increase among teachers in very-low-performing schools represents an increase of approximately one-third of a standard deviation, while the 4 percentage-point increase among teachers in better-performing schools represents an increase of two-fifths of a standard deviation.

It appears that teachers with the lowest-performing students felt significant pressure to increase the level of mathematics content they were covering in their class-

Figure 2-10

Teachers in Very-Low- and Better-Performing Elementary Schools Reported the Greatest Improvement in Time Spent on Grade-Level Mathematics Material (1994-2001)



Note: This chart indicates the estimated change in the proportion of time teachers in a particular grade devoted to grade-level material (i.e., material considered appropriate for students in that grade or the grade immediately above or below it). Estimates shown are from a hierarchical linear model that controls for the demographic characteristics of the teachers responding to the survey. For grades four through eight, where valid test-score data are available, the models also control for the incoming achievement level of the students in that grade (see Appendix B). For more detailed information on the categorization of schools by their achievement, see Figure 2-4.

rooms. This is an important finding, since these are the same schools where pacing has been shown to slow down the most in the upper grades. Yet, teachers in the better-performing schools, where they faced ostensibly little pressure to change the content covered in their mathematics classes, also spent significantly more time on grade-level material in 2001. Perhaps these schools had the greatest capacity to alter their curriculum, when faced with increased accountability and specific standards. There were no statistically significant differences by the racial composition of schools after accounting for differences across schools in their achievement levels in 1994.

There is no evidence that the shift in emphasis toward time on reading and reading comprehension was concentrated in the very-low-performing schools (see Figure 2-11). Thus, it appears that CPS teachers in all schools reacted to the signal created by high-stakes testing by aligning their curriculum more closely with the content of the test. No differences were found based on the racial composition of the schools.

How We Determined Whether Teachers Were Teaching “Grade-Level” Mathematics Content

In each survey year, elementary school teachers who taught mathematics were asked, “approximately what percent of your total instructional time is devoted to” the following 34 topic areas (see list below). Smith, Smith, & Bryk analyzed the ITBS to determine what topics in the list were associated with different levels of student performance.¹

The 34 questions about math content coverage were then divided into grade-level categories based on their level of difficulty. Difficulty levels reflect the percentage of students at a given grade level who were able to answer correctly an item associated with that topic on the ITBS. Based on teacher responses to individual items within each grade category, the proportion of time spent on material at each grade level was computed and standardized to equal 100 percent. For this analysis, the proportion of time spent teaching “on grade level” is defined as the proportion of time spent on material one grade below, at, or one grade above the grade taught. For a third-grade teacher, for example, the proportion of time spent on grade-level material includes time spent on second-, third-, and fourth-grade material. In the survey itself, items were grouped according to skill areas, such as “basic facts and concepts,” “measurement,” “numbers and operations,” “computation,” and “algebra,” and not according to the grade level of the material; thus, they appeared in a different order on the survey than shown here.

First-Grade Material

Counting objects and places in line
Identifying shape names and relationships
Naming and ordering numbers
Simple addition
Simple subtraction
Identifying operations
Solving word problems using addition, subtraction

Second-Grade Material

Coins and money
Reading a clock
Using number lines and rulers

Third-Grade Material

Multidigit addition
Multiplication facts
Multidigit subtraction
Measuring common objects
Counting by a factor greater than one
Converting between units of time

Fourth-Grade Material

Simple division
Place value and regrouping
Finding length, perimeter
Shifting between words and numerals
Solving word problems using multiplication, division

Fifth-Grade Material

Solving equations with one unknown
Estimating, approximating the closest number to solve a problem
Finding area and volume from pictures
Multidigit multiplication

Sixth-Grade Material

Long division
Solving inequalities
Associative, communicative, distributive properties
Rounding
Decimal place value
Primes, factors, multiples
Converting between units of measurement

Seventh-Grade Material

Solving ratio, proportion problems
Operations with decimals
Solving percent problems (find what percent, part, whole)
Operations with negative numbers
Exponents and roots
Operations with fractions
Absolute value

Eighth-Grade Material

Statistics
Probability
Equations of lines
Simplifying algebraic expressions
Graphing equations

Beyond Eighth-Grade Material

Solving two equations, two unknowns
Solving quadratics
Solving interest problems
Solving mixture and coin problems
Solving word problems
Solving distance problems

¹ Smith, Smith, & Bryk (1998).

How We Measured Reading Alignment and Emphasis on Reading Comprehension

In 1997, 1999, and 2001, elementary reading and language arts teachers were asked, “approximately what percent of your total instructional time is devoted to” the following topics (see the list to the right). In the survey, items were grouped according to the subject areas of reading, writing, and other language arts and appeared in exactly the order they are presented here. The items varied slightly from year to year. The items shown to the right are taken from the 2001 survey.

We developed our measure of *reading alignment* using Rasch analysis. The measure was created by combining multiple survey items to capture the degree to which teachers aligned their curriculum with topics covered on the Iowa Tests of Basic Skills (ITBS). The items in this measure include the emphasis given to things such as differentiating fact from opinion, drawing inferences, and analyzing and interpreting literature. These topics mirror closely the topics covered on the reading comprehension portion of the ITBS and around which the Summer Bridge summer-school curriculum is built.

We also created a way to measure the emphasis teachers placed on *reading comprehension* by dividing these topics into three categories: reading comprehension, writing, and other language arts topics; the proportion of time teachers spent on reading comprehension became the basis for the measure. As illustrated to the right, reading comprehension includes such things as recognition of important ideas and supporting details in a text, comprehension of facts and details, and analyzing and interpreting literature. Writing topics includes spelling; proper grammar, punctuation, and conventions; persuasive writing; narrative writing; and free or creative writing. Other language arts topics includes study skills, handwriting, word processing, public speaking, and listening skills. Teachers were able to choose among eight categorical responses regarding the amount of time they spent on each subject or skill:

- 1 = taught at a previous grade level
- 2 = will be taught at a future grade level
- 3 = 1 to 5 percent
- 4 = 6 to 10 percent

- 5 = 11 to 20 percent
- 6 = 21 to 35 percent
- 7 = 36 to 50 percent
- 8 = more than 50 percent

Teachers' answers were standardized to equal 100 percent, and the proportion of time teachers devoted to reading as opposed to writing or other language arts topics was explored.

Reading Comprehension

- Systematic phonics (decoding skills, phonemic awareness)
- Vocabulary acquisition
- Reading text with fluency (e.g., with speed and proper intonation)
- Understanding a basic story line or narrative (fiction)
- Analyzing and interpreting literature
- Recognition of important ideas and supporting details (expository texts)
- Differentiating fact from opinion
- Drawing inferences from expository texts
- Synthesizing ideas from several texts
- Comprehension of facts and details
- Identifying the main idea in a paragraph or text
- Remembering the sequence of significant events
- Understanding the author's perspective

Writing

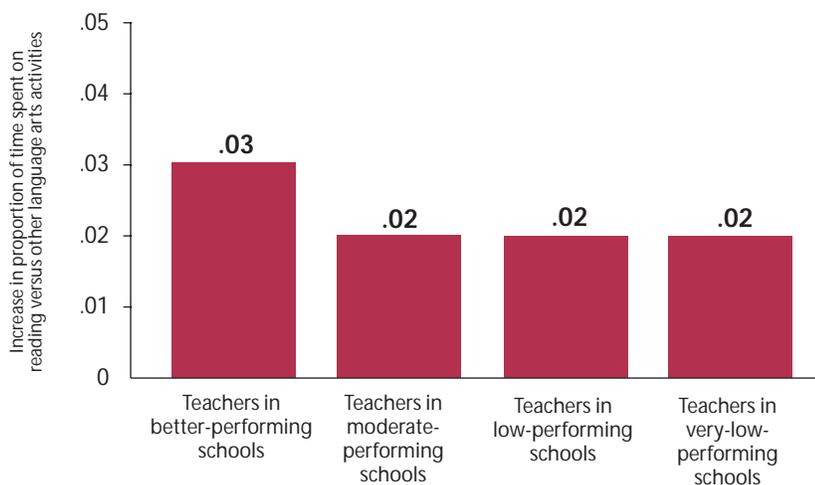
- Spelling
- Proper grammar, punctuation, and conventions
- Writing process
- Persuasive writing
- Expository writing
- Narrative writing
- Free or creative writing

Other Language Arts Topics

- Study skills
- Cursive handwriting and penmanship
- Word processing
- Finding information on the Internet
- Public speaking
- Listening skills

Figure 2-11

Teachers in All Schools Spent More Time on Teaching Reading Versus Other Language Arts Activities



Note: For more detailed information on the categorization of schools by their achievement, see Figure 2-4.

Interviewed Teachers Also Confirmed That They Aligned Curriculum to the ITBS

Longitudinal survey data suggest that many Chicago teachers shifted their skills focus in reading and mathematics toward the skills that were being assessed on the ITBS. One problem with survey data, however, is that we do not know whether teachers' reports reflect actual changes in behavior. One interpretation may be that teachers changed what they taught in order to raise test scores and better prepare their students for the ITBS. Another interpretation may be that teachers had become so familiar with what was on the tests and with the district's emphasis on reading, that they were more likely to report emphasizing this content even if they had made few actual changes in their classrooms (e.g., teachers know the right answer).

Interviewed teachers in low-performing schools were asked to describe the most effective strategies they found for raising students' test scores on the ITBS. Their responses suggest that many teachers have indeed made changes in what skills they teach and in their content coverage in order to prepare students for the ITBS. The changes were not confined to language arts and mathematics teachers. For example, one de-

partmentalized science teacher explained that as the ITBS approached, she switched to teaching mathematics:

For the last month or more, might be six weeks, I did math with all the full classes. I just stopped science . . . there was too much content to be covered, and she [the math teacher] was not able to [cover it all]. So I said, "Okay, whatever she is not able to cover, I'm going to cover before the test." . . . I did just math for six weeks . . . I did percents. I did mean, median, and mode. I did probability, which is a very interesting topic, and if you understand it, it's very easy to do. I did area. I did volume, perimeter, and ratios. (Sixth-grade teacher)

Many teachers talked specifically about feeling compelled to teach skills such as estimation or word problems or give more emphasis to nonfiction passages once they knew what was on the test. As several teachers explained:

I do things throughout the year, everyday, that are directly related to the IOWA [ITBS] test. So, like for example [quizzes] I always get a computation problem, problem-solving problem, an estimation problem, a graph problem, and a general-concept problem, maybe a couple more computations. . . . In the beginning of the year, it's a lot more computation. Towards the end of the year, it's a lot more word problems or estimation. (Eighth-grade teacher)

[Last year on] the IOWA test . . . there was one [passage] from a novel . . . but everything else is from *National Geographic*, science, social studies, . . . so I tried to incorporate other kinds of nonfiction material into the reading instruction. We don't have much around here that's nonfiction. (Sixth-grade teacher)

I . . . pick up my curriculum, so I know that my students have been exposed to everything that they're going to be tested on. . . . (Eighth-grade teacher)

As already noted in Chapter 1, most teachers did not view the restructuring of their curriculum negatively. As one sixth-grade teacher said:

[The policy] taught me that I need to teach smarter. And what I mean by that is, teach what they need to know to be successful and not waste a lot of time on things that they will not be tested on. I can do that after the ITBS if I feel it's important and they need that; then I would make sure they got it after the testing was over. . . . You know, it made me a better teacher because I can see what I need to teach. . . . If I know main idea is going to be three questions on that test, then I know I want to cover main idea.

There were a few teachers, however, who lamented that the policy had taken away from other things they would liked to have done in their classrooms:

There were things like special projects that I would have loved to have done with different units in science and social studies . . . writing projects . . . But you're trying to make sure that you reinforce those basic skills all the time. (Sixth-grade teacher)

[Has the policy influenced your content?] Yeah, like I said, . . . It's more skill oriented than practical life skills. And I would probably devote more time to the . . . practical aspects of math rather than just the skills. . . . It's just the time; there's so many skills that have to be covered. Every day is a new skill. (Eighth-grade teacher)

What Have We Learned?

Is there evidence that teachers spent more time preparing students for standardized tests and teaching students the content and skills that were on those tests? Taken together, there is a great deal of evidence that teachers changed what they taught as a result of increased accountability and high-stakes testing. Teachers spent more time preparing students to take standardized tests, placed a greater emphasis on



John Booz

reading comprehension, and spent more time on grade-level mathematics material. There are several important findings that are worth noting:

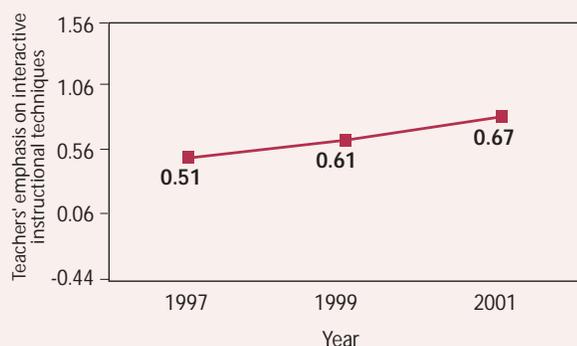
- Time on test preparation increased from 1994 to 1999, but then declined slightly in 2001. During the same period, teachers continued to increase the time spent on grade-level mathematics material and reading comprehension.

It is not surprising that time on test preparation increased after the implementation of high-stakes testing. A significant question raised in our analysis, however, is the extent to which shifts in the emphasis placed on test preparation reflect a short-term or long-term approach for teachers. Using direct test-preparation activities, such as administering practice tests or aligning classroom assessments, requires relatively little effort on the

Changes in Instructional Practice

This chapter focuses on the ways Chicago Public Schools teachers changed their instructional behavior in response to increased pressure to raise student test scores. We limited our analysis to whether teachers shifted their instructional emphases toward spending more time on test-preparation activities and reading and mathematics content emphases. A major concern about high-stakes testing policies is that not only will such policies influence the content covered in the classroom, but they may lead teachers to emphasize more traditional forms of instruction (e.g., lecture, drill, memorization, seat work, and worksheets). Additionally, these policies may encourage teachers to move away from interactive kinds of instruction that may produce more sustained learning, such as having students work on longer projects, brainstorm, debate ideas, or collaborate in groups. As Newmann, Bryk, and Nagaoka argue:

Between 1997 and 2001, Teachers Did Not Substantially Change How Often They Used More Interactive Instructional Techniques



Note: The measure used here is intended to capture the degree to which a teacher uses a more inquiry-based approach to teaching, including the use of group projects, discussion, debate, and brainstorming. The measure includes items that ask teachers to indicate how often they assign one-week projects, have students discuss or debate ideas for more than half the period, or have students work in cooperative groups. The measure was constructed using Rasch analysis, contains 13 items, and has an internal reliability of 0.81 logits. The figure is scaled to show half a standard deviation above and below the 1997 mean.

students who were in classrooms where teachers provided them with assignments that asked them to engage in more in-depth, constructive, and challenging inquiry did better on basic-skills tests than their peers who received more traditional assignments that focused on didactic learning.

The increasingly serious consequences for low student performance on standardized tests has been coupled with renewed attention to the question of how best to organize classroom instruction. Within the “back-to-basics” movement, it is widely believed that more sustained attention to didactic methods is essential. From this perspective, the best way to teach is to present students with the desired information and ask them to memorize it. Through various drills, exercises, and tests, students are expected to recall and repeat what they have memorized.¹

Research suggests the tradeoff between the mastery of basic skills and more authentic intellectual work may not be as difficult as it might first appear. In their 2001 report, Fred Newmann and colleagues looked at the standardized test performance of students who were asked to do more “authentic intellectual work”—work that asked students to construct knowledge, use a broad knowledge base, and that was more applied. This project was part of the Annenberg Research evaluation. They found that, contrary to popular assumptions,

part of teachers—especially since so many testing companies have ready-made test-preparation materials available for teachers to use. Altering curriculum and content coverage requires a relatively greater investment of time and energy on the part of teachers, since it involves designing new units and writing new lesson plans. Yet, content alignment may be somewhat more effective in raising test scores than direct test preparation

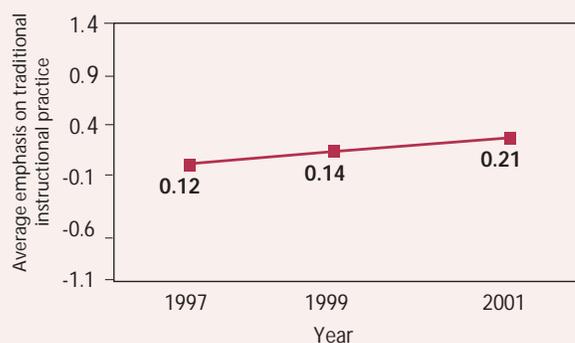
in the long run. Once students become familiar with the format of the test, additional time spent teaching test-taking strategies and taking practice tests likely has little added benefit, while devoting more time to teaching the specific skills tested on the exam may be more likely to lead to improved test performance. Teachers may begin to shift their emphasis towards content alignment and away from test preparation, both because they

Despite this finding, many are concerned that popular wisdom will prevail and that teachers will move away from, rather than toward, this kind of intellectually demanding work. For this reason, we thought it was important to examine trends in the measures of interactive and didactic instruction developed as part of the Consortium on Chicago School Research's work on instructional practice over time. We do not have prepolicy measures available for these two measures. Exploring changes between 1997 and 2001 can give us a sense of the general trends in these two measures, although the results should be interpreted with caution. The measures explored here were constructed so that they did not reflect increases in the time spent on test preparation over time; these increases are explored separately and in more detail within this report.

We explored two measures of instructional practice. Our measure of interactive instruction measures the degree to which a teacher uses discussion, hands-on activities, and student-centered projects in class. This measure includes items that ask teachers to indicate how often they assign one-week projects, have students discuss or debate ideas for more than half the period, or have students work in cooperative groups. The measure contains 13 items and has a measure reliability of .81. The second, a measure of didactic instruction, measures the degree to which a teacher uses tightly structured exercises and requires students to memorize, recite, and demonstrate facts, definitions, and procedures. This measure consists of eight items and has a reliability of .66. It asks teachers to indicate how often they ask students to memorize facts and procedures or complete workbook or textbook exercises in class.

These figures indicate that there has been no substantive shift in emphasis in either of these two measures since 1997. Increases in both measures were statistically significant, yet both measures increased by less than one-tenth of a standard deviation between 1997 and 2001. While teachers have not substantially increased their use of didactic instructional techniques as many had feared, neither has the policy encouraged teachers to use interactive techniques more often. It is important to keep in mind when interpreting these results that teachers have traditionally relied heavily on the use of didactic instructional techniques, and it is therefore unlikely that extremely large increases would be observed.

Between 1997 and 2001, Teachers Did Not Substantially Change How Often They Relied on Traditional Instructional Practice



Note: This measure explores the degree to which a teacher uses tightly structured exercises that require students to memorize, recite, and demonstrate facts, definitions and procedures. The measure was constructed using Rasch analysis and consists of eight items with an internal reliability of 0.66 logits. The figure is scaled to show half a standard deviation above and below the 1997 mean.

¹ Newmann, Bryk and Nagaoka (2001), 9.

no longer believe that the benefits of test preparation are as great as they once were and because the teachers themselves have had ample time to alter their curriculum to correspond more closely to the test.

- Test preparation increased in all grades, but the most dramatic changes in the time spent on grade-appropriate mathematics content and reading alignment occurred in the upper grades.

Upper-grade teachers made the largest behavioral changes. For example, they greatly increased instruction on grade-appropriate mathematics content, reading comprehension, and test preparation. Among third-grade teachers, the most substantial changes were increases in test preparation.¹⁰ Interestingly, teachers' behavioral changes corresponded to the observed changes in student test scores. The largest student test-score increases, since the effort to end social promotion began, occurred among sixth- and eighth-grade students. Smaller changes occurred in third-grade test scores. One interpretation of this trend might be that the strategies available to teachers in the upper grades were different and may have been more effective in raising test scores than those available to teachers in the primary grades. Primary-grade teachers may be faced with a different set of instructional problems in addressing the learning needs of very-low-achieving students than teachers in the upper grades. In third grade, for example, many low-performing students are still struggling to learn basic decoding skills in reading. The problem for these teachers is diagnosing successfully the source of students' reading difficulties. In the upper grades, however, it appears that one of the main problems may have been with pacing of classes and content coverage. As a result, it may have been easier for teachers in the upper grades to make

changes that would lead to increases in students' test scores than in the lower grades because the promotion initiative was not accompanied by investments in building teachers' knowledge and capacity to diagnose and address basic reading problems.

- Teachers in the lowest-performing schools changed the most.

