

Making the Grade (or Not)

Success and Failure in NCLB's World

Gaps in school achievement, as measured, for example, in the eighth grade, have deep roots—deep in out of school experiences and deep in the structures of schools. Inequality is like an unwanted guest who comes early and stays late.

Paul E. Barton, Parsing the Achievement Gap

On a cool August day at the 2003 Minnesota State Fair, Republican governor Tim Pawlenty helped a South St. Paul fifth-grader named Jeremy look up his school's brand-new report card on the Internet. Calling the program Accountability on a Stick—in reference to the fact that people attending the Minnesota State Fair can find almost anything fried and stuck on a stick—Pawlenty praised the report cards and the positive effects that they would have on accountability for Minnesota's public schools:

Traditionally, the measure of our commitment to schools has always been just, "How much are we spending?" That's a good and important measure, but it's an incomplete measure. We also want the measurement to be, "What are we getting for the

money? What are we getting in terms of student learning and performance and accountability?”¹

The report card for Jeremy’s elementary school included summaries of student test scores, demographics, teacher qualifications, and—at the centerpiece of the initiative—a rating of between one (worst) and five (best) stars.

The foundation for these star ratings—fulfilling No Child Left Behind’s requirement to publicize achievement—was a school’s success or failure to make adequate yearly progress (AYP).² In this, the system’s first year, only elementary and combined elementary/middle schools were rated. Stars were assigned in both reading and math, based on the results of third- and fifth-grade achievement tests. For the ratings, schools were compared to other, similar schools based on size and percentage of students qualifying for free or reduced-price lunch. The star ratings were normalized, meaning that only the top performing schools within similar comparison groups could attain the highest ratings. Schools that failed to make AYP in any area could do no better than two stars, regardless of how well they did on the test results for any other grade or subgroup.

In this pilot year, Jeremy’s South St. Paul School received three stars, as did the vast majority of Minnesota’s elementary schools. This pattern held true the following year, when middle and high schools were incorporated into ratings system. The result was largely by design, since high ratings were awarded on a competitive basis. “Schools,” noted the president of the Minnesota teachers’ union, “are graded on a curve, and therefore it would be statistically impossible for every school to perform at the top.”³ Even in Minnesota, not all schools are allowed to be above average.

In many other ways, Jeremy’s school was average as well. It enrolled roughly the same percentage of white students (84 percent) as Minnesota’s public schools as a whole (81 percent), slightly fewer African American students, and slightly more Hispanic students. Just under a third of the students were eligible for free or reduced-price lunch, slightly higher than the state average.

So why did Jeremy’s school receive only an average rating? His school might have represented a snapshot of average Minnesota, with

its test scores reflecting more Jeremy's peers and the level of community resources than anything that the schools were or were not doing. It could also be that Jeremy's teachers and principals have been performing up to a decent standard, but nothing more. Or, it could be neither of these, or some of both. The ratings offer little guidance, especially for those schools that failed to make AYP and therefore received no more than two stars. Given the critical role played by student and community characteristics in educational production (as discussed in chapter 2), it is very difficult to extract the performance of any school from the sociodemographic characteristics of its student body. This is equally true in looking at test scores, success or failure to make AYP, or any list of "blue ribbon" schools based on the results of a state's standardized tests.

This chapter begins the book's empirical analyses, offering a detailed look at the relationships among community characteristics, school leadership, and the specific testing provisions of No Child Left Behind. I focus on AYP as it relates to school and student characteristics and the patterns of leadership among public school principals. The questions for this chapter are

Why do schools succeed or fail to make AYP?

Given the dominance of ethnicity and resource inequality in determining achievement test scores, do school principals' actions matter at all to measured educational achievement?