Teachers in all schools aligned their classroom content to correspond closely to the ITBS content. Changes in mathematics content coverage and in the proportion of time spent on reading comprehension in language arts were observed across schools of all performance levels and racial compositions. However, the most substantial changes in mathematics content coverage were observed among the very-low-performing schools and the better-performing schools, where the problem of pacing was greatest before policy implementation. At the same time, most of the increases in test-preparation time were observed in the low- and very-low-performing schools. The better-performing schools in the system showed little change in the time they devoted to test preparation, despite the new accountability system.

Clearly, high-stakes testing sent a strong signal about the content teachers should cover in their classrooms, which was heeded by teachers across the system. Teachers in the very-low-performing schools responded the most aggressively with content changes, and they also increased substantially the time they spent on test preparation. It is not clear whether the response of teachers in very-low-performing schools reflected their own and their colleagues' responses to school accountability measures and the threat of academic probation, or whether their responses reflected individual teachers' responding to the need to

help students meet the test-score cutoffs established by the end of social promotion.

The responses of teachers in the very-low-performing schools demonstrate the dual-edged sword of high-stakes testing. On the one hand, advocates could point to these findings to argue that CPS's high-stakes testing policy worked to push instruction and increased opportunities to learn for schools with the lowest achievement. On the other hand, teachers in these schools also increased significantly the amount of time devoted to pure test preparation—suggesting that normal instruction had been narrowed.

Are the Changes Good or Bad: Some Final Thoughts

One of the greatest concerns raised about the use of high-stakes exams for accountability purposes was the degree to which it provided incentives for teachers to “teach to the test.” In this chapter, we looked at teaching to the test in two ways. First, we explored changes in the time spent on test preparation, such as teaching test-taking strategies and administering practice exams. Second, we explored the time spent on the content and skills covered by the exam.

Any district involved in high-stakes testing must understand that teachers will respond to pressure to raise student test scores by increasing time on direct test-preparation activities that may not necessarily lead to any real increases in student learning. There is a cost to this, and it must be factored into any evaluation of the benefits and costs of the policy. CPS teachers devoted a considerable amount of time to test

preparation before they were faced with increased accountability, and they spent even more test-preparation time postpolicy. This created clear instructional costs in terms of the time available for student learning. Such costs must be considered in light of any potential benefits of the increased accountability.

The findings that teachers devoted more class time to reading comprehension and grade-level mathematics material may not, however, be entirely undesirable. To the extent that the allocation of instructional effort in both reading and mathematics was undesirable prior to policy implementation, the postpolicy changes to content alignment can be viewed positively. This finding is particularly important since previous Consortium on Chicago School Research studies found that mathematics pacing in CPS slowed dramatically in the upper grades, especially in high-poverty schools.¹¹ Similarly, numerous educators have expressed concern about the poor reading comprehension skills of many CPS students. As was shown in Chapter 1, most teachers seemed to believe that this reallocation of instructional time was appropriate.

Reallocation of instructional effort is problematic to the degree that it increases learning opportunities only in tested subjects and skills, resulting in the exclusion of other subjects and skills that are deemed important. In the presence of increased accountability to raise mathematics and reading scores, subjects such as social studies, science, and art may be neglected. We were not able to look directly at evidence for this in this chapter although teacher interviews suggest that this may have happened, at least to some degree.



John Booz

CHAPTER

3

The Student's Perspective:

Trends in student reports of academic support from teachers and parents, engagement in school, and perceptions of academic expectations and efficacy

by

Susan Stone and Melissa Roderick

If we are to believe teachers' reports, a lot changed in the Chicago Public Schools (CPS) between 1994 and 2001. Teachers reported that the ending social promotion initiative made them more sensitive to their students' needs, motivated their students, and led parents to be more involved in their children's education. Many teachers believed that these policies influenced instruction in positive ways. And there is evidence that teachers, particularly in low-performing schools and in the upper grades, were spending more time on reading and grade-level mathematics material. At the same time, many teachers spent significantly more time preparing students to take standardized tests. How significant are these changes? And, most importantly, to what extent did they influence students' experiences of school? To address these questions, we turn to a different source—the students themselves.

This chapter focuses on three central questions:

1. Between 1994 and 2001, what changes occurred in sixth- and eighth-graders' reports of the support they received from teachers and parents, their involvement in their schoolwork, the challenge of their classroom environments, and their engagement in school?
2. After the institution of high-stakes testing, was there evidence that students with the lowest skills reported different levels of engagement and support in school?
3. Were changes in student experiences most pronounced in the schools that had been most affected by high-stakes testing (i.e., very-low-performing and low-performing Chicago schools)?

Looking at Trends in Student Reports Based on Consortium on Chicago School Research Surveys

In 1994, 1997, 1999, and 2001, the Consortium on Chicago School Research surveyed all sixth- and eighth-grade CPS students (see *Data Used in This Chapter*). Throughout this period, researchers at the

Data Used in This Chapter

This chapter uses survey data collected in 1994 (prior to the implementation of the ending social promotion policy) and in 1997, 1999, and 2001, (after the implementation of the policy) to explore trends in student reports of their school environments since the inception of the policy. In these years, the Consortium on Chicago School Research surveyed sixth and eighth graders (students in two of the three promotional grades) about many different aspects of their experiences in school and created Rasch measures of the student perception of teacher and parent support, academic challenge, and student engagement in school. These measures were equated across years so results of each of the survey administrations would be comparable. While the 1994 surveys were generally used as a baseline, in some instances there were no comparable items or measures available in 1994. When no information was available in 1994, changes from 1997 to 1999 were explored. There is evidence that the full effect of the policy did not take place until several years after high-stakes testing was implemented. Therefore, changes using surveys conducted in 1997 as a baseline were not unwarranted; these findings, however, should be interpreted cautiously.

Surveys were administered to all sixth- and eighth-grade students in the Chicago Public Schools. Responses were received from about 30,000 sixth and eighth graders in each survey year (about 50 percent of the total number of sixth and eighth graders in each of those years).¹ As noted in Chapter 1, of the elementary schools in the system, 56 percent were represented in 1994 survey responses, 88 percent were represented in 1997, and 80 percent in 1999. For analyses, we drew on survey data from sixth- and eighth-grade students in regular elementary schools who participated in surveys in 1994 and at least one additional survey year.

**Chicago Public Schools
Promotional Test-Score Cutoffs**

	1997	1998	1999	2000
Third Grade				
Cutoffs	2.8	2.8	2.8	2.8
National Norms	3.8	3.8	3.8	3.8
Sixth Grade				
Cutoffs	5.3	5.3	5.3	5.5
National Norms	6.8	6.8	6.8	6.8
Eighth Grade				
Cutoffs	7.0	7.2	7.4	7.7
National Norms	8.8	8.8	8.8	8.8

¹ Respondent students are not statistically different from the systemwide whole in terms of demographic or achievement characteristics.

Consortium have worked to identify a range of critical indicators based both on research on effective schools and on analyses of what indicators within schools are linked to higher performance in Chicago.¹ In particular, research has consistently found that students learn more and are more successful when educators provide high levels of personal support, often referred to as “personalism,” and strong expectations for students’ work, or “academic press.”² Positive trends in these measures, then, would suggest that classrooms are becoming more conducive to student learning. Other critical indicators that would support positive achievement include time spent on homework completion, parental support for schoolwork, involvement in after-school activities, and engagement in classwork.³ In this chapter, we draw upon this previous research and the Consortium’s longitudinal database to look at trends in students’ reports of:

- The personal support they received for their schoolwork from teachers and parents;
- Their level of work outside of class time, including participation in after-school activities and time spent on homework;
- Their perceptions of the academic focus of their classrooms; and
- Their level of engagement in school.

We looked at trends over time in students’ reports by both a student’s level of achievement and the achievement level of the school the student attended. If the ending social promotion initiative resulted, as proposed, in teachers paying more attention to students with the lowest skills, we would expect to see more substantial changes over time in these students’ reports of their experiences in school. On the other hand, some critics argue that students with the lowest skills will likely disengage from school and schoolwork if they perceive that standards are outside their reach.⁴ In short, both views suggest that students of differing skill levels might experience the policy differently. Therefore, we grouped students into achievement categories based on the estimated risk of retention they



John Booz

would have faced given the 1997 promotion-gate test-score cutoffs (see *A Risk Category Approach* on pages 60-61). For example, sixth graders in 1997 had to reach a 5.3 cutoff for promotion. Grade level on national norms for a sixth grader was 6.8. This test-score cutoff was 1.5 grade equivalents below grade level on national norms. We characterized students in sixth grade at “high risk” of retention if their estimated fifth-grade reading test score placed them 1.5 years or more below the 5.3 cutoff (or three years or more below grade level). A sixth grader was characterized as being at “moderate risk” if his estimated prior year test score was between two and three years below grade level and at “low risk” if between two years to one year below grade level. Finally, sixth graders were characterized at “no risk” if their test score was one year below to grade level, ensuring that they had already met the promotion criteria, and “at grade level or above” if their estimated prior test scores placed them reading at 6.8 or above (over a year above grade level on entry into sixth grade). Of course, students in 1994 did not have to

A Risk Category Approach

This chapter looks at trends over time in student reports by both students' achievement level and the achievement and racial composition of the school they attended. If high-stakes testing shaped either the level of support students received or their attitudes toward their schoolwork, we would expect to see more substantial changes over time in the experience of students with the lowest skills. In order to investigate this question, in 1994, 1997, 1999, and 2001, we compared the responses of students with similar reading test scores. We grouped students by their predicted test

scores in the year before they entered that grade (e.g., by a student's estimated fifth-grade test scores for students in sixth grade).

The Chicago Public Schools' Test-Score Cutoffs, from *Ending Social Promotion: Results from Summer Bridge*

Under the Chicago Public Schools' policy, third, sixth, and eighth graders need to meet test-score cutoffs on the Iowa Tests of Basic Skills (ITBS) in reading and mathematics in order to be promoted. The test-score cutoffs are set using the grade equivalent (GE) metric. One month corresponds to 0.1 GE; therefore, 10 months equal one academic year or one GE. The ITBS reports results in GEs on national norms where a student is considered on grade level if, when taking the test in the eighth month of the school year, she obtains a score of that grade plus eight months. Thus, a third grader is considered to be reading at grade level on national norms if her ITBS score is a 3.8.

Between 1997 and 2000, the CPS promotion 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, which was 1.5 years below grade level at 6.8. In 2000, this was raised to 5.5. The eighth-grade cutoff was initially set at 7.0, 1.8 years below grade level at 8.8. In 1998, the cutoff for eighth grade was raised to 7.2. In 1999, it was raised to 7.4, and it was raised again, to 7.7, in 2000.

Our analysis, therefore, encompassed four different cohorts of children—those who were in the sixth and eighth grades in the school years 1994, 1997, 1999, and 2001. In order to make these different cohorts as comparable as possible, we placed students into five categories based on their "risk of retention" given the 1997 cutoffs for promotion set by CPS. The CPS test-score cutoffs are based on a student's scores on the reading and mathematics portions of the Iowa Tests of Basic Skills (ITBS) and are measured in "grade equivalents." A student who is at "grade level" on the ITBS when the test is given at the end of the school year would receive a score of that grade level plus eight months (6.8 for sixth grade and 8.8 for eighth grade). In 1997, the promotion test-score cutoffs in both reading and mathematics were set at 5.3 for sixth grade and 7.0 for eighth grade (see *The Chicago Public Schools' Test-Score Cutoffs, from Ending Social Promotion: Results from Summer Bridge*). The Chicago Public Schools has raised the standard for promotion in each year in the eighth grade. The promotion test-score cutoff

face the promotion criteria; in other words, a high-risk student in 1994 did not face the threat of retention. By grouping students in every year the same way, however, we can look at trends in reports by students of similar achievement levels.

As we noted in the introduction, students at different grades faced different cutoffs. Our use of student "distances" from the test-score cutoff allows for comparison of students who were affected similarly by the policy across grades. In estimating trends among students of different achievement groups, our analysis also took into account other student characteristics, includ-

ing gender, race, ethnicity, students' ages, whether a student was retained in a given survey year, whether or not a student was excluded from testing, and whether a student was enrolled in bilingual education. These statistical adjustments allowed us to estimate trends in the responses of students of different achievement levels accounting for the fact that the demographic characteristics of CPS have been changing over time (see *Why Adjust Student Responses for Changes in the Demographic and Achievement Characteristics of Students and Schools?* on page 62).

was 7.0 in 1997 and was raised to 7.4 in 1999 and fully 8.0 in 2001. Clearly, students in 1994 were not subject to the promotion policy, thus, “high-risk students” in 1994 represent students who would have faced a high risk of retention given their reading test scores if the 1997 policy had been in place. And, eighth graders in 1999 and 2001 faced slightly more stringent criteria for promotion. In short, these groupings facilitated our ability to identify trends among similarly performing students over time.

Students were categorized at “high risk” of retention if their prior year’s predicted test score was greater than 1.5 grade equivalents below the test-score cutoff for promotion (5.3 for sixth grade). Thus, in 1997, a student who entered sixth grade at high risk (1.5 grade equivalents below the 5.3 required for promotion) had an estimated end-of-year reading test score more than three years below grade level. “Moderate-risk” students were defined as students whose reading scores were estimated to be between .5 (half a year) and 1.5 years below the test-score cutoff on entry into the grade, and “low-risk” students were defined as those students who were slightly above or below the cutoff (.5 below to .5 above). Students were defined as “no risk” if their prior year test score was already above the cutoff but below grade level, and “at grade level” if their prior year test score was at or above grade level.

Risk Category	Combined Year (94, 97, 99, 01) Totals		Description
	Sixth grade (n=71,923)	Eighth grade (n=65,257)	Predicted fifth- or seventh-grade test score is . . .
High risk	12% (8,248)	14% (9,421)	1.5 years below the 1997 test cutoff (3 years or more below grade level)
Moderate risk	31% (22,531)	25% (16,427)	.5 to 1.5 years below the 1997 test cutoff (2 to 3 years below grade level)
Low risk	29% (21,380)	26% (17,022)	.5 below to .5 above the 1997 cutoff (1 to 2 years below grade level)
No risk	15% (11,125)	19% (12,076)	.5 to 1.5 above the 1997 test cutoff (1 year below to grade level)
Grade level	8% (5,464)	13% (8,606)	1.5 above or greater than the 1997 test cutoff (grade level to above grade level)
(Missing test scores)	4% (3,175)	3% (1,714)	

Throughout this time period, test scores were rising in the Chicago Public Schools. This meant that between 1994 and 2001, there were fewer students in the high- and moderate-risk categories, and more students in the no-risk and grade-level categories. For example, in 1994, 14 percent of sixth graders could be considered at high risk of retention, meaning that their fifth grade test scores were at least 1.5 years below the test cut off, and 33 percent were considered at moderate risk. By 2001, 8 percent of sixth graders were considered at high risk, and 29 percent at moderate risk. The smaller numbers of students with very low test scores most likely means that we are underestimating changes in students’ responses; we would expect that our very low test-score groups would be comprised of more students with more problematic school performance over time.

Changes in Students’ Perceptions of Support: Support from Teachers and Parents, and Involvement in After-School Programs

In the first chapter, we found that teachers felt that the ending social promotion policy had made parents more concerned about their children’s school performance and had made teachers more sensitive to individual students’ needs and problems. In this section, we look at trends in three critical measures of academic support—personal support for schoolwork from teachers, support from parents for schoolwork, and involvement in after-school academic support programs.

Trends in Personal Support for Schoolwork from Teachers

In each survey year, the Consortium’s sixth- and eighth-grade surveys ask students to respond to a series of questions about the personal support they receive from their teachers for their schoolwork. Students are asked, for example, the extent to which they believe their teacher is willing to give them extra help, believes they do well in school, or notices if they have trouble learning something. (for a full list of questions, see *How Support from Teachers for Their Schoolwork (Personalism) Was Measured* on page 63). Students’ answers are then combined into a summary measure on a one to 10 scale that can be compared across time.

Why Adjust Student Responses for Changes in the Demographic and Achievement Characteristics of Students and Schools?

The focus of this chapter is on trends over time in students' academic support and experiences in school. However, if the demographic characteristics of students were also changing during this time period, we might see trends in student attitudes that are unrelated to their achievement levels. For example, if Latino students are more likely to be positive about their teachers, we would expect that student reports of personal support for their schoolwork would be higher in 2001 than in 1994 because there were more Latino students in the Chicago Public Schools in 2001 than in 1994. By controlling for the demographic and achievement characteristics of students at each time period, we can adjust our estimate of trends in student attitudes for other underlying changes in the demographic composition and achievement levels of students and schools during this period.

When looking at overall trends among students of different achievement groups, our analysis takes into account a student's gender, race and ethnicity, age relative to their grade level, whether a student's test score was excluded from the promotion policy, whether they were retained, whether the student was enrolled in bilingual education, the year in which they participated in the survey, and an interaction term of the student's risk category and survey year.

We also included information on the characteristics of the school, including the percentage of students who were excluded from testing, the racial composition of the school (predominantly African-American, predominantly Latino, mixed race, predominantly minority, or integrated) and baseline (1994) school achievement levels. Thus, trends in this chapter can be interpreted as trends over time of students with similar achievement levels (e.g., students with very-low-achievement test scores) if the demographic characteristics of students (race and ethnicity, gender, age, test exclusion and retention status, and participation in bilingual education) and schools had remained similar over time (see Appendix C for the model).

In the first set of analyses, we estimated trends in student responses regardless of what kind of school they attended. In the second set of analyses, we estimated trends in student reports for students of different risk levels, controlling for the demographic and achievement characteristics of their schools. The goal of this second set of analyses was to disentangle how much of the trends we observed were driven by changes in student responses regardless of the school they attended and how much could be attributed to the fact that students attended schools with varying racial and achievement characteristics. For example, we could observe that students with very low skills reported more personal support for their schoolwork either because low-achieving students in all CPS schools received more attention or because very-low-performing schools were paying more attention to all of their students, regardless of their achievement levels. Thus, in our second set of analyses we estimated trends in student reports by both student achievement levels and by the achievement level of their school.

In both sets of analyses, we relied on a statistical method called Hierarchical Linear Modeling to estimate how the reports of students of different achievement levels changed once we accounted for the fact that students of all achievement levels might have become more or less positive in higher- versus lower-performing schools.

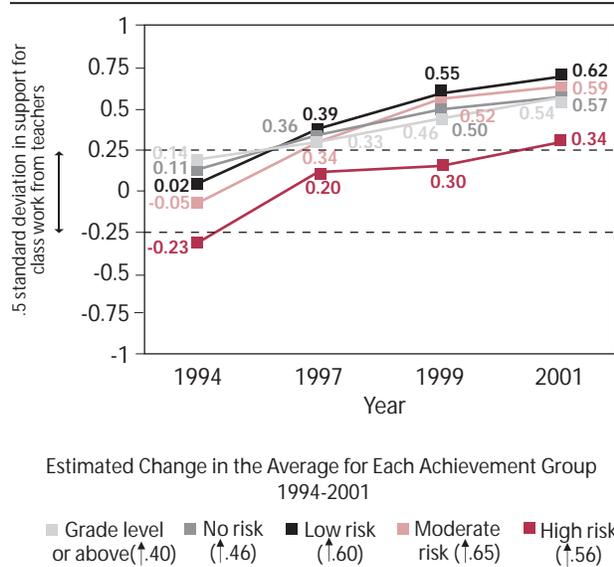
In 1994, students with the lowest achievement test scores (high-risk students) reported significantly less personal support for their schoolwork from their teachers than students whose test scores placed them close to or above grade level (no risk to above-grade-level students). In the sixth grade, all groups were more positive in 2001 than in 1994 about the level of personal support they received from their teachers (see Figure 3-1). However, the most dramatic increases occurred among students with the lowest skills. Indeed, among sixth graders at high risk of retention, in 2001, the average measure of their reports of the personal support they received from teachers for their schoolwork was fully .56 standard deviations higher than the average of students with similar test scores in 1994. In 2001, the average measure of personal support from teachers for schoolwork was .65 standard deviations higher among students at moderate risk of retention than the average of their counterparts with similar test scores in 1994. This change meant that by 2001, the initial gap between low- and higher-achiev-

ing students in their perception of support from teachers had closed significantly. Low-achieving sixth graders in 2001 were as positive about the academic support they received from teachers as were the highest-performing students in 1994.

The same trend was observed among eighth graders with the lowest test scores (see Figure 3-2). Eighth graders at high, moderate, or low risk of retention reported much higher levels of personal support from teachers for their schoolwork in 2001 than in 1994. Eighth graders with the highest test scores, however, reported smaller increases in levels of support from teachers in 2001 than in 1994. As a result, while in 1994, eighth graders with the lowest skills reported significantly lower levels of personal support from teachers for schoolwork, by 2001, high- and moderate-risk students actually reported higher levels than their higher-achieving counterparts.

Figure 3-1

Sixth Graders Reported Much Higher Levels of Personal Support from Teachers for their School Work (1994-2001)



Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall 1994 average. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

How Support from Teachers for Their Schoolwork (Personalism) Was Measured: Core Questions Asked of Students in 1994, 1997, 1999, and 2001

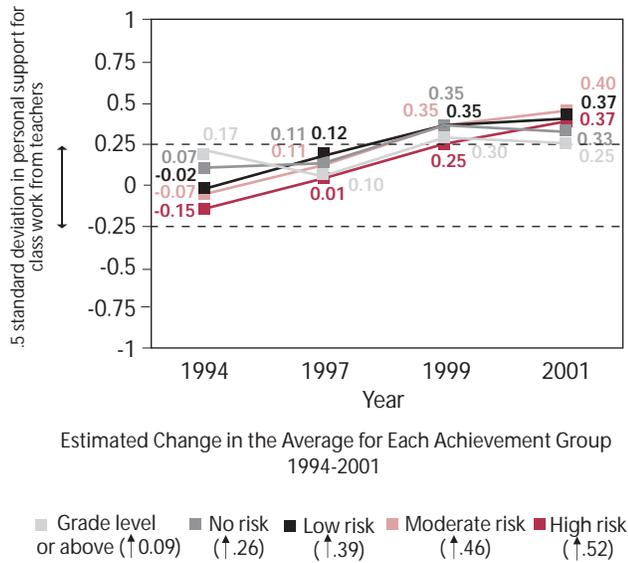
How much do you agree (strongly disagree, disagree, agree, strongly agree) with the following statements about your reading/language arts or mathematics teacher?

- Is willing to give me extra help on schoolwork if I need it.
- Really listens to what I have to say.
- Helps me catch up if I am behind.
- Notices if I have trouble learning something.
- Believes I can do well in school.
- Relates this subject to my personal interests.

In 1994, the measure of personalism developed by the Consortium included additional questions that were later dropped to shorten the survey and improve the psychometric properties of the measure. In order to make the measure of personalism comparable across survey years, difficulties of items common across years were "anchored" on a single set of values, and the measures were then produced.

Figure 3-2

Eighth Graders with the Lowest Skills Reported Much Higher Levels of Personal Support from Teachers for Their School Work (1994-2001)



Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall 1994 average. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

Did changes in student reports differ across schools? The increase in high-risk students' perceptions of the personal support they received from their teachers for their schoolwork suggests that since 1994, these students were experiencing increases in personal support for their schoolwork. Students with low test scores were concentrated in the lowest-performing schools, and we know that these schools were under significant pressure to raise student test scores. The fact that students with low achievement were more likely to attend low-performing schools raises the question: To what extent did the increase we observed in students' reports of personal support for schoolwork reflect the fact that teachers in low-performing Chicago schools might have become more supportive of their students? Or, was the increase in high- to moderate-risk students' perception of support from teachers also observed across schools of different achievement levels?

We used a multivariate analysis to estimate how student reports changed in the prepolicy versus postpolicy period in both schools of different achievement levels and among students of different risk levels (see *Why Adjust Student Responses for Changes in the Demographic and Achievement Characteristics of Students and Schools?* on page 62). For this analysis, we looked only at the prepolicy (1994) versus average postpolicy (1997, 1999, and 2001) difference in our measure of personal support for schoolwork. Thus, our analysis compared student reports in 1994 to the average of student reports in all of the postpolicy survey years (1997, 1999 and 2001). As in Chapter 2, we grouped schools on the basis of their achievement level in 1994 (prior to the policy). We called a school “very low performing” if less than 15 percent of their students in 1994 were reading at national norms, and “low performing” if between 15 and 24 percent of their students in 1994 were reading at or above national norms. We called schools “moderately performing” if between 25 and 35 percent of their students in 1994 were reading at or above grade level at national norms. “Better-performing” schools had more than 35 percent of their students reading at or above national norms.

Figures 3-3 and 3-4 show the estimated change in the level of academic support from teachers reported by students at moderate risk of retention if those students attended schools of different achievement levels. Students in low-performing schools, regardless of their achievement level, showed slightly larger increases in their perception that their teachers paid attention to their learning. However, students in all schools reported significant increases. Thus, the overall large increase we observed in low-achieving students' perception of their teachers' support for schoolwork reflected both the fact that all students in the schools under the most pressure to raise test scores reported larger increases in support for their schoolwork, and the fact that low-achieving students across the board reported more support for their schoolwork from teachers.

How Parental Support for Schoolwork Was Measured: Questions Asked of Students in 1994, 1997, 1999, and 2001

During this school year, how often (never, one to two times, three to five times, more than five times) have you discussed the following with your parents or other adults living with you?

- Selecting courses or programs at school
- School activities or events of interest to you
- Things you've studied in class
- Going to college
- Your grades

How often (never, once in a while, most of the time, all of the time) does a parent or other adult living with you

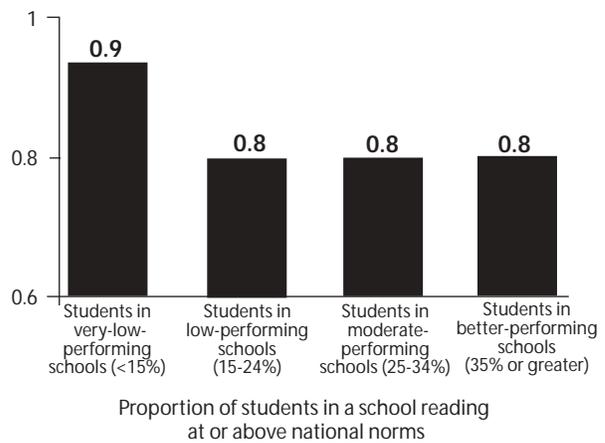
- Help you with your homework?
- Check to see if you have done your homework?
- Praise you for doing well in school?
- Talk about why you were not doing homework?
- Encourage you to work hard at school?
- Encourage you to take responsibility for things you have done?

Trends in Student Reports of Parental Support for Schoolwork

An important source of academic support for a student is the extent to which parents are involved in monitoring and supporting their child's education. In each survey year, sixth and eighth graders were asked how often they discuss school activities and things studied in class with their parents, how often their parents encourage them to work hard in school or praise their work in school, and how often their parents help with or check homework and discuss their school performance (see *How Parental Support for Schoolwork Was Measured*).

As with personal support for schoolwork from teachers, students with the lowest skills reported much higher levels of parental support and attention to their schoolwork in 2001 than in 1994. In 1994, high- and

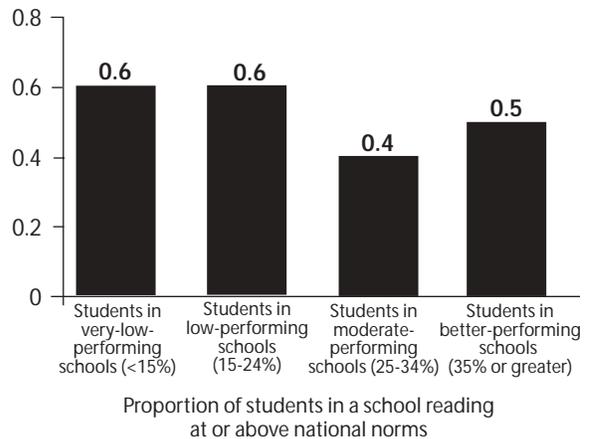
Figure 3-3
Sixth Graders in the Lowest-Performing Schools Reported the Largest Increases in Personal Support from Teachers for Their Schoolwork After 1997



Note: This graph shows the estimated level of students' reports of academic and personal support from teachers when estimating effects for students of different achievement levels and different schools after we have accounted for school characteristics. The model includes additional information on students' demographic characteristics at each time point. In 1994, 26% of moderate risk sixth-grade students attended very-low-performing schools (<15% of students reading at grade level on national norms) at grade level at national norms, 45% attended low-performing schools (15-24% at grade level), 14% attended moderate-performing schools (between 25-34% at grade level) and 15% attended better-performing schools (35% or greater at grade level).

Figure 3-4

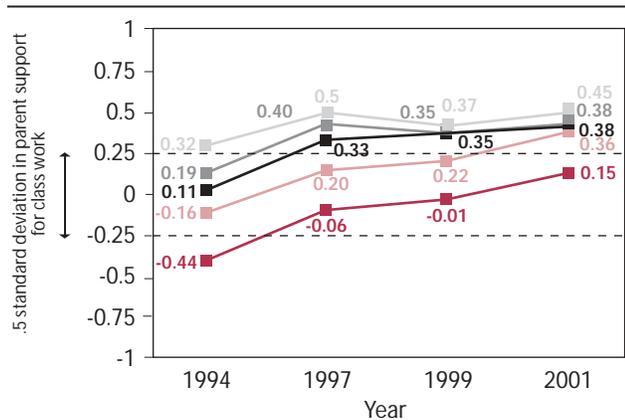
Figure 3-4
Eighth Graders in All Schools Reported Increases in Personal Support from Teachers for Their Schoolwork after 1997



Note: This graph shows the estimated level of students' reports of academic and personal support from teachers when estimating effects for students of different achievement levels and different schools after we have accounted for school characteristics. The model includes additional information on students' demographic characteristics at each time point. In 1994, 26% of moderate risk sixth-grade students attended very-low-performing schools (<15% of students reading at grade level on national norms) at grade level at national norms, 45% attended low-performing schools (15-24% at grade level), 14% attended moderate-performing schools (between 25-34% at grade level) and 15% attended better-performing schools (35% or greater at grade level).

Figure 3-5

Low-Achieving Sixth Graders Reported Much Higher Levels of Parental Support for School Work (1994-2001)



Estimated Change in the Average for Each Core Achievement Group 1994-2001

Grade level or above (↑0.13) No risk (↑0.19) Low risk (↑0.27) Moderate risk (↑0.52) High risk (↑0.59)

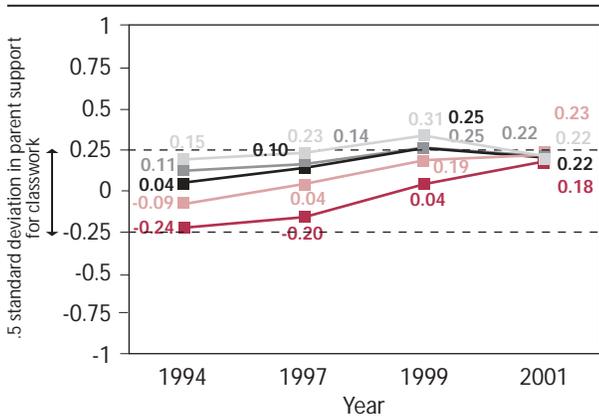
Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall average of 1994. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

moderate-risk students reported very low levels of parental monitoring and support for their schoolwork (see Figures 3-5 and 3-6). But by 2001, sixth and eighth graders with low skills (high- and moderate-risk students) reported much higher levels of parental involvement than in 1994. Smaller increases were observed among students with achievement test scores at or above grade level (no risk and grade-level students). As a result, while parental monitoring and support of schoolwork differed dramatically across students of different achievement levels in 1994, by 2001, there were significantly fewer differences in the extent to which students of different achievement levels reported that their parents were involved in checking and discussing schoolwork with them on a regular basis.

Across-school differences. Increases in parental support for schoolwork among low-achieving students were concentrated in the lowest-performing schools. Students in low-performing schools did report increases in parental support for their schoolwork, and this increase was significantly larger than the average increase we estimated in higher-performing schools, with 25 percent or more reading at national norms. We show results for the sixth graders who were at moderate risk of retention, across schools of different achievement levels (see Figure 3-7). Similar patterns were observed in the eighth grade.

Figure 3-6

Low-Achieving Eighth Graders Also Reported Higher Levels of Parental Support for School Work (1994-2001)



Estimated Change in the Average for Each Core Achievement Group 1994-2001

Grade level or above (↑0.07) No risk (↑0.11) Low risk (↑0.18) Moderate risk (↑0.32) High risk (↑0.43)

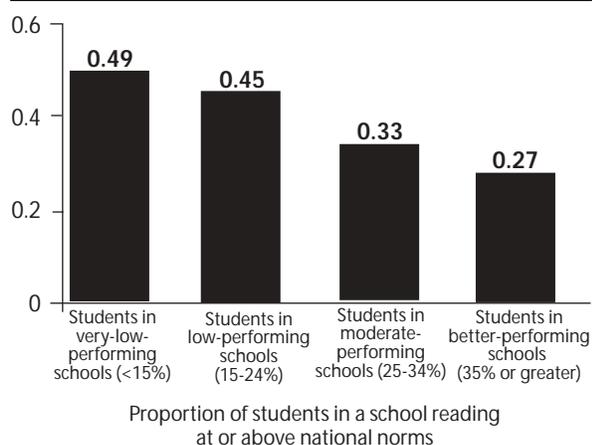
Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall average of 1994. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

Trends in After-School Participation

Direct parental involvement and students' perceptions of their academic and personal support from teachers are two critical measures of the social and academic support available to students. After-school participation is also an indicator of whether a student is involved in the school and making academic achievement a priority. As discussed in Chapter 1, one of the key academic supports available to students was the Lighthouse after-school program. Lighthouse began in 1997 and expanded rapidly. By 2001, the Lighthouse after-school program served 125,000 students in 400 elementary schools, operating, on average, for 70 days (or 14 weeks) each year. In each survey year, the Consortium student surveys asked, "This year, how often

Figure 3-7

Sixth Graders in the Lowest-Performing Schools Reported Significantly Higher Parental Support After 1997



Note: This graph shows the estimated level of students' reports of academic and personal support from teachers when estimating effects for students of different achievement levels and different schools after we have accounted for school characteristics. The model includes additional information on students' demographic characteristics at each time point. In 1994, 26% of moderate risk sixth-grade students attended very-low-performing schools (<15% of students reading at grade level on national norms) at grade level at national norms, 45% attended low-performing schools (15-24% at grade level), 14% attended moderate-performing schools (between 25-34% at grade level) and 15% attended better-performing schools (35% or greater at grade level).

(never, once in a while, once a week, almost everyday, everyday) have you attended after-school programs for help with schoolwork?" For ease of interpretation, we have reported the percentage of students who reported that they attended an after-school program almost everyday or everyday. Similar trends are observed if we look at the average of students' answers to this question. Because the Lighthouse program did not operate all year and because Consortium surveys were completed later in the school year, this question might underestimate the percentage of students who actually participated in after-school programs. In addition, initially some schools used Lighthouse resources to extend the school day, in which case students may not have perceived a transition to a formal after-school program.

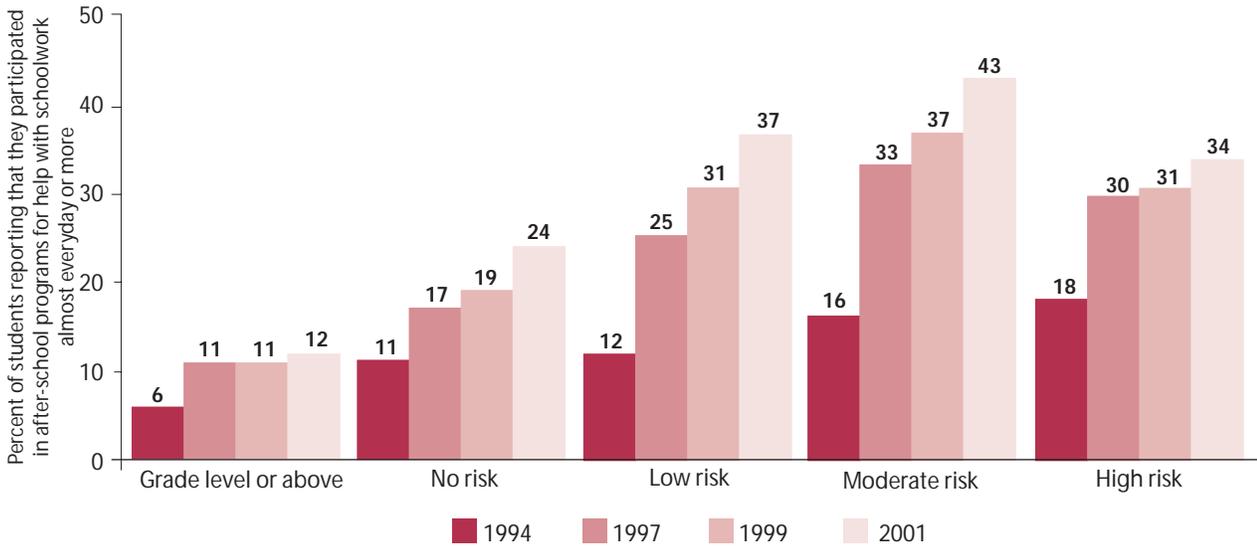
Between 1994 and 2001, there was a marked increase in the proportion of sixth and eighth graders who reported regularly participating in after-school programs for help with schoolwork. Students with low

In 1994, only 16 percent of eighth graders at moderate risk reported attending an after-school program almost everyday or everyday. By 2001, 43 percent of eighth graders at moderate risk and over one-third of eighth graders at high risk of retention reported regular attendance in after-school programs for help with schoolwork.

but not the lowest achievement (moderate- and low-risk students) showed the greatest increases in reports of after-school participation, particularly in the eighth grade. In 1994, only 16 percent of eighth graders at moderate risk reported attending an after-school program almost everyday or everyday (see Figure 3-8). By 2001, 43 percent of eighth graders at moderate risk and over one-third of eighth graders at high risk of retention reported regular attendance in after-school programs for help with schoolwork. Similarly, the proportion of sixth graders at moderate to low risk of retention who reported regular participation in after-school programs doubled between 1994 to 2001 (see Figure 3-9). Interestingly, sixth graders with the lowest skills (high-risk students) reported frequent participation in after-school programs in 1994, and this had risen slightly by 2001. We do not know if the higher rates (25 percent) of participation in after-school programs in 1994 by sixth graders reflected a pre-existing program. It is important to note that by 2001, the highest rates of after-school participation were reported by moderate- as opposed to high-risk students.

Figure 3-8

Many More Eighth Graders Reported Participating Regularly in After-School Programs



This is somewhat surprising since the Lighthouse after-school program was intended to target students at risk of retention. This higher rate of after-school participation among students with low but not the lowest skills might reflect the fact that schools have a hard time getting the most at-risk students to attend after-school programs. On the other hand, it might reflect a triaging strategy whereby schools focused resources on students they felt might benefit most from participation.⁵

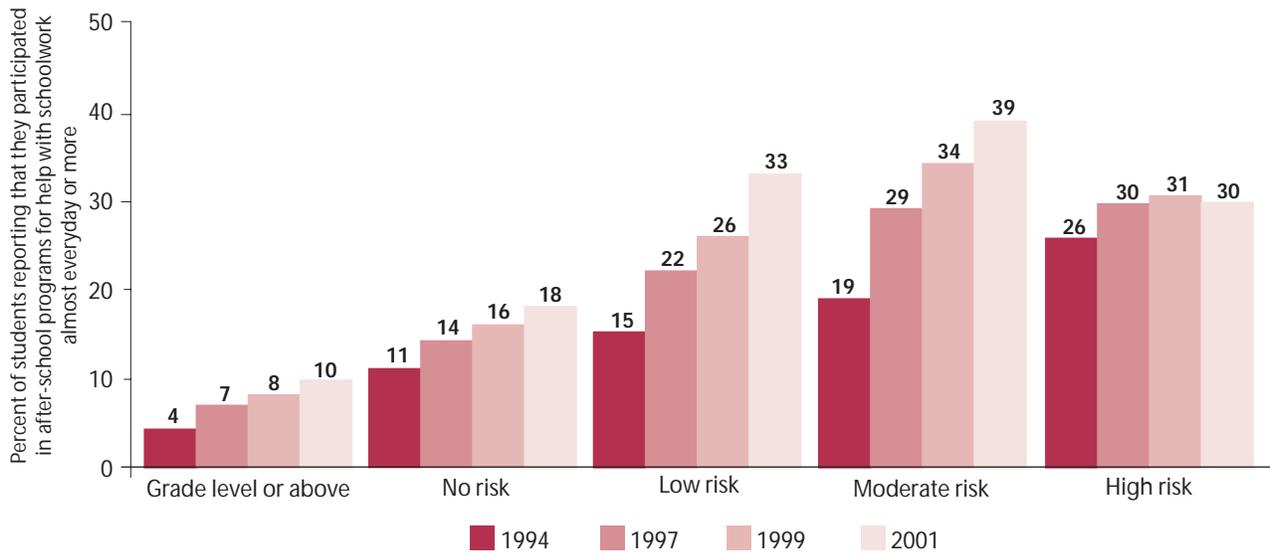
The Lighthouse after-school program was initially targeted at very-low-performing schools. Not surprisingly, student reports of after-school participation after 1997 were much higher in low- and very-low-performing schools, although students in all schools reported significant increases in after-school attendance (see Figure 3-10). We show across-school differences in trends in after-school participation among moderate-risk eighth graders. Results for sixth graders are similar.

Trends in Academic Expectations: Academic Press and Homework Completion

If students experienced higher levels of support for their schoolwork from parents and teachers, did they also experience concomitant increases in performance expectations? In every survey year, the Consortium has asked students a detailed set of questions designed to tap students' perceptions of the academic expectations or press of their teachers and schools. "Academic press" is often defined as the extent to which school members experience a normative emphasis on academic success and conformity to standards of achievement.⁶ Our summary measure combined students' answers to questions regarding whether their primary subject teachers expected them to complete homework every night, expected them to do well in school, cared if the student received bad grades and did his/her best, thought it was important to do well, and encouraged the student to do extra work when she didn't understand the material.⁷

Figure 3-9

Many More Sixth Graders Reported Participating Regularly in After-School Programs



Students’ perceptions of the academic expectations and press they received from teachers varied significantly across achievement levels in 1994. In that year, the average summary measure of high-risk students’ reports of their teachers’ academic expectations was nearly half of a standard deviation lower than among

students with achievement test scores at or above grade level.

Unlike our measure of personal support for their class work from teachers, there was no consistent increase in students’ reports of academic expectations from teachers between 1994 and 2001 in either the

How Academic Press Was Measured

Questions Asked of Students in 1994, 1997, 1999, and 2001

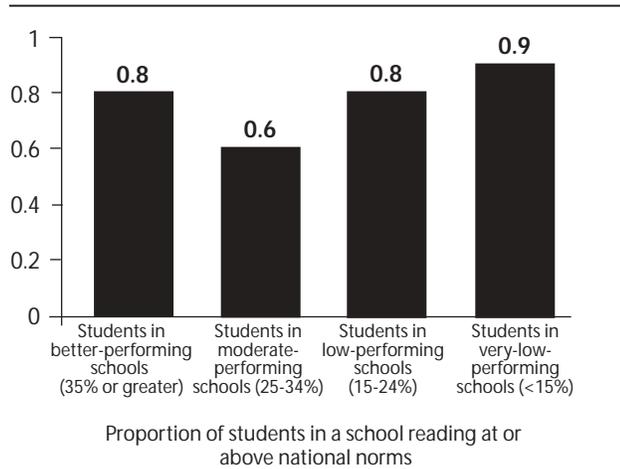
How much do you agree (strongly disagree, disagree, agree, strongly agree) with the following statements about your reading/language arts or mathematics teacher?

- Encourages me to do extra work when I don’t understand something.
- Praises my efforts when I work hard.
- Expects me to do my best all the time.
- Expects me to complete my homework every night.
- Cares if I get bad grades in this class.
- Cares if I don’t do my work in this class.
- Thinks that it is very important that I do well in this class.

In 1994, the measure of academic press developed by the Consortium included additional questions that were later dropped to shorten the survey and improve the psychometric properties of the measure. In order to make the measure of academic press comparable across survey years, difficulties of items common across years were anchored on a single set of values, and the measures were then produced.

Figure 3-10

Eighth Graders in Low-Performing Schools Reported the Largest Increases in Participation in After-School Programs (After 1997)



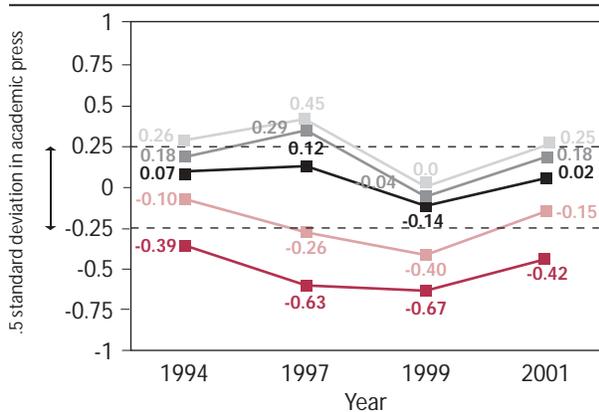
Note: This graph shows the estimated level of students' reports of academic and personal support from teachers when estimating effects for students of different achievement levels and different schools after we have accounted for school characteristics. The model includes additional information on students' demographic characteristics at each time point. In 1994, 26% of moderate risk sixth-grade students attended very-low-performing schools (<15% of students reading at grade level on national norms) at grade level at national norms, 45% attended low-performing schools (15-24% at grade level), 14% attended moderate-performing schools (between 25-34% at grade level) and 15% attended better-performing schools (35% or greater at grade level).

sixth or eighth grades (see Figures 3-11 and 3-12). Among eighth graders, low-achieving (high-risk and moderate-risk) students' reports of academic press declined from 1994 to 1997 and then improved slightly from 1997 to 2001. From 1994 to 2001, there was a consistent decline in the academic expectations reported by higher-performing eighth graders—those students whose test scores were above the promotion cutoff at entrance into eighth grade. In 2001, the measure of academic press among the highest-achieving eighth graders (grade-level students) was over one-fifth of a standard deviation lower than in 1994. If we compare 1994 (prepolicy) to the average of students' reports in 1997-2001 (postpolicy), we would estimate small declines in academic press among all groups, with the largest declines occurring among the highest-performing students.

Across-school differences. The fact that higher-achieving students reported lower levels of academic press in the 1997, 1999, and 2001 surveys could be

Figure 3-11

There Was Little Change in Sixth Graders' Reports of Academic Press



Estimated Change in the Average for Each Core Achievement Group 1994-2001

■ Grade level or above (↓0.01) ■ No risk (0.00) ■ Low risk (↓0.04) ■ Moderate risk (↓0.04) ■ High risk (↓0.03)

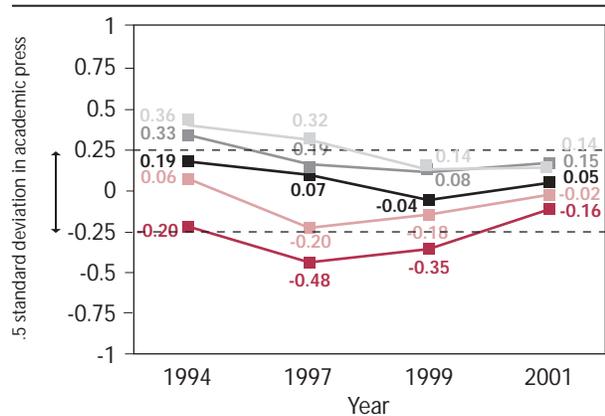
Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall 1994 average. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

interpreted to mean that as low-achieving schools focused on test scores, the higher-achieving students in those schools experienced less academic pressure. This did not, however, appear to be the case. In fact, when we estimated differences in students' reports across schools, we found that eighth graders in better-performing schools (those schools with greater than 25 percent of their students reading at or above grade level on national norms in 1994) reported lower levels of academic press after 1994 (see Figure 3-13).

Why would academic press, or students' reports of the academic expectations of teachers, remain relatively stable over time when these same students had been reporting significantly more academic and personal support from their teachers? One explanation is that academic press measures teachers' expectations for performance while personal support for schoolwork measures teachers' actual behavior in helping students with work, noticing difficulty, and paying attention to students' progress. To restate, items in this measure

Figure 3-12

The Highest Achieving Eighth Graders Reported Lower Levels of Academic Press 1994-2001



Estimated Change in the Average for Each Core Achievement Group 1994-2001

Grade level
 No risk
 Low risk
 Moderate risk
 High risk
 or above (↓-0.22) (↓-0.17) (↓-0.14) (↓-0.08) (↑0.04)

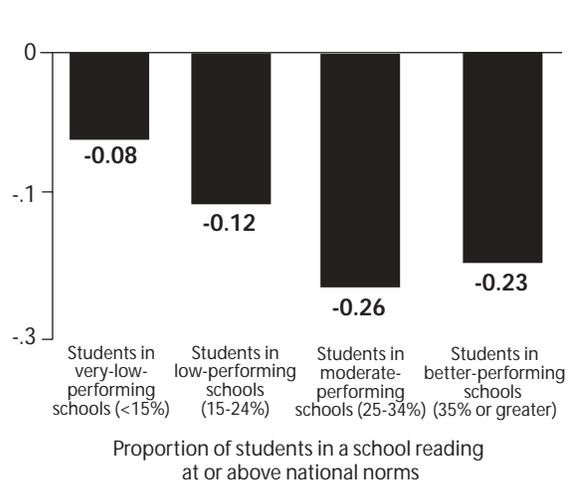
Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall 1994 average. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

include students' answers to questions such as "teachers praise my efforts" and "my teachers expect me to do my best all the time." The ending social promotion policy may have changed teachers' behavior in the classroom toward students with low skills but not their expectation that these students would perform at higher levels. This would explain why low-achieving students showed little change in their perceptions of their teachers' academic expectations, but would not explain declines in students' perceptions of academic press in higher-performing schools.

A second explanation is that while teachers changed the level of personal attention they provided to the lowest-performing students, they did not actually change what they did within their classrooms enough to affect students' perceptions of their teachers' academic expectations. If teachers were spending more time teaching test-related content and basic skills, students might have experienced little change in the academic focus of their classroom environments, and

Figure 3-13

Eighth-Grade Students in Initially Better-Performing Schools Showed Significant Declines in Academic Press (after 1994)



Note: This graph shows the estimated level of students' reports of academic and personal support from teachers when estimating effects for students of different achievement levels and different schools after we have accounted for school characteristics. The model includes additional information on students' demographic characteristics at each time point. In 1994, 26% of moderate-risk sixth-grade students attended very-low-performing schools (<15% of students reading at grade level on national norms), 45% attended low-performing schools (15-24% at grade level), 14% attended moderate-performing schools (between 25-34% at grade level) and 15% attended better-performing schools (35% or greater at grade level).

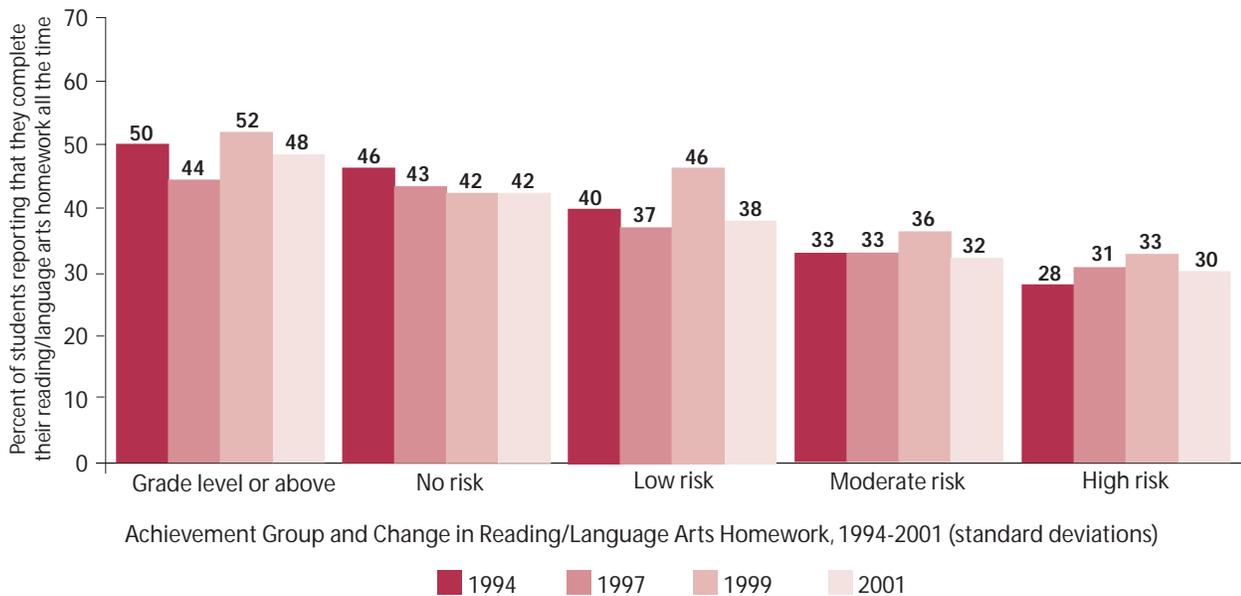
students with the highest skills or in higher-performing schools might actually have experienced declines in their perception of the academic challenge of their coursework.

In order to investigate this question further, we looked at trends in homework completion. In 1994, 1997, 1999, and 2001, eighth graders were asked to report how often (almost never, almost half of the time, most of the time, all of the time) they completed homework in their reading/language arts and mathematics classes. Sixth graders were asked this question for the first time in 1997. Because trends in the sixth grade were quite similar, we have shown results for eighth grade only.

Homework completion in reading/language arts classes. From 1994 to 2001, we observed few changes in eighth graders' reports of completing their homework in reading/language arts classes (see Figure 3-14). The percentage of eighth graders who reported completing their homework "all of the time" rose

Figure 3-14

There Was Little Change in Eighth Graders' Reports of How often They Completed Their Reading/Language Arts Homework (1994-2001)



Note: The scale on this graph shows each group's average measure in standard deviation units from the overall average of 1994. Trends have been adjusted for differences in student characteristics across time.

slightly from 1994 to 1999, and then declined to 1994 levels in 2001. We found similarly flat trends in homework completion in reading/language arts classes among sixth graders from 1997 to 2001.

Homework completion in mathematics classes.

The picture in mathematics was slightly more positive. In the eighth grade, there was a slight upward trend in students' reports of how often they completed their mathematics homework, with students with the lowest skills showing the greatest improvement (see Figure 3-15). In the sixth grade, there was little change in students' reports of mathematics homework completion.

To summarize, students' reports of their teachers' academic expectations and students' reports of their own behavior in meeting those expectations suggest that little progress was made in the post-1994 period toward improving the extent to which students felt that they were encouraged to learn and expected to achieve academically. It is important to remember that our measure of academic press and homework completion did not tap whether teachers had assigned more

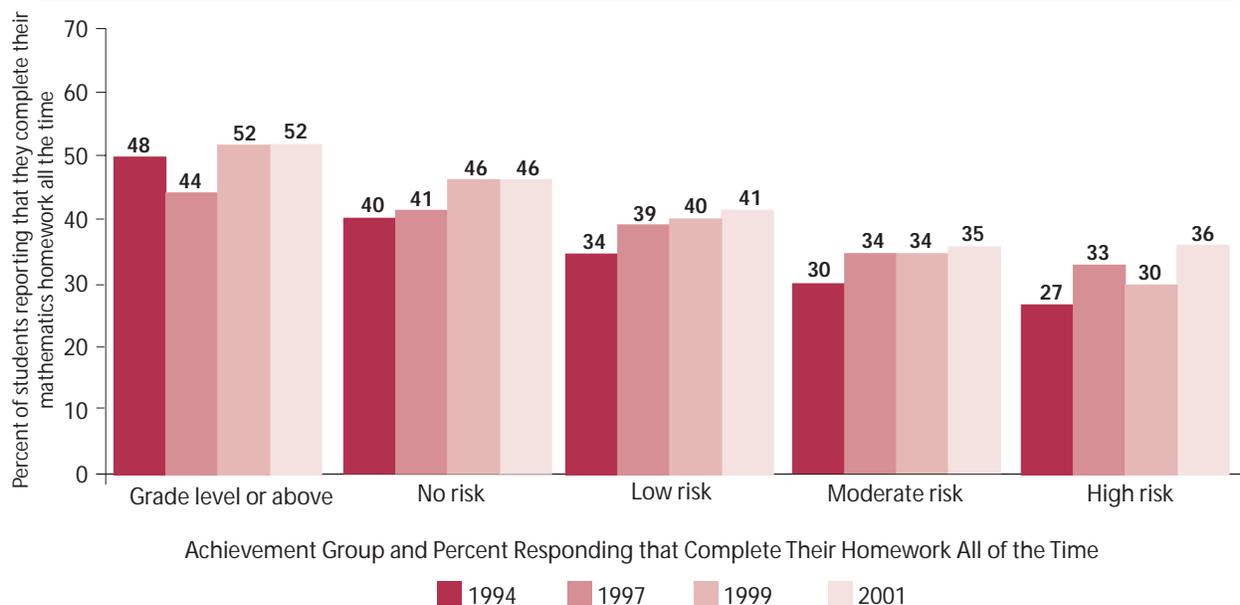
difficult work to students—in other words, whether the content of their courses had changed. Instead, our measure evaluated whether students reported changes in the amount of homework they were assigned, the amount of homework they completed, and the level of their teachers' expectations of work effort.

How Engagement Was Measured: Questions Asked of Eighth-Grade Students in 1994, 1997, 1999, and 2001

How much do you agree (strongly disagree, disagree, agree, strongly agree) with the following statements about your language arts/mathematics class?

- I often count the minutes until class ends.
- Sometimes I get so interested in my work I don't want to stop.
- I usually look forward to this class.
- I am usually bored with what we study in this class.
- The topics we are studying are interesting and challenging.
- I work hard to do my best in this class.

Eighth Graders with the Lowest Skills Reported the Most Improvement in Completing Mathematics Homework All of the Time (1994-2001)



Trends in Students' Levels of Engagement in their Schoolwork: Engagement and Self-Efficacy

While students in Chicago might not have reported higher expectations from teachers after the advent of high-stakes testing, the ending social promotion policy most certainly acted as a form of “academic press” for achievement. A central premise of efforts to end social promotion was that the threat of retention would motivate students to work harder and value achievement. Drawing on motivational theory, critics of such policies argue that the use of such “extrinsic incentives” may actually work to undermine students’ sense of engagement in their schoolwork and feeling of competence or efficacy.⁸ Indeed, a central concern over high-stakes testing is that the threat of retention, when combined with changes in classroom practices such as those we observed in the previous chapter (e.g., increased emphasis on test preparation), will reduce student engagement in school and increase student anxiety, particularly among those students who are most at risk of retention.⁹ In order to assess these claims, we looked at trends in two Consortium survey measures designed

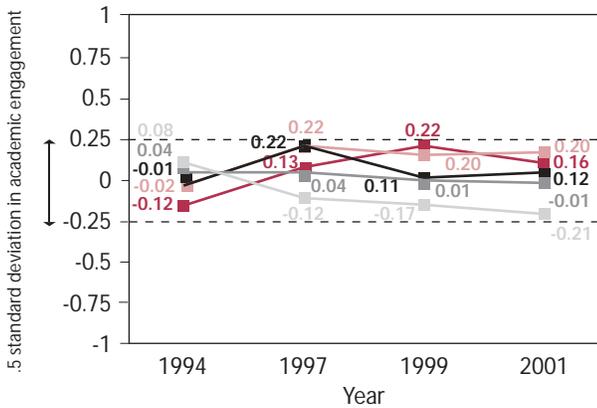
to capture students’ sense of enjoyment in their schoolwork and sense of efficacy and valuation of achievement.

Academic engagement. Since 1994, the Consortium’s student surveys asked eighth graders to respond to a series of questions designed to tap their level of commitment to doing well in school and their level of engagement or connection to their schoolwork. Sixth graders were asked these questions beginning in 1997. Students were asked, for example, to what extent did they agree (strongly agree, agree, disagree, strongly disagree) to questions such as, “I work hard to do my best in class,” “The topics we are studying in this class are interesting and challenging,” and “I usually look forward to class.” Students’ answers to specific questions were then combined into a summary measure of students’ engagement in their schoolwork.

In 1994, Chicago’s eighth graders reported similar levels of engagement in their schoolwork regardless of their academic status. Between 1994 and 2001, however, the academic engagement of high- and moderate-risk students increased by nearly a quarter of a

Figure 3-16

High-Risk Eighth Graders Reported Higher Levels of Academic Engagement While No-Risk Students' Levels of Academic Engagement Declined



Estimated Change in the Average for Each Core Achievement Group 1994-2001

■ Grade level or above (↓-0.29) ■ No risk (↓-0.05) ■ Low risk (↑0.13) ■ Moderate risk (↑0.22) ■ High risk (↑0.28)

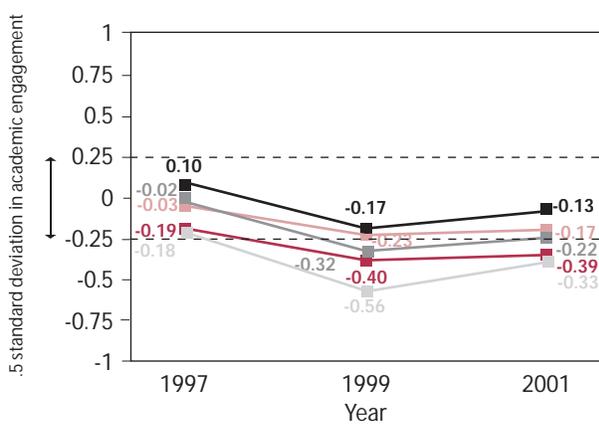
Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall 1994 average. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

standard deviation (see Figure 3-16). At the same time, eighth graders with higher achievement (grade-level students) were much less likely to report working hard and being engaged in their class work in 2001 than in 1994. As a result, by 2001, the lowest-achieving eighth graders in Chicago were more likely to report working hard and finding the work that they did in class enjoyable than students with achievement test scores at or above grade level. We do not observe these trends in the sixth grade. In the sixth grade, from 1997 to 2001, student reports of their academic engagement declined among all groups so that all sixth graders were less likely to report looking forward to and enjoying their class work (see Figure 3-17). In both grades, these changes were statistically significant but substantively small (in the range of 0.2 to 0.3 standard deviations).

Across-school differences. Like academic press, the decline in academic engagement among better-performing eighth graders appeared to be driven by changes in student reports of the challenge and enjoyment of their schoolwork in schools that were initially

Figure 3-17

Sixth Graders' Academic Engagement Declined After 1997



Estimated Change in the Average for Each Core Achievement Group 1997-2001

■ Grade level or above (↓-0.15) ■ No risk (↓-0.20) ■ Low risk (↓-0.23) ■ Moderate risk (↓-0.20) ■ High risk (↓-0.20)

Note: The 1994 data was not available. The scale on this graph shows each groups' average measure in standard deviation units from the overall 1994 average. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

above probation or the threat of probation (schools with 25 percent or more of their students reading at or above national norms in 1994). For example, on

How Self-Efficacy Was Measured: Questions Asked of Students in 1997, 1999, and 2001

How much do you agree (strongly disagree, disagree, agree, strongly agree) with the following statements about your reading/language arts class or mathematics class?

No matter how hard I try, there is some class work I'll never understand.

I am certain I can master the skills taught in this class.

I can do even the hardest work in this class if I try.

If I have enough time, I can do a good job on all my class work.

I can do better work than I'm doing now.

I care if I get a bad grade in this class.

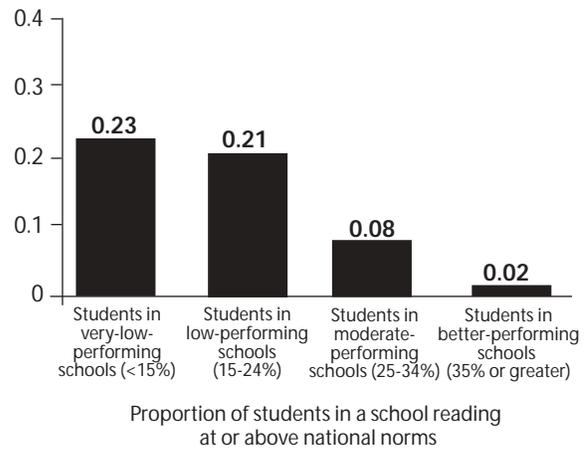
average, eighth graders at moderate risk of retention experienced an increase in their reports of academic engagement from 1994 to 2001, but this increase was largely concentrated among students in the lowest-performing schools. (see Figure 3-18).

Self-efficacy. Research on achievement finds that students are motivated to achieve a goal if they value the outcome either because of some intrinsic interest or extrinsic rewards, and if they believe they can achieve that outcome, which is often called self-efficacy. Beginning in 1997, the Consortium asked students a series of questions designed to tap their perception of their academic competency. Because these questions were not asked in 1994, we cannot compare students' reports prior to the institution of high-stakes testing initiatives. Students reported on the degree to which they agreed (strongly agree, agree, disagree, strongly disagree) with statements such as, "I care if I get a bad grade in this class;" "If I have enough time I can do a good job on all my class work;" and "No matter how hard I try, there is some class work I'll never understand." These items were combined into a single measure that we termed "students' sense of efficacy and valuation of achievement."

In 1997, students' reports of self-efficacy varied by their levels of achievement. For example, the average eighth-grade score on this measure for students with the lowest skills (high-risk students) was nearly two standard deviations lower in 1997 than among students with the highest skills (grade-level students). (See Figure 3-19.) Between 1997 and 2001, eighth-grade students with the lowest skills (high-risk students) reported much higher (0.40 standard deviations) levels of efficacy toward their schoolwork. At the same time, eighth graders with better skills (low-risk, no-risk, and grade-level risk students) felt less confident that they could master the schoolwork than in 1997.

Figure 3-18

Eighth Graders in Low-Performing Schools Reported Higher Levels of Engagement in School After 1997



Note: This graph shows the estimated level of students' reports of academic and personal support from teachers when estimating effects for students of different achievement levels and different schools after we have accounted for school characteristics. The model includes additional information on students' demographic characteristics at each time point. In 1994, 26% of moderate-risk sixth-grade students attended very-low-performing schools (<15% of students reading at grade level on national norms), 45% attended low-performing schools (15-24% at grade level), 14% attended moderate-performing schools (between 25-34% at grade level) and 15% attended better-performing schools (35% or greater at grade level).

Because we do not have a prepolicy (1994) measure of self-efficacy, we could not estimate the prepolicy and postpolicy change across schools.

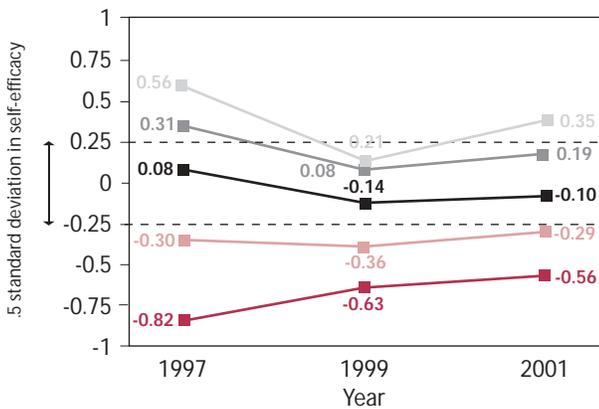
In the sixth grade, we also observed declines in students' sense of efficacy among all achievement levels



John Booz

Figure 3-19

Eighth Graders at High Risk of Retention Felt More Efficacious Toward Their Schoolwork After 1997, While Grade-Level Students Felt Less Efficacious



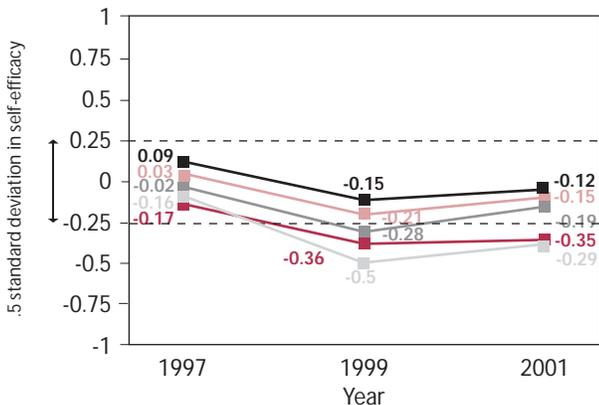
Estimated Change in the Average for Each Core Achievement Group 1997-2001

Grade level (↓-0.22) No risk (↓-0.12) Low risk (↓-0.18) Moderate risk (↑0.01) High risk or above (↑0.27)

Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall 1994 average. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

Figure 3-20

Sixth Graders Reported Feeling Less Efficacious toward Their Classwork in 2001 Than in 1997



Estimated Change in the Average for Each Core Achievement Group 1997-2001

Grade level (↓-0.18) No risk (↓-0.13) Low risk (↓-0.21) Moderate risk (↓-0.18) High risk or above (↓-0.18)

Note: The scale on this graph shows each groups' average measure in standard deviation units from the overall 1994 average. Trends have been adjusted for differences in student characteristics across time. See Appendix C for a description of the statistical model and the method used to place changes into standard deviation units.

(see Figure 3-20). We did not observe improvements among sixth graders with the lowest skills. Thus, with the exception of high-risk eighth graders (approximately 15 percent of CPS eighth graders), there appeared to be declines in students' sense of efficacy toward their schoolwork from 1997 to 2001.

Summary: What We Have Learned

Both advocates and opponents of high-stakes testing policies make claims about the consequences of these policies based on how either adults or students might respond to the incentives created by the threat of retention. Yet, there has been little quantitative research examining whether the institution of high-stakes testing aimed at either students or schools actually impacted students' behavior, experiences in school, and attitudes toward their schoolwork. In this chapter, we used Consortium surveys to examine trends in students' responses to critical indicators of academic support, levels of engagement in school, and perceptions of teachers' expectations for students' academic success and level of work effort. CPS's ending social promotion policy and school accountability programs began in 1996. We considered student responses in 1994 as a baseline measure of students' experiences prior to these policy changes. Students' responses in 1997 gave us an estimate of short-term effects, while survey results in 1999 and 2001 allowed us to identify trends and longer-term effects.

First, there was strong evidence that between 1994 and 2001, low-achieving students' perceptions of the extent to which their teachers gave them personal support for their schoolwork increased dramatically. In 1994, students with low skills reported significantly lower personal support from teachers and their parents than their higher-achieving counterparts. Low-achieving students also reported only occasionally participating in after-school programs for academic help.

This situation changed dramatically over the late 1990's. By 2001, low-achieving students across all schools reported much higher levels of support for their

schoolwork from teachers and parents. The largest increases occurred between 1994 and 1997, in the period when Chicago's high-stakes testing initiative was implemented, and continued to increase in 1999 and 2001. These changes were so dramatic that in many cases the gap in levels of academic support reported by students with low versus average to high skills narrowed considerably. Participation in after-school programs also increased, particularly among eighth-grade students and students who were at moderate to low risk of retention—those students with low but not the lowest skills.¹⁰

We cannot definitively conclude that there is a causal link between trends in these measures and the advent of high-stakes testing. However, the fact that surveyed and interviewed teachers agreed that the policy made them pay attention to students with low skills and increased parental involvement in education, combined with these dramatic changes in students' reports, suggests that the level of attention that low-achieving students received did improve in the post-1996 period. Higher levels of teacher and parental supports did not, however, appear to be accompanied by concomitant improvement in students' perceptions that their teachers held them to higher expectations, or that they felt more challenged and engaged in their course work. In general, these perceptions generally remained flat or slightly declined over this period and varied by both student achievement and grade levels.

Critics of the ending social promotion policy contend, as discussed earlier, that high-stakes testing will lead students with the lowest skills to feel discouraged and become disengaged from school. Indeed, this concern was expressed by Valerie Lee and her colleagues in a 1999 Consortium report on the importance of academic press and personal support for students. The authors used the promotion policy as an example where

high levels of press without personal support could set up students for failure.

Press can be enhanced by the stakes attached to academic success and failure. An excellent illustration of this is represented by the standards and high-stakes student-assessments system, such as Chicago's, that tie grade-level promotion and retention to student performance on standardized tests. The hope is that students who confront these stakes will respond by working harder and learning more. On the other hand, when confronted with higher expectations and high stakes for performance, students who do not perform well may lose motivation, become alienated and disengaged, and eventually drop out of school. Some observers are concerned that such potentially negative outcomes of academic press may be most prevalent in schools that enroll substantial proportions of low-achieving students.¹¹

It is, therefore, important to place the findings of the chapter in the context of prior research. Valerie Lee and her colleagues found that achievement gains occur when schools combine a positive social environment with high expectations for student performance. They concluded, "findings strongly suggest that efforts to improve academic achievement by primarily emphasizing social support in or out of school will not be sufficient unless these efforts are accompanied by strong academic press in school."¹² From this perspective, one interpretation is that low-achieving students received more support to do the same level of work—a change that may have led to greater short-term achievement and engagement, but may not have contributed to long-run improvements in these students' opportunities to learn since these supports were not accompanied by increases in academic expectations or "press."



John Booz

INTERPRETIVE SUMMARY

Since 1996, many major school districts and several states have adopted elements of Chicago Public Schools' (CPS) ending social promotion policy—making student progress toward the next grade dependent upon demonstrated achievement on standardized tests. Many states and districts have made high school graduation contingent upon test performance. And, the No Child Left Behind federal legislation has brought this tough approach to standards and accountability measures to every school system in the nation. Now schools will be evaluated on the basis of their average performance and annual progress as well as on the progress of their most vulnerable students.

The debate over the potential impact of these policies on education is highly polarized. Policy makers argue that such approaches will equalize educational opportunities, set high standards for all students, and improve instruction. Opponents argue that high-stakes testing will only exacerbate existing inequalities in school performance and students' opportunities, leaving students in poor-performing schools left to curriculum focused solely on preparing them for tests. Richard Elmore, for example, has written an eloquent and vigorous critique of the new federal approach, arguing that test-based accountability without investment in professional development, including improving teachers' content knowledge and skills, will only exacerbate existing inequalities.

“The working theory behind test-based accountability is seemingly—perhaps fatally—simple. Students take tests that measure their academic performance in various subject areas. The results trigger certain consequences for students and schools. . . . Having stakes attached to test scores is supposed to create incentives for students and teachers to work harder and for school and district administrators to do a better job of monitoring their performance. . . . The threat of such measures is supposed to be enough to motivate students and schools to even higher levels of performance. . . . Test-based accountability without substantial investment in capacity—internal accountability and instructional improvements in schools—is unlikely to elicit better performance for low-performing students and schools. Furthermore, the increased pressure of test-based accountability without substantial investments in capacity is likely to aggravate the existing inequalities between low-performing and high-performing schools and students.¹

This report takes a rigorous, empirically based, and multifaceted look at educators and students’ views of the impact of ending social promotion in the Chicago Public Schools. We examined teachers and principals’ assessments of the impact of the policy, tracked changes in instructional practice over time, and examined trends in critical indices of students’ experiences in schools. A particular focus was to examine how the effects of the policy were distributed across students and schools. In this section, we discuss what we learned about the impact of the high-stakes accountability initiative. We then highlight the questions our findings raise for assessing and analyzing the impact of high-stakes accountability policies more generally.

. . . CPS teachers and principals viewed the policy of socially promoting students as more negative than retention.

- **Most CPS educators were positive about the impact of ending social promotion.**

Most Chicago teachers and principals assessed the impact of ending social promotion positively. Surveyed and interviewed

teachers reported that the policy focused their instruction and the instructional efforts within their school building, which motivated parents to be more involved in their children’s education and motivated students. Teachers and principals were particularly positive about the addition of after-school and summer-school programs that provided extended and focused instructional time for the lowest-performing students. Even five years after the institution of the policy, most educators remained positive about these instructional impacts and supports.

Many teachers were positive about ending social promotion because its emphasis on basic reading and mathematics skills reflected their own philosophy that basic skills should be the highest educational priority. At a minimum, CPS teachers and principals viewed the policy of socially promoting students as more negative than retention. And, while some teachers felt that more needed to be done to assist retained children, most educators viewed retention, the most controversial component of the 1996 policy change, as having positive educational benefits. Principals were ambivalent, although slightly more negative than teachers, about whether the high-stakes promotion policy had instructional tradeoffs in terms of limiting attention to higher-order skills and diverting attention away from prevention and other initiatives.

- **Time spent on test preparation increased substantially after the institution of high-stakes accountability, particularly in the promotion-gate grades and in low-performing schools.**

Most prior research on the effects of instructional accountability has found that teachers respond by both spending more time teaching students to take tests and aligning their instructional content with the content and skills emphasis of the test. Our analysis of trends in teachers' reports of the time they spent on test preparation and on reading and mathematics content largely supports these findings. Teachers devoted a considerable amount of time to test preparation even before the institution of high-stakes testing. But by 1999, almost half of CPS elementary school teachers reported spending more than 20 hours per year preparing students to take standardized tests. Not surprisingly, the time teachers reported spending on test preparation increased most substantially in

the grades targeted for promotion decision, particularly the third and eighth grades, and in those schools that faced school-level accountability sanctions. We estimated that in the lowest-performing schools, the average time teachers devoted to test preparation doubled between 1994 and 1999. Based both on interviewed teachers' reports and statistical analysis of teachers' survey responses, we estimated that in low-performing schools, the combination of time spent on test preparation and time devoted to taking tests could have amounted to a loss of between three to four weeks of instructional time per year. This created

clear instructional costs in terms of the time available for student learning. Such costs must be considered in light of any potential benefits of the increased accountability.

- **Teachers Shifted Instructional Emphases in Reading and Mathematics.**

We also found evidence that CPS teachers increased the time they spent teaching reading and were more likely to emphasize those mathematics and readings skills tested on the ITBS. In the language arts, teachers devoted more instructional time to the teaching of reading and reading comprehension. In mathematics, we observed in-

creases in students' exposure to grade-level material. In both reading and mathematics, changes in instructional content were greatest in the seventh and eighth grades.

Is this content alignment positive? To the extent that the allocation of reading and mathematics instructional efforts was

undesirable prior to policy implementation, this shift in content emphasis might be viewed positively. A new CPS reading initiative, begun in 2001, and developed by national reading experts, stressed that teachers should spend more time on reading, devoting at least two hours of instructional time per day to teaching reading skills. This new initiative, then, argues that time spent teaching reading is a high priority for this school system. Our analysis suggests that the accountability policy had already encouraged teachers to begin devoting more instructional time to reading.

... in low-performing schools, the combination of time spent on test preparation and time devoted to taking tests could have amounted to a loss of between three to four weeks of instructional time per year.

Previous work done by researchers at the Consortium on Chicago School Research also found that prior to the 1996 reforms, mathematics pacing in schools slowed dramatically in the upper grades, especially in high-poverty schools. Upper-grade students were seldom exposed to grade-level-appropriate mathematics content. Thus, the observed increases in mathematics pacing in the upper grades suggest that students may have benefited from an increased opportunity to learn. We observed the largest improvement in mathematics content coverage in the lowest-performing schools where the earlier research identified the most significant problems. At the same time, our analysis suggests that mathematics pacing had remained a significant concern. Despite improvements, in 2001 we still observed a dramatic decline in the introduction of new topics beginning in the third grade and continuing throughout elementary school. While high-stakes testing may have influenced the content teachers covered in their classrooms, it did not do enough to solve the pacing problem in Chicago schools.

- **Time on test preparation declined between 1999 and 2001, while changes in content emphasis continued to increase.**

Uniquely, in this report, we were able to examine trends in instructional practice both prior to, immediately after, and five years after the initial implementation of high-stakes accountability in 1996. Time spent on test preparation increased substantially from 1994 to 1997, and increased at a lower rate between 1997 and 1999. However, in 2001, the percentage of teachers reporting that they spent more than 20 hours a year on test preparation declined, particularly in the upper grades. Yet teachers continued to make changes in content emphasis through 2001. For example, teachers' reports of time spent on reading comprehension increased substantially between 1997 and 1999, and continued to increase

in 2001. These trends may suggest that test preparation may be a short-term strategy used by teachers when initially faced with increased accountability, but that ultimately, teachers may make more substantive instructional changes as they begin to adjust to high-stakes environments. Teachers may have implemented even more of these types of changes as CPS test scores rose and many schools moved off of probation status.

- **There is substantial evidence that low-achieving sixth- and eighth-grade students experienced greater academic support after the institution of the promotion policy.**

High-stakes accountability in the Chicago Public Schools influenced how teachers allocated their instructional time. One of the most important and surprising findings in this report is that high-stakes accountability also appeared to affect who received attention from teachers, parents, and the district. In 1994, prior to the policy, students with the lowest achievement test scores reported significantly lower levels of academic support from both teachers and parents than their counterparts with higher achievement. Low-achieving students also reported participating only occasionally in after-school programs for academic help. Between 1994 and 2001, we observed dramatic improvement in the extent to which low-achieving sixth- and eighth-grade students felt academically supported by teachers, perceived parental support of and attention to their schoolwork, and participated in after-school programs for help with schoolwork. These changes, in many cases, resulted in a substantial narrowing of the gap in levels of academic supports reported by students with low- versus average- to high-achievement test scores.

We found evidence, moreover, that increases in academic support occurred for low-achieving students across the school system, regardless of the

. . . increases in academic support occurred for low-achieving students across the school system, regardless of the achievement level of the school they attended.

achievement level of the school they attended. Students in the lowest-achieving schools reported higher levels of academic support from teachers and greater involvement in after-school programs between 1994 and 2001. High-stakes accountability concentrated instructional pressure on teachers in low-performing schools, but high stakes for students were experienced by all teachers and parents of low-achieving children regardless of the school they attended. While it might not make sense for a teacher in a better-performing school to make substantial changes in content emphasis and test preparation in response to a few students in her class being at risk of retention, it would make sense for that teacher to pay more attention to those students.

- **There is little evidence that students reported substantially greater academic expectations on the part of their teacher or more engagement in their coursework between 1994 and 2001.**

In this report, our ability to probe deeply into whether the quality of teaching changed over this period was limited. In interviews, teachers rarely indicated that they had responded to high-stakes testing by changing how they taught, by attempting to engage students differently in the classroom, or by investing in professional knowledge of how to teach reading and mathematics. We found little evidence, moreover, that students reported that they experienced higher academic expectations from their teachers, greater academic

demands in the form of homework, or higher levels of engagement in their schoolwork between 1994 and 2001.

The lowest-achieving eighth graders did show small, but statistically significant, increases in academic engagement—the extent to which they reported working hard and enjoying their class work. However, students in initially higher-performing schools reported declines in both academic press (i.e., expectations for performance) and academic engagement over this time period. There is little consistent evidence, moreover, that students in the lowest-performing schools felt that their teachers held them to high expectations, pushed them to learn, or that students themselves were more engaged in school. Trends in student reports cannot tell us whether the quality of instruction changed over this period, but they do suggest that students did not perceive that they were being held to higher expectations. One interpretation of these findings, consistent with Richard Elmore’s argument, is that high-stakes accountability without substantial investments in improving instructional capacity could not have changed how teachers engaged students in learning. Alternatively, CPS’s use of a basic-skills test and a policy that focused accountability on only very-low-performing schools and students might not have been what was required to encourage teachers to raise expectations of students or change how they were engaging students in the classroom.

- **Changes in instructional practice and students’ experiences were substantially greater in the upper grades.**

Our larger evaluation of the ending social promotion initiative found consistently that sixth- and eighth-grade passing rates showed significantly greater improvement than third-grade passing rates.² Roderick, Jacob, and Bryk estimated the achievement value added in promotion-gate

grades (test-score increases over and above that predicted from a student's prior growth trajectory) for groups of students both prior to and after 1996. In both the sixth and eighth grades, achievement-test-score gains increased substantially following the introduction of high-stakes testing. There was little evidence of positive effects in third-grade reading.³

The results of this report lent additional insight and support to the finding that achievement gains were larger in the upper grades. The most significant instructional changes occurred in the seventh and eighth grades, and we find significant evidence for changes in students' experiences in the sixth and eighth grades. In contrast, the most significant instructional trend in the third grade was a substantial increase in the number of teachers preparing students to take standardized tests, so that by 2001, the percentage of teachers devoted to test preparation was highest in the third grade. While eighth-grade teachers reported significant increases in test preparation, they also reported the increases in the extent to which they exposed students to grade-level mathematics material and in time spent on teaching reading comprehension.

What Do These Findings Mean for High-Stakes Accountability Initiatives?

As this report is released, many school systems and states throughout the United States are struggling with how to manage the new national legislation for accountability in No Child Left Behind. Below we highlight some general findings that are relevant to educators and policy makers as they work to implement effective educational reforms in the context of increased accountability. First, results suggest that accountability policies can and do encourage teachers and school administrators to pay greater attention to the lowest-performing students in their classrooms. After the introduction of high-stakes testing in Chicago, low-achieving students perceived significant increases in the academic support they received from

teachers, and these increases were observed among students in both higher-performing and lower-performing schools.

Second, there is evidence to suggest that accountability policies can influence what teachers teach in their classrooms. Teachers in Chicago altered the content they covered in both reading and mathematics in response to the accountability initiative, devoting more time to material covered on the ITBS, a finding that emphasizes the importance of selecting an appropriate test.

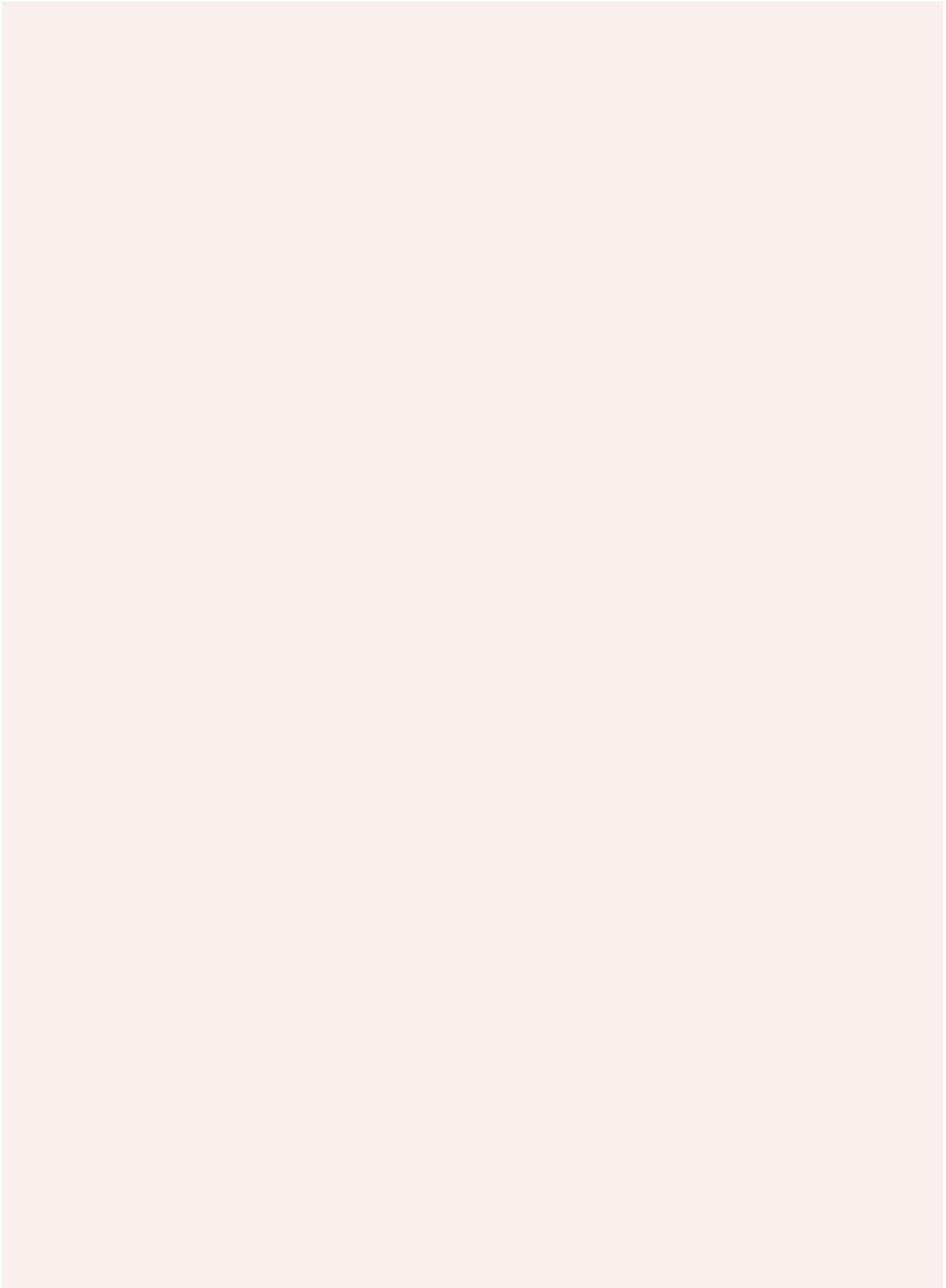
While most Chicago teachers were comfortable with a basic-skills emphasis, the ITBS is not based on publicly stated standards. In fact, state standards (on which there has been an increased emphasis since 1999) require students to apply skills effectively to solve problems, communicate answers, and analyze data and results, and to do so not only in reading and mathematics, but in writing, science, and social studies as well. This suggests that in Chicago, a test that assessed a broader range of skills and content areas may have been a more appropriate choice by which to hold students accountable. As many already have, states and districts will need to continue to invest considerable time and energy to identify and/or select appropriate assessments for measuring the progress of their students and schools.

Third, while accountability policies may help focus teachers' energy and attention on the appropriate content and students, states and districts should also be aware that teachers may need help in directing their energies in productive and academically beneficial ways. For example, in Chicago, although teachers spent more time on reading, it is not clear whether they were also teaching reading more effectively. In interviews, teachers rarely indicated that they had responded to high-stakes testing by changing how they taught, by attempting to engage students differently in the classroom, or by investing in professional knowledge of to whom reading and mathematics should be taught. Similarly, although teachers gave more support to low-achieving students, this support did not necessarily translate into increases in academic expect-

tations for students or greater academic engagement. Accountability policies that are accompanied by significant investments in building teachers' capacity and skills will likely meet with greater success.

Finally, any district and state involved in using achievement tests to evaluate school performance must understand that teachers and school communities will respond to the pressure to raise test scores by increasing time on direct test-preparation activities that may not necessarily lead to real increases in students' learning and future success. This creates clear instructional costs because it takes valuable time away from students' learning. In the worst case scenario, test preparation can allow schools to buffer themselves from the need

to invest in long-term changes that may alter instruction in positive ways and lead to real benefits for students. School districts must place a high priority on finding ways to minimize the costs associated with time spent on test preparation and maximize the potential benefits of increased accountability by encouraging teachers to look carefully at their instructional practices. States and districts must build the instructional capacity of schools and provide the supports necessary for teachers to make long-term investments in instructional improvements, the benefits of which will continue after the initial pressure of accountability policies wane.



APPENDICES

APPENDIX A

Hierarchical Linear Models

The model of teachers' assessments of the policy was estimated using responses from all teachers in the system who completed the survey in 1999. The analysis used a three-level Hierarchical Linear Model (HLM). Level 1 accounts for the measurement error in the Rasch measure. Level 2 explores the influence of teacher background characteristics on teachers' assessments of the policy. The list of teacher background characteristics explored includes indicators of teacher race, education, and gender; years of teaching experience; grade level taught; and indicators of whether the teacher was a self-contained classroom teacher or a subject-specific teacher. Indicators for missing values were also included, and missing values were set at zero.

The outcomes in these analyses are measures produced by Rasch analysis from survey items. Each Rasch measure is a random variable, comprised of a latent, true measure and a random measurement error term. This can be expressed as an equation:

$$Y_{ij} = \pi_{ij} + e_{ij}$$

where $Y_{ij} \sim N(\bar{Y}, \sigma_{ij}^2)$

and $e_{ij} \sim N(0, \sigma_{ij}^2)$

Y_{ij} is the observed outcome for student i in school j

π_{ij} is the latent measure for student i in school j

e_{ij} is the measurement error associated with Y_{ij}

Ordinarily, the errors at Level 1 have a constant variance, but in this case, each of the Rasch measures can have a different amount of measurement error. In order to correct for this heteroscedasticity, we multiply both sides of the equation by the inverse standard error ($\frac{1}{se_{ij}}$), where se_{ij} is the standard error of measurement for student i in school j on the given measure. The Level 1 equation thus becomes:

Level 1:

$$Y_{ij}^* = \pi_{ij}^* \left(\frac{1}{se_{ij}} \right) + e_{ij}^*$$

where Y_{ij}^* is the observation divided by its standard error

π_{ij}^* is the latent measure adjusted for measurement error

$e_{ij}^* \sim N(0,1)$

Note that the Level 1 error is now distributed standard normal. The π_{ij}^* is the latent measure adjusted for measurement error and comes down to become our outcome at Level 2.

Level 2:

$$\pi_{ij} = \beta_{0j} + \beta_{1j}GATE_{ij} + \beta_{2j}SUBJECT_{ij} + \beta_{3j}OSUBJECT_{ij} + \beta_{4j}LOWEXP_{ij} + \beta_{5j}MODEXP_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISPANIC_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}MALE_{ij} + \beta_{10j}BACHELOR_{ij} + r_{ij}$$

π_{ij} is either the measure of teacher assessment of the policy or the measure of teacher concerns about retention adjusted for measurement error. GATE indicates that the teacher teaches in grades three, six, or eight (the promotion-gate grades). SUBJECT indicates that the teacher is a departmentalized classroom instructor, such as a math teacher, rather than a self-contained classroom teacher. OSUBJECT indicates that the teacher teaches auxiliary classes such as art or physical education rather than being a self-contained classroom teacher. Self-contained classrooms are the omitted category. LOWEXP indicates the teacher has fewer than five years of teaching experience, while MODEXP indicates the teacher has between five and fifteen years of experience. Teachers with more than fifteen years of teaching experience are the omitted category. BLACK indicates that the teacher is black, HISPANIC indicates that the teacher is Latino, and OTHRACE indicates that the teacher is another race (e.g., Asian, Native American, or biracial/multiethnic). White teachers are the excluded category. MALE indicates the teacher is male rather than female, and BACHELOR indicates that the teacher has received no more than a bachelor's degree.¹

The individual teacher characteristics are all grand-mean centered, so that the intercept β_{0j} can be interpreted as the mean level of the support variable (Y_{ij}) in a school with average teacher characteristics in 1999.

Level 3:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}EXCLR99_j + \gamma_{02}HISP_j + \gamma_{03}INTG_j + \gamma_{04}MIXED_j + \gamma_{05}MNTY_j + \gamma_{06}MODRET_j + \gamma_{07}HIRET_j + \gamma_{08}VHIRET_j + v_{0j}$$

Again, β_{0j} can be interpreted as the mean level of support in school j . EXCLR99 indicates the percentage of students in a school excluded from the policy for reasons such as special-education status. HISP is a dummy variable indicating that the school is predominantly Latino (more than 85 percent), and INTEG is a dummy variable indicating that the school is more than 30 percent white. MIXED is a dummy variable indicating that the school is between 15 and 30 percent white, and MNTY is a dummy variable indicating that the school is less than 15 percent white, but not more than 85 percent African-American or 85 percent Latino. Predominantly African-American schools are the omitted category (over 85 percent African-American). MODRET indicates that the school had a moderate retention rate in 1998 (6-16 percent), HIRET indicates a high retention rate (16-33 percent), and VHIRET indicates a very high retention rate (more than 33 percent). The excluded category was schools with low retention rates (less than 6 percent). The intercept represents the mean value of the coefficient districtwide and the error terms represent school-specific effects.

¹ Although not shown in the model, indicators for missing values were also included.

Appendix B

Hierarchical Linear Models

Models were estimated using survey responses from all teachers in the system. We combined survey responses of teachers in each of the years into a single data set and explored aggregate changes over time. We controlled for teacher background characteristics at Level 2. The vector of teacher background characteristics includes indicators of teacher race, education, gender, years of teaching experience, grade level taught, and indicators of whether the teacher was a self-contained classroom teacher or a subject-specific teacher. Indicators for missing values were also included, and missing values were set to zero.¹

The outcomes in these analyses are measures produced by Rasch analysis from survey items. Each Rasch measure is a random variable, comprised of a latent, true measure and a random measurement error term. This can be expressed as an equation:

$$Y_{ij} = \pi_{ij} + e_{ij}$$

$$\text{where } Y_{ij} \sim N(\bar{Y}, \sigma_{ij}^2)$$

$$\text{and } e_{ij} \sim N(0, \sigma_{ij}^2)$$

Y_{ij} is the observed outcome for student i in school j

π_{ij} is the latent measure for student i in school j

e_{ij} is the measurement error associated with Y_{ij}

Ordinarily, the errors at Level 1 have a constant variance, but in this case, each person's score on each of the Rasch measures can have a different amount of measurement error. In order to correct for this heteroscedasticity, we multiply both sides of the equation by the inverse standard error ($\frac{1}{se_{ij}}$), where se_{ij} is the standard error of measurement for student i in school j on the given measure. The Level 1 equation thus becomes:

Level 1:

$$Y_{ij}^* = \pi_{ij}^* \left(\frac{1}{se_{ij}} \right) + e_{ij}^*$$

where Y_{ij}^* is the observation divided by its standard error

π_{ij}^* is the latent measure adjusted for measurement error

$$e_{ij}^* \sim N(0,1)$$

Note that the Level 1 error is now distributed standard normal. The π_{ij}^* is the latent measure adjusted for measurement error and comes down to become our outcome at Level 2.

Level 2:

$$\pi_{ij}^* = \beta_{0j} + \beta_{1j}(1997)_{ij} + \beta_{2j}(1999)_{ij} + \beta_{3j}(2001)_{ij} + \sum_q B_{qj} Z_{qij} + r_{ij}$$

π_{ij}^* is the outcome of interest adjusted for measurement error (e.g., pacing measure, test-preparation item, etc.) and Z_{qij} is a vector of background characteristics of the teacher including the grade and subject taught. The individual teacher characteristics are all grand-mean centered, so that the intercept β_{0j} can be interpreted as the mean level of the outcome (π_{ij}^*) in a school with average teacher characteristics in 1994. The binary variables 1997_{ij} , 1999_{ij} and 2001_{ij} indicate the year of the survey response. Thus, β_{1j} can be interpreted as the change in the outcome (π_{ij}^*) between 1994 and 1997, β_{2j} as the change in the outcome (π_{ij}^*) between 1994 and 1999, and β_{3j} as the change in the outcome (π_{ij}^*) between 1994 and 2001. These binary variables are not centered. When no 1997 measures are available, only binary variables for 1999 and 2001 are included, and the intercept can then be interpreted as the mean level of the outcome in 1997.²

Level 3:

$$\beta_{0j} = \gamma_{01} + \mathbf{v}_{0j}$$

$$\beta_{1j} = \gamma_{11} + \mathbf{v}_{1j}$$

$$\beta_{2j} = \gamma_{21} + \mathbf{v}_{2j}$$

$$\beta_{3j} = \gamma_{31} + \mathbf{v}_{3j}$$

$$\beta_{qj} = \gamma_{q1}$$

β_{0j} can be interpreted as the mean level of outcome Y in school j in year 1994, β_{1j} can be interpreted as the average change in outcome Y in school j between 1994-1997, β_{2j} as the average change in outcome Y in school j between 1994-1999, and β_{3j} as the average change in outcome Y in school j between 1994-2001. The intercepts ($\gamma_{01}, \gamma_{11}, \gamma_{21}, \dots, \gamma_{q1}$) represent the mean value of the coefficient districtwide, and the error terms represent school-specific effects.

When modeling schools' level responses, Level 3 becomes:

Level 3:

$$\beta_{0j} = \gamma_{01} + \sum_s \gamma_{0s} X_{sj} + \theta_{0j}$$

$$\beta_{1j} = \gamma_{11} + \theta_{1j}$$

$$\beta_{2j} = \gamma_{21} + \theta_{2j}$$

$$\beta_{3j} = \gamma_{31} + \sum_s \gamma_{3s} X_{sj} + \theta_{3j}$$

$$\beta_{qj} = \gamma_{q1}$$

where X is a vector of school-level characteristics.

Model for dichotomous outcomes and composite measures:

For single-item indicators, such as "test preparation," or composite measures, such as "reading comprehension," where the outcome is simply the sum of items teachers report, there is no standard error available. For these indicators, the outcome variable is simply the dichotomous outcome or the composite measure, and Level 2 becomes the Level 1 equation. For dichotomous outcomes, Level 2 is a logistic regression.

Appendix C

Hierarchical Linear Models Used for Identifying Trends in Students' Attitudes

In Chapter 3, we reported results from two sets of analyses used to estimate trends in student survey responses. The first analytic set estimated trends in each survey year (1994, 1997, 1999, and 2001) for students with similar “achievement levels” as measured by their distance from the 1997 test-score cutoff for promotion (or “risk” under the policy, see Box 3-2). In the first set of models, we controlled for the students’ characteristics. In the second set of models, we estimated changes in students’ attitudes in the prepolicy versus postpolicy period (1994 versus the average of 1997, 1999, and 2001), controlling for student and school characteristics. Students were included in our analyses if they completed surveys and if they attended a regular elementary school that participated in surveys both in 1994 and at least one additional survey year.

For each set of analyses, the model we used depended on whether the outcome variable was a Rasch measure or a single-item indicator, such as the time students reported spending on homework. Measures created using Rasch analysis (personalism, parent support, academic press, engagement, and efficacy) have known standard errors. We take advantage of this additional information to adjust for measurement error by using three-level hierarchical linear models where Level 1 is a measurement model.

Measurement Model

Each Rasch measure is a random variable, comprised of a latent, true measure and random measurement error. This can be expressed as an equation:

$$Y_{ij} = \pi_{ij} + e_{ij}$$

$$\text{where } Y_{ij} \sim N(\bar{Y}, \sigma_{ij}^2)$$

$$\text{and } e_{ij} \sim N(0, \sigma_{ij}^2)$$

Y_{ij} is the observed outcome for student i in school j

π_{ij} is the latent measure for student i in school j

e_{ij} is the measurement error associated with Y_{ij}

Ordinarily, the errors at Level 1 have a constant variance, but in this case, each of the Rasch measures can have a different amount of measurement error. In order to correct for this heteroscedasticity, we multiply both sides of the equation by the inverse standard error ($\frac{1}{se_{ij}}$), where se_{ij} is the standard error of measurement for student i in school j on the given measure. The Level 1 equation thus becomes:

Level 1:

$$Y_{ij}^* = \pi_{ij}^* \left(\frac{1}{se_{ij}} \right) + e_{ij}^*$$

where Y_{ij}^* is the observation divided by its standard error

π_{ij}^* is the latent measure adjusted for measurement error

$$e_{ij}^* \sim N(0,1)$$

Note that the Level 1 error is now distributed standard normal. The π_{ij}^* is the latent measure adjusted for measurement error and comes down to become our outcome at Level 2.

Level 2:

At Level 2, we estimate how students' measures vary by their demographic characteristics, level of risk (i.e., distance from the promotion cutoff under the policy), and the year in which they participated in the survey.

$$\begin{aligned} \pi_{ij}^* = & \beta_{0j} + \beta_{1j} \text{AGE}_{ij} + \beta_{2j} \text{HISPANIC}_{ij} + \beta_{3j} \text{WHITE/OTHER}_{ij} + \beta_{4j} \text{MALE}_{ij} + \beta_{5j} \text{BILINGUAL}_{ij} + \\ & \beta_{6j} \text{EXCLUDE}_{ij} + \beta_{7j} \text{RETAINED}_{ij} + \beta_{8j} \text{HIGHRISK}_{ij} + \beta_{9j} \text{MODERATERISK}_{ij} + \beta_{10j} \text{LOWRISK}_{ij} + \beta_{11j} \text{GRADELEVEL}_{ij} \\ & + \beta_{12j} \text{MISSRISK}_{ij} + \beta_{13j} 1997_{ij} + \beta_{14j} 1999_{ij} + \beta_{15j} 2001_{ij} + \beta_{16j} (\text{HIGHRISK} * 1997)_{ij} \\ & + \beta_{17j} (\text{HIGHRISK} * 1999)_{ij} + \beta_{18j} (\text{HIGHRISK} * 2001)_{ij} + \beta_{19j} (\text{MODERATERISK} * 1997)_{ij} \\ & + \beta_{20j} (\text{MODERATERISK} * 1999)_{ij} + \beta_{21j} (\text{MODERATERISK} * 2001)_{ij} + \beta_{22j} (\text{LOWRISK} * 1997)_{ij} \\ & + \beta_{23j} (\text{LOWRISK} * 1999)_{ij} + \beta_{24j} (\text{LOWRISK} * 2001)_{ij} + \beta_{24j} (\text{GRADELEVEL} * 1997)_{ij} \\ & + \beta_{25j} (\text{GRADELEVEL} * 1999)_{ij} + \beta_{26j} (\text{GRADELEVEL} * 2001)_{ij} + r_{ij} \end{aligned}$$

AGE indicates the students' age as of the current survey year. HISPANIC, WHITE AND OTHER, and MALE are dummy variables for students' race and gender with African-American and female students as the comparison groups. BILINGUAL, EXCLUDE, and RETAINED are dummy variables indicating whether the student is enrolled in the bilingual program, is excluded under the policy (i.e., the students' test scores are excluded from reporting because of special circumstance such as special-education status), or whether the student is currently retained under the policy, respectively. We include three dummy variables for the year of survey administration (1997, 1999, 2001), with 1994 as the reference category. HIGHRISK, MODERATERISK, LOWRISK, GRADELEVEL, MISSING RISK (i.e., student is missing test scores) are dummy variables indicating each student's risk under the policy. Students at no risk of retention are the left-out category. Finally, we include 12 interaction terms (e.g., HIGHRISK*1997) between survey year and student-risk category. Thus, in this model, β_{13j} indicates the change in the average of student reports at each school from 1994 to 1997 among students who are not at risk of retention, and $\beta_{13j} + \beta_{16j}$ indicates the change for a student at high risk of retention. With the exception of the intercept β_{0j} and terms denoting years or interactions, all student characteristics are fixed and grand-mean centered.

Level 3:

$$\beta_{0j} = \gamma_{01} + \nu_{0j}$$

At this level, we include a random school effect.

Modeling Single-Item Indicators

For single-item indicators (e.g., mathematics and reading/language arts homework completion, after-school programming) without known standard errors, we used a two-level hierarchical model. In the models, the Level 2 and Level 3 equations simply become Level 1 and Level 2 equations.

Level 1:

$$\begin{aligned}
 Y_{ij} = & \beta_{0j} + \beta_{1j} \text{AGE}_{ij} + \beta_{2j} \text{HISPANIC}_{ij} + \beta_{3j} \text{WHITE/OTHER}_{ij} + \beta_{4j} \text{MALE}_{ij} + \beta_{5j} \text{BILINGUAL}_{ij} + \\
 & \beta_{6j} \text{EXCLUDE}_{ij} + \beta_{7j} \text{RETAINED}_{ij} + \beta_{8j} \text{HIGHRISK}_{ij} + \beta_{9j} \text{MODERATERISK}_{ij} + \beta_{10j} \text{LOWRISK}_{ij} + \beta_{11j} \text{GRADELEVEL}_{ij} \\
 & + \beta_{12j} \text{MISSRISK}_{ij} + \beta_{13j} 1997_{ij} + \beta_{14j} 1999_{ij} + \beta_{15j} 2001_{ij} + \beta_{16j} (\text{HIGHRISK} * 1997)_{ij} \\
 & + \beta_{17j} (\text{HIGHRISK} * 1999)_{ij} + \beta_{18j} (\text{HIGHRISK} * 2001)_{ij} + \beta_{19j} (\text{MODERATERISK} * 1997)_{ij} \\
 & + \beta_{20j} (\text{MODERATERISK} * 1999)_{ij} + \beta_{21j} (\text{MODERATERISK} * 2001)_{ij} + \beta_{22j} (\text{LOWRISK} * 1997)_{ij} \\
 & + \beta_{23j} (\text{LOWRISK} * 1999)_{ij} + \beta_{24j} (\text{LOWRISK} * 2001)_{ij} + \beta_{24j} (\text{GRADELEVEL} * 1997)_{ij} \\
 & + \beta_{25j} (\text{GRADELEVEL} * 1999)_{ij} + \beta_{26j} (\text{GRADELEVEL} * 2001)_{ij} + r_{ij}
 \end{aligned}$$

Level 2:

$$\beta_{0j} = \gamma_{01} + \nu_{0j}$$

Standardization

We reported trends in measures and single-item indicators in terms of standardized change from 1994 to help maintain consistency in reporting across measures as well as to give readers a sense of the magnitude (an effect size) of change over time. For every measure and indicator, we generated standard deviations from “unconditional” hierarchical linear models. These models have the same general form as those presented above, except these include no predictors. To calculate, we subtracted the predicted mean on each variable of a given year and risk category from the overall mean of 1994 and divided by the relevant, individual-level standard deviation generated from the unconditional hierarchical linear models. In general, effect sizes between .20 and .50 are considered small, those between .50 and .80 are considered moderate, and those above .80 are considered large. An effect size of .50 indicates that an observer would be more likely than chance (60 percent) to guess the year and risk category of a student, based on his or her score.

Appendix D

Hierarchical Linear Models Used for Estimating Postpolicy Trends in Student Attitudes

In our second set of analyses, we aimed to disentangle by how much student responses impacted the trends we observed, regardless of the schools the students attended. Also, we aimed to determine how much the racial and achievement characteristics of the school affected the trends in student responses. For example, we could observe that students with very low skills reported more personal support for their schoolwork either because low-achieving students in all Chicago Public Schools received more attention or because very-low-performing schools in Chicago were paying more attention to all of their students, regardless of their achievement levels.

The models we used to control for such school characteristics were quite similar to those we used to estimate trends of students over time (see Appendix C). The Level 1 model remains the same. However, instead of including dummy variables for each survey year, we grouped students in terms of whether they responded to the survey in the years after the inception of the policy (1997, 1999, and 2001). The Level 2 model becomes:

Level 2:

$$\begin{aligned} \pi_{ij} = & \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}HISPANIC_{ij} + \beta_{3j}WHITE/OTHER_{ij} + \beta_{4j}MALE_{ij} + \beta_{5j}BILINGUAL_{ij} + \\ & \beta_{6j}EXCLUDE_{ij} + \beta_{7j}RETAINED_{ij} + \beta_{8j}HIGHRISK_{ij} + \beta_{9j}LOWRISK_{ij} + \\ & \beta_{10j}GRADELEVEL_{ij} + \beta_{11j}MISSRISK_{ij} + \beta_{12j}POSTPOLICY_{ij} + \beta_{13j}(HIGHRISK * POSTPOLICY)_{ij} \\ & + \beta_{14j}(LOWRISK * POSTPOLICY)_{ij} + \beta_{15j}(NORISK * POSTPOLICY)_{ij} + \\ & \beta_{16j}(GRADELEVEL * POSTPOLICY)_{ij} + r_{ij} \end{aligned}$$

AGE indicates students' ages as of the current survey year. HISPANIC, WHITE AND OTHER, and MALE are dummy variables for students' race and gender, with African-American and female students being the left-out comparison groups. BILINGUAL, EXCLUDE, and RETAINED are dummy variables that indicate, respectively, whether the student was enrolled in the bilingual program, was excluded under the policy (i.e., the student's test scores were excluded from reporting due to special circumstances, such as special-education status), or was retained at that time under the policy. POSTPOLICY indicates that the student was surveyed in 1997, 1999, or 2001, with 1994 as the reference category. HIGHRISK, LOWRISK, NORISK, GRADELEVEL, and MISSING RISK are dummy variables indicating each student's risk under the policy. Students at moderate risk of retention are the reference group. Finally, we included four interaction terms (e.g., HIGHRISK*POSTPOLICY). With the exception of β_{0j} and terms denoting years of interactions, student characteristics are fixed and grand-mean centered.

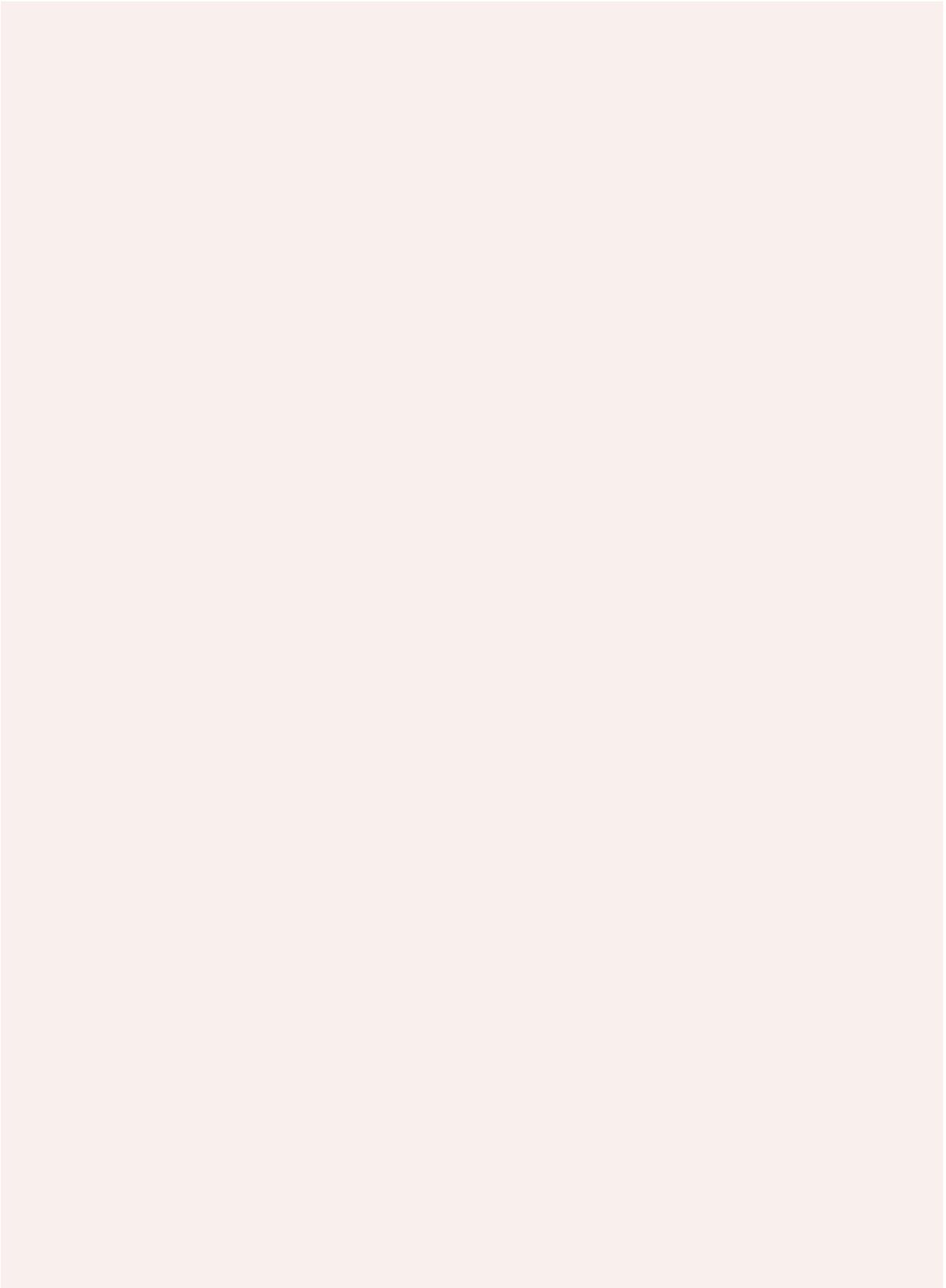
Level 3:

$$\begin{aligned} \beta_{0j} = & \gamma_{01} + \gamma_{02}EXCL99_j + \gamma_{03}HISP_j + \gamma_{04}INTG_j + \gamma_{05}MIXED_j + \gamma_{06}MNTY_j + \gamma_{07}0-15\%_j + \\ & \gamma_{08}24-35\%_j + \gamma_{09}>35\%_j + v_{0j} \end{aligned}$$

EXCLR99 indicates the percentage of students in a school excluded from the policy for reasons such as special-education status. HISP is a dummy variable indicating that the school was predominantly Latino (more than 85 percent), INTEG is a dummy variable indicating that the school is more than 30 percent white. MIXED is a dummy variable indicating that the school was between 15 and 30 percent white, and MNTY is a dummy variable indicating that the school was less than 15 percent white, but not more than 85 percent African-American or 85 percent Latino. Predominantly African-American schools were the omitted category (over 85 percent African-American). We used three dummy variables to control for whether in 1994 a school's student body was 0 to 15 percent, 25 to 35 percent, or over 35 percent at or above national norms based on the Iowa Tests of Basic Skills. The comparison category is 16 to 24 percent. In short, from these models, we can estimate the overall trends of students of different risk levels, and we can also determine to what extent 1994 school performance and racial composition impacted these trends over time.

Modeling Single-Item Indicators

For single-item indicators (e.g., math and reading/language arts homework completion, after-school programming) without known standard errors, we used a two-level hierarchical model. In the models, the Level 2 and Level 3 equations simply became Level 1 and Level 2 equations.



ENDNOTES

Introduction

¹ Easton, Rosenkranz, and Bryk (2001); Hess (1999).

² Jacob (2001), Dissertation; Haladyna, Nolen, and Haas (1991); Koretz (1988); Madaus (1991); Shepard (1990, 1991); Urdan and Paris (1994).

³ Hochschild and Scott (1998); Darling-Hammond and Wise (1985); Jones, Jones, and Hardin (1999); Rosenholtz (1987); Smith (1991); Urdan and Paris (1994).

⁴ Darling-Hammond and Wise (1985); Firestone, Mayrowetz, and Fairman (1998); Jones, Jones, and Hardin (1999); Rosenholtz (1987); Smith (1991); Urdan and Paris (1994).

⁵ For example, Darling-Hammond and Wise (1985); Rosenholtz (1987); Smith (1991).

⁶ Roderick et al. (1999); Roderick, Jacob, and Bryk (2003).

⁷ Betts and Costrell (2001); Wheelock et al. (2000).

⁸ Catterall (1989); Jacob (2001); Roderick and Engel (2000).

⁹ Koretz (1999); Linn (1994).

¹⁰ From the beginning, the policy did not apply strictly to all students in these grades. Students are not included under this policy if they are in special education or have been in bilingual education for fewer than four years. A student who is in special education follows the student's Individual Education Plan that includes a promotion plan and may require the student to attend Summer Bridge or be retained. The decision not to include students who were in bilingual or special education meant that many students were not included under the policy. In 1997, only 70 percent of third graders and over 80 percent of sixth and eighth graders were included. Inclusion rates have decreased over time. See Roderick et al. (1999); Easton (2000).

¹¹ Roderick, et al. (1999); Roderick et al. (2000).

¹² Roderick et al. (1999).

¹³ The standard for eighth graders was initially set at 7.0. In 1998 this standard was raised to 7.2 and in 1999 to 7.4.

¹⁴ House (1998); Roderick et al. (1999).

¹⁵ The standard has been raised subsequently to 20 percent of students at or above national norms. In addition, the criterion for getting off probation had been 20 percent and now is 25 percent.

¹⁶ Hess (1999).

¹⁷ Hess (1999).

¹⁸ Bryk et al. (1998), "Academic Productivity of CPS;" Bryk et al. (1998) *Charting Chicago School Reform*.

¹⁹ Lee et al. (1999); Smith (1998); Smith, Smith, and Bryk (1998); Smith, Lee, and Newmann (2001).

Chapter 1

¹ Jones, Jones, and Hardin (1999); Rosenholtz (1987); Smith (1991); Urdan and Paris (1994).

² The question asked teachers, “When you think about students who are at risk of being retained or have already been retained, please indicate the extent to which you agree or disagree with the following statements.” In addition to the question about referring students to tutoring or after school programs, teachers were asked about such things as spending additional time outside of school with students or slowing the pace of class to accommodate these students. While teachers agreed with most of these statements, at least to some degree, more teachers agreed with the statement about referring students to after-school programs than with any other statement.

³ Holmes (1989); Holmes and Matthew (1984); Reynolds (1992).

⁴ Alexander, Entwisle, and Dauber (1994); Peterson, DeGracie, and Ayabe (1987).

⁵ Alexander, Entwisle, and Dauber (1994); Holmes (1989); Holmes and Matthews (1984); Peterson, DeGracie, and Ayabe (1987); Pierson and Connell (1992); Reynolds (1992).

⁶ Alexander, Entwisle, and Dauber (1994); Holmes (1989); Holmes and Matthews (1984); Pierson and Connell (1992); Reynolds (1992).

⁷ Grissom and Shepard (1989); House (1998); Roderick (1994).

⁸ See, for example, Tomchin and Impara (1992); Smith and Shepard (1988); Shepard and Smith (1989).

⁹ These same questions were not asked of teachers.

¹⁰ These summaries were created using Rasch analysis.

¹¹ The items used to construct this scale are listed in Table 1-1.

¹² These data are an unpublished tabulation of survey demographic data.

¹³ Estimates from this model are not shown here. It is identical to the model outlined in Appendix A, except that rather than including three dummy variables for school retention rate, this model includes dummy variables for school performance in 1994. The results are the same if we also control for the retention rate of the school by including a continuous retention-rate variable at Level 2.

¹⁴ Over one-third of schools (36 percent) had retention rates that could be considered moderate (between 6 and 16 percent of students retained) and another one-third of schools retained between 16 and 33 percent of their class (high retention rates). Finally, a small percentage of schools in 1998 (4 percent) retained more than one-third of third, sixth, and eighth graders that year.

Cutoffs for retention rates were based on an analysis of the distribution of retention rates across schools and consideration of how many students, on average, were retained in a class. If the average class size is 28 (the pupil-teacher ratio in CPS for a non-special education classroom in the elementary grades), a

retention rate of less than six percent translated into fewer than two students in each of the third, sixth, and eighth grade classes. A retention rate of six to 16 percent translates into two to five students per class, and a retention rate of 16 to 33 percent translates into a retention rate of between five and 10 students per class.

¹⁵ Roderick et al. (2000).

Chapter 2

¹ Our categorization of test preparation draws on the work of Perry (2000).

² Similar trends were observed across all response categories.

³ See Appendix B for a description of the HLM used to obtain these estimates. We characterized schools on the basis of their achievement test scores in 1994. Schools were characterized as *very low performing* if, in 1994, fewer than 15 percent of their student body had ITBS reading scores at or above national norms. This was the standard that was used to place schools on academic probation. In 1994, 24 percent of the city’s 456 elementary schools were very low performing. Schools were identified as *low performing* (36 percent of schools) if between 15 to 24 percent of their students’ scores were at or above national norms in reading. The designation *moderate performing* (16 percent of schools) was used for schools with 25 to 35 percent of their student body scoring at or above national norms on the ITBS reading test. Schools were considered to be *better performing* (24 percent of schools) if more than 35 percent of their students were performing at or above national norms.

⁴ As can be seen from the examples given, teachers often did not give a precise estimate of the time they spent on test preparation but provided answers from which we derived an estimate (e.g., 20 minutes, five days a week). For each teacher we interviewed, we derived estimates based on their description of the time they spent on test preparation.

⁵ If we assume that there are a total of 900 hours of instructional time over the course of a school year, this is 6 percent of the instructional time available. Furthermore, previous Consortium work has shown that in a typical year there are really only about 500 hours of instruction time available, taking into account the time spent on start-up routines, special programs and events, holiday-slowdown periods, and test-preparation periods. In this case, 53 hours represents approximately 11 percent of the instructional time available. See Smith (1998). Because the estimate of 500 hours already includes the time teachers spent on test preparation, the 11 percent figure probably overestimates the proportion of instructional time lost, but 6 percent likely underestimates it. The actual instructional time lost is probably closer to 8 or 9 percent.

⁶ See, for example, Rosenholtz (1987); Smith (1991); Urdan and Paris (1994).

⁷ Smith, Smith, and Bryk (1998).

⁸ Smith, Smith, and Bryk (1998).

⁹ In a multivariate analysis that accounts for changes in the entering mathematics score of students in each grade, as well as the characteristics of teachers, we find that there was a sta-

tistically significant increase in the percentage of time teachers devoted to grade-level material in both the seventh and eighth grades.

¹⁰ For example, while math pacing in the third grade increased by a little more than 2 percentage points between 1994 and 2001, it increased by 6 percentage points in the eighth grade. The increase in math pacing was statistically significant for eighth grade but not for third grade. Similarly, although both increases were statistically significant, the proportion of time spent on reading comprehension increased by 2 percentage points among third-grade teachers between 1997 and 2001, while it increased by 4 percentage points among eighth-grade teachers.

¹¹ Smith, Smith, and Bryk (1998).

Chapter 3

¹ Lee et al. (1999); Sebring et al. (1996); Wenzel et al. (2001).

² Lee et al. (1999); Bryk et al. (1998), *Academic Productivity of CPS*; Noddings (1988); Phillips (1997); Shouse (1996).

³ Booth and Dunn (1996); Cooper et al. (1998); Cooper et al. (1999); Eccles and Roeser (1999); Sebring et al. (1996).

⁴ Betts and Costrell (1999).

⁵ In a 2001 school survey of Lighthouse, many principals stated that they had a difficult time getting the most at-risk students to attend. CPS Audit Department (2001), *Profile of After School Programs*.

⁶ Lee et al. (1999).

⁷ The summary measure of academic press we use in this report relies only on student reports of academic expectations. In a previous Consortium report that linked academic press to achievement, Valerie Lee and her colleagues (1999) used a

measure of academic press that combined the student measure we include here with teacher reports of the extent to which the school climate was organized around high expectations.

⁸ Ames (1992).

⁹ Kellaghan, Madaus, and Raczek (1996); Linn (2000); Mehrens (1998); Wheelock et al. (2000).

¹⁰ We do not fully understand the link between increases in after-school participation and improvement in students' reports of academic and personal support from teachers. Even though our measure of academic and social support uses student answers to questions about their reading/language arts and mathematics teachers, it is possible that students who participated in the after-school program may feel greater academic and social support in general and might not be able to distinguish between help and support after school and within class.

¹¹ Lee et al. (1999), 10.

¹² Lee et al. (1999), 17.

Interpretive Summary

¹ Elmore (2002), 33.

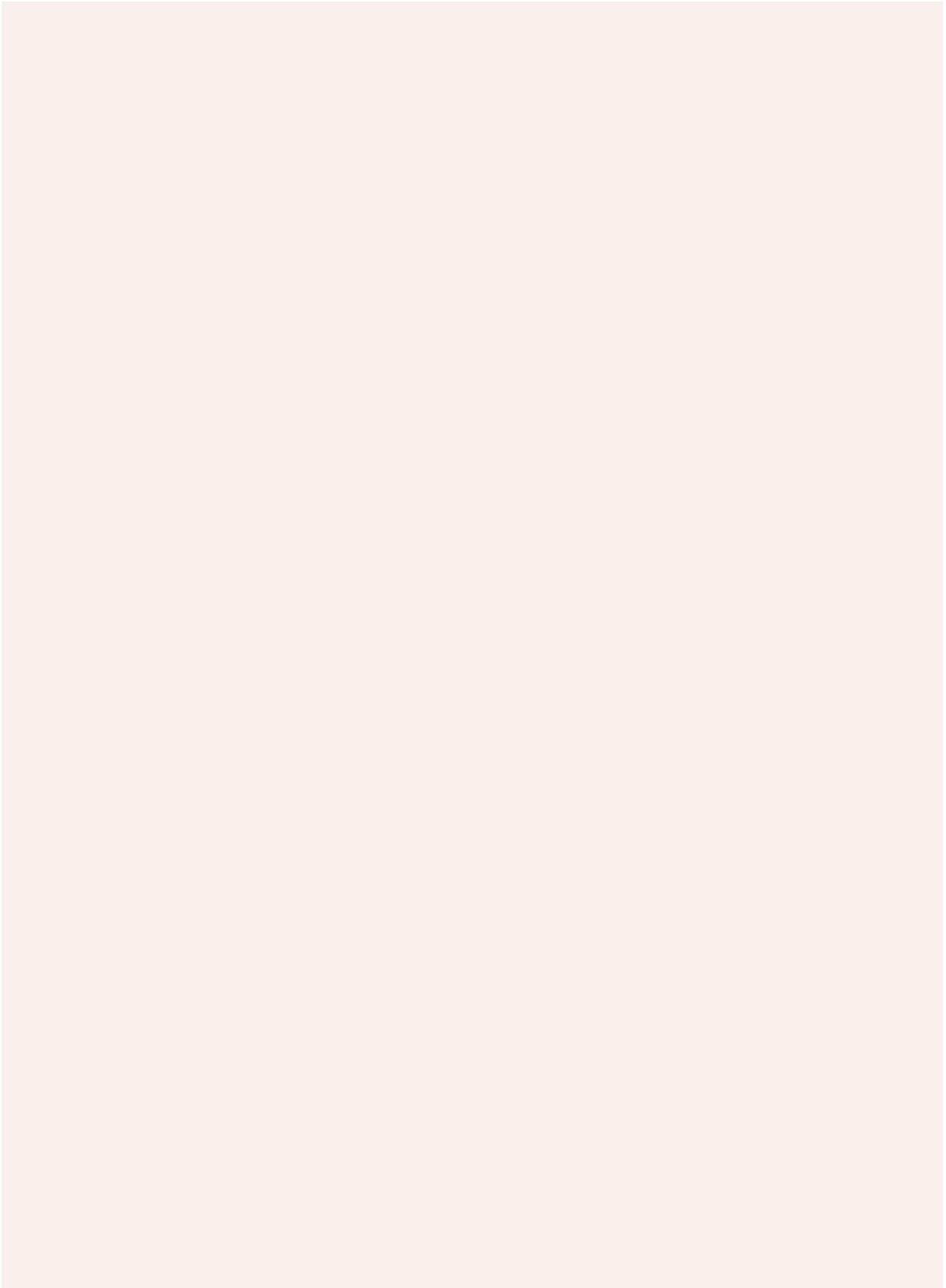
² Roderick et al. (1996).

³ Roderick, Jacob, and Bryk (2003).

Appendix B

¹ When 1994 survey measures or items were not available, change was measured from 1997 to 2001.

² Although not shown in the model, indicators for missing values were also included.



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.
- Ames, Carole. 1992. "Classrooms: Goals, structures, and student motivation." *Journal of Educational Psychology*, 84(3): 261-71.
- Betts, Julian R. and Costrell, Robert. 2001. "Incentives and equity under standards-based reform." In Ravitch, D. (Ed.) *Brookings Papers on Education Policy*. Washington, D.C. : The Brookings Institution.
- Booth, Alan and Judith Dunn. 1996. *Family-school links: How do they affect educational outcomes?* Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Bryk, Anthony S. and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.
- Bryk, Anthony S., Penny B. Sebring, David Kerbow, Sharon G. Rollow, and John Q. Easton. 1998. *Charting Chicago school reform: Democratic localism as a lever for change*. Boulder, CO: Westview Press.
- Bryk, Anthony S., Yeow Meng Thum, John Q. Easton, and Stuart Luppescu. 1998. "Academic Productivity of Chicago Public Schools: A Technical Report." Chicago: Consortium on Chicago School Research.
- Catterall, James S. 1989. "Standards and school dropouts: A national study of tests required for high school graduation." *American Journal of Education*, 98: 1-34.
- Chicago Public Schools. 2001. "Profile of after school programs" (an internal audit report, Office of the CEO, audit department: unpublished document).
- Cooper, Harris, James Lindsay, Barbara Nye, and Scott Greathouse. 1998. "Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement." *Journal of Educational Psychology*, 90(1): 70-83.
- Cooper, Harris, Jeffrey Valentine, Barbara Nye, and James Lindsay. 1999. "Relationships between five after-school activities and academic achievement." *Journal of Educational Psychology*, 91(2): 369-78.

- Darling-Hammond, Linda and Arthur E. Wise. 1985. "Beyond standardization: State standards and school improvement." *The Elementary School Journal*, 85(3): 315-36.
- Easton, John Q., Todd Rosenkranz, and Anthony S. Bryk. 2001. "Annual CPS test trend review 2000." Chicago: Consortium on Chicago School Research.
- Eccles, Jacquelynne and Robert Roeser. 1999. "School and community influences on human development." In M. Bornstein and M. Lamb. (Eds). *Developmental psychology: An advanced textbook* (4th ed.) Mahwah, NJ: Lawrence Erlbaum Associates, 503-54.
- Elmore, Richard F. 2002. "Unwarranted intrusion." *Education Next*, Spring 2002, www.educationnext.org/20021/30.pdf.
- Firestone, William A., David Mayrowetz, and Janet Fairman 1998. Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2): 95-113.
- Firestone, William A. and David Mayrowetz. 2000. "Rethinking 'high stakes': Lessons from the United States and England and Wales." *Teachers College Record*, 102(4): 724-49.
- Grissom, James B. and Lorrie A. Shepard. 1989. "Repeating and dropping out of school." In Shepard, Lorrie A. and M.L. Smith (Eds.) *Flunking grades: The policies and effects of retention*. London: Falmer Press, 34-63.
- Haladyna, Thomas M., Susan B. Nolen, and Nancy S. Haas. 1991. "Raising standardized achievement test scores and the origins of test score pollution." *Educational Researcher*, 20(5): 2-7.
- Hess, G. Alfred 1999. "Expectations, opportunity, capacity, and will: The four essential components of Chicago school reform." *Educational Policy*, 13(4): 494-517.
- Hochschild, Jennifer and B. Scott. 1998. "Trends: Governance and reform of public education in the United States." *Public Opinion Quarterly*, 62(1): 79-120.
- Holmes, Charles T. 1989. "Grade level retention effects: A meta-analysis of research studies." In Shepard, Lorrie A. and M.L. Smith (Eds.) *Flunking grades: The policies and effects of retention*. London: Falmer Press, 16-33.
- Holmes Charles T. and Kenneth Matthews. 1984. "The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis." *Review of Education Research*, 54(2): 225-36.
- House, Ernest R. 1998. "The predictable failure of Chicago's student retention program." University of Colorado, Boulder: School of Education.
- Jacob, Brian A. and Steven Levitt. 2003. "Rotten apples: An investigation of the prevalence and predictors of teacher cheating." *Quarterly Journal of Economics*, 118(3): 843-77.
- Jacob, Brian A. 2001. "Three essays on education and urban School Reform." PhD dissertation. University of Chicago. *Economics*. Manuscript submitted for publication.
- Jacob, Brian A. 2001. "Getting tough? The impact of high school graduation tests on student outcomes." *Education Evaluation and Policy Analysis*, 23(2): 99-122.
- Jones, M. Gail, Brett D. Jones, and Belinda Hardin. 1999. "The impact of high-stakes testing on teachers and students in North Carolina." *Phi Delta Kappan*, 81(3): 199-203.
- Kellaghan, Thomas, George F. Madaus, and Anastasia Raczek. 1996. *The use of external examinations to improve student motivation*. Washington, D.C.: American Educational Research Association.
- Koretz, David. 1988. "Arriving in Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction?" *American Educator*, 12(2): 8-15, 46-52.
- Koretz, David. 1999. "Foggy lenses: Limitations in the use of achievement tests as measures of educators' productivity." Presented at the National Academy of Sciences Conference, Devising Incentives to Promote Human Capital. Irvine, CA.

- Lee, Valerie E., Julia B. Smith, Tamara E. Perry, and Mark A. Smylie. 1999. "Social support, academic press, and student achievement: A view from the middle grades in Chicago." Chicago: Consortium on Chicago School Research.
- Linn, Robert L. 2000. "Assessments and accountability." *Educational Researcher*, 29(2): 4-16.
- Long, J. Scott. 1997. *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.
- Madaus, George F. 1988. "The influence of testing on the curriculum." In *Critical issues in curriculum, NSSE yearbook, part I*. Chicago: University of Chicago Press, 83-121.
- Madaus, George F. 1991. The effects of important tests on students: Implications for a national examination or system of examinations. *Phi Delta Kappan*, 73(3): 226-231.
- Maddala, G.S. *Introduction to econometrics*. Second edition, New York: Macmillian.
- Mehrens, William A. 1998. Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, 6 (13), <http://epaa.asu.edu/epaa/v6n13.html>.
- Newman, Fred M., Anthony S. Bryk, and Jenny Nagaoka. 2001. "Authentic intellectual work and standardized tests: Conflict or coexistence?" Chicago: Consortium on Chicago School Research.
- Noddings, Nel. 1988. An ethic of caring and its arrangements for instructional arrangements. *American Journal of Education*, 96(2): 215-30.
- Perry, Adrienne. 2000. "Coaching for a reading test? Yes, it is possible and necessary." *Reading Improvement*, 37(3): 119-25.
- 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): 300-07.
- Reynolds, Arthur J. 1992. "Grade retention and school adjustment: An exploratory analysis." *Educational Evaluation and Policy Analysis*, 14(2): 101-21.
- Roderick, Melissa. 1994. "Grade retention and school dropout: Investigating the association." *American Educational Research Journal*, 31(4): 729-59.
- Roderick, Melissa. 1995. "Grade retention and school dropout: Policy debate and research questions." *Phi Delta Kappan Research Bulletin*, 15: 1-6
- Roderick, Melissa, Anthony S. Bryk, Brian A. 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, Jenny Nagaoka, Jennifer 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 and Mimi Engel. 2001. "The grasshopper and the ant: Motivational responses of low achieving students to high-stakes testing." *Journal of Educational Evaluation and Policy Analysis*, 23(3): 197-227.
- 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. 2002. "The impact of high-stakes testing on student achievement in promotional gate grades." *Educational Evaluation and Policy Analysis*.

- Rozenholtz, Susan J. 1987. "Education reform strategies: Will they increase teacher commitment?" *American Journal of Education*, August, 534-62.
- Sebring, Penny, Anthony S. Bryk, Melissa Roderick, Eric Camburn, Stuart Luppescu, Yeow Meng Thum, Julia Smith, and Joseph Kahne. 1996. "Charting reform in Chicago: The students speak." Chicago: Consortium on Chicago School Research.
- Shepard, Lorrie A. and Mary Lee Smith. 1986. "Synthesis of research on school readiness and kindergarten retention." *Educational Leadership*, 44(3): 78-86.
- Shepard, Lorrie A. and Mary Lee Smith. 1987. "Effects of kindergarten retention at the end of first grade." *Psychology in the Schools*, 24(4): 346-57.
- Shepard, Lorrie A. and Mary Lee Smith, (Eds.) 1989. *Flunking grades: Research and policies on retention*. New York: Falmer Press.
- Shepard, Lorrie A. and Mary Lee Smith. 1990. "Synthesis of Research on Grade Retention." *Educational Leadership*, 47: 884-88.
- Shepard, Lorrie A. 1990. "Inflated test score gains: Is the problem old norms or teaching the test?" *Educational Measurement: Issues and Practice*, 9(3): 15-22.
- Shepard, Lorrie A. 1991. "Will national tests improve student learning?" *Phi Delta Kappan*, 73(3): 232-38.
- Smith, BetsAnn. 1998. "It's about time: Opportunities to learn in Chicago's elementary schools." Chicago: Consortium on Chicago School Research.
- Smith, BetsAnn, Sophie Degener, and Melissa Roderick. 2001. "Staying after to learn: A first look at Chicago's Lighthouse Program." Paper presented at the American Educational Research Association's Annual Conference. Seattle, WA.
- Smith, Julia B., Valerie Lee, and Fred Newmann. 2001. "Instruction and achievement in Chicago elementary schools." Chicago: Consortium on Chicago School Research.
- Smith, Julia B., BetsAnn Smith, and Anthony S. Bryk. 1998. "Setting the pace: Opportunities to learn in Chicago's elementary schools." Chicago: Consortium on Chicago School Research.
- Smith, Mary L. 1991. "Put to the test: The effects of external testing on teachers." *Educational Researcher*, 20: 8-11.
- Smith, Mary L. and Lorrie A. Shepard. 1988. "Kindergarten readiness and retention: A qualitative study of teachers' beliefs and practices." *American Educational Research Journal*, 25(3): 307-33.
- Steward, Mark B. 1983. "On least squares estimation when the dependent variable is grouped." *Review of Economic Studies*, 50(163): 737-54.
- STATA Base Reference Manual*. 2003. "Tobit: Tobit, censored-normal and interval regression." College Station, TX: Stata Press, 193-205.
- Tobin, James. 1958. "Estimation of relationships for limited dependent variables." *Econometrica*, 26: 24-36.
- Tomchin, Ellen M. and James C. Impara. 1992. "Unraveling teachers' beliefs about grade retention." *American Educational Research Journal*, 29: 199-223.
- Urdu, Timothy C. and Scott G. Paris. 1994. "Teachers' perceptions of standardized achievement tests." *Educational Policy*, 8(2): 137-156.
- Wenzel, Stacy, Mark Smylie, Penny B. Sebring, Elaine Allensworth, Tania Gutierrez, Sara Hallman, Stuart Luppescu, and Shazia Miller. 2001. "Development of Chicago Annenberg Schools: 1996-1999." Chicago: Consortium on Chicago School Research.
- Wheelock, Anne, Damian Bebell, and Walt Haney. 2000. "Student self-portraits as test-takers: variations, contextual differences, and assumptions about motivation." *Teachers College Record*. www.tcrecord.org.

This report reflects the interpretations of its authors. Although the Consortium's Steering Committee provided technical advice and reviewed an earlier version of this report, no formal endorsement by these individuals, their organizations, or the full Consortium should be assumed.

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*

Steering Committee

John Ayers, Cochair
Leadership for Quality Education

Victoria Chou, Cochair
University of Illinois at Chicago

INSTITUTIONAL MEMBERS

CHICAGO PRINCIPALS AND ADMINISTRATORS ASSOCIATION

Clarice Berry

CHICAGO PUBLIC SCHOOLS

Christy Carter
for the Chicago Board of Education

Dan Bugler

Office of Research, Evaluation and Accountability

Barbara Eason-Watkins

for the Chief Executive Officer

ACADEMIC ACCOUNTABILITY COUNCIL

Jorge Oclander

CHICAGO TEACHERS UNION

Deborah Lynch

ILLINOIS STATE BOARD OF EDUCATION

Connie Wise
for the Superintendent

INDIVIDUAL MEMBERS

Gina Burkhardt

North Central Regional Educational Laboratory

Louis M. Gomez

Northwestern University

Anne C. Hallett

Cross City Campaign for Urban School Reform

G. Alfred Hess, Jr.

Northwestern University

Janet Knupp

Chicago Public Education Fund

James Lewis

Roosevelt University

Rachel Lindsey

Chicago State University

George Lowery

Roosevelt University

Angela Perez Miller

University of Illinois at Chicago

Donald R. Moore

Designs for Change

Sharon Ransom

University of Illinois at Chicago

Barbara A. Sizemore

DePaul University

James Spillane

Northwestern University

Steve Zemelman

Leadership for Quality Education

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.

Directors

John Q. Easton

Consortium on Chicago School Research

Melissa Roderick

University of Chicago

Albert L. Bennett

Roosevelt University

Penny Bender Sebring

University of Chicago

Anthony S. Bryk

University of Chicago

Mark A. Smylie

University of Illinois
at 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 0-9726035-2-2