No Child Left Behind in State and Nation

The details of implementation in the world of No Child Left Behind constitute a somewhat unique and often contentious result of a process of bargaining between state and federal education officials.⁴ Though states devise and implement their own plans, these plans are structured by the law and must be approved by the U.S. Department of Education. There are three main areas where state policies and context can shape the number of schools that make or fail to make AYP. The first is the underlying rigor of the tests used in these assessments, which are likely based on accountability measures already in

use prior to NCLB's passage. The second is the decision regarding where to set the minimum cutoff point for separate evaluation of students of nonmajority ethnicity, those with limited English proficiency, and those with special education needs, along with related questions of whether to use confidence intervals and other techniques to smooth out the data. The third is the composition of the student bodies across states and communities, especially the relative percentages of a state's students that qualify for separate evaluation under NCLB's rules. Some of these factors fall under the states' control; some do not.

The variation among state plans resulting from this negotiated interface in the law's implementation presents an interesting challenge for researchers. Focusing on interstate differences in these details to explain patterns of school success and failure risks aggregating data that may not be meaningfully comparable. Yet ignoring interstate differences can lead to a narrow focus, particularly if one is examining the relationship between these particular provisions of the law and its ultimate success or failure.⁵ What is needed is a body of research that addresses the ways in which state choices about NCLB specifics impact school outcomes but that also includes a more focused analysis of how the commonalities of the challenges of measuring achievement confront school principals, teachers, and education officials across states.

This book takes the second approach: raising and exploring the challenges of defining and identifying educational quality in a specific context—particularly using any standardized measures of achievement—and using these evaluations to make specific suggestions about how to make the law work better. In this approach, interstate differences in the details of NCLB implementation will not be the key factor in NCLB's success or failure. Rather, NCLB will likely stand or fall on the degree to which the framework of the national policy does or does not consider the core challenge of measuring what is good in education. This question has not been sufficiently addressed in all of the debate about NCLB. If the findings in this book are useful, then a next step might be a cross-state comparison of the patterns that emerge in this more focused study.

To do this, I combine what we currently know about how NCLB is

playing out in the country with an extensive analysis of original data from one state, Minnesota. Though most of this book's data for empirical analysis come from one state, this book is not really about Minnesota but rather about key issues that are playing out or are likely to play out across the country. I combine the in-depth exploration of the data from Minnesota with information about what others have found relating to other cities and states, thereby providing a window onto a very complex phenomenon.

Focusing on the on-the-ground effects of No Child Left Behind in Minnesota in particular offers several advantages. First, the state is an early and active adopter of NCLB's provisions, making it a better case for finding positive benefits from NCLB than states that do not consistently comply with law. Compared to other states, it has been a leader in compliance and implementation of No Child Left Behind.⁶ Minnesota also often receives praise for having high-quality schools, traditionally scoring quite high in state-to-state comparisons of national test scores. Thus, looking at what Minnesota's school leaders think about NCLB and at how they are responding to the law offers a sort of best-case scenario for high-stakes testing and accountability.

The empirical data for most of this book combine the results of the Minnesota Schools Survey, an original survey of 1,434 Minnesota public and charter school principals conducted in the fall of 2003, with the results of follow-up interviews conducted shortly thereafter.⁷ Survey questionnaires were sent to all public and charter school principals in Minnesota in November and December 2003.⁸ I followed up these surveys with e-mail interviews of a small but diverse group of survey respondents.⁹ While I do not attempt to draw any broad conclusions from these follow-up interviews, the responses are useful for adding some detail to the comprehensive survey results and at times complicate, challenge, or contextualize the results of the empirical tests. I will draw on the interview results for these more limited purposes.

Though Minnesota lacks a high concentration of students of minority ethnicity, the state offers the chance to combine analyses of diverse urban settings, less diverse suburban settings, and rural settings, which are often omitted from educational policy analysis. Finally, the state's moderate size allowed me to conduct a survey (a

census, actually) of all of the state's public school principals. Such a comprehensive survey, combined with the high response rate, allows greater confidence in the observed patterns than might otherwise be possible.

Of course, the data for this study come from only one state—not coincidentally, the one in which I live. The question arises, in the vocabulary of political science, of generalizability—that is, the degree to which studies in one specific context provide information about what will happen in different contexts. Minnesota exhibits some notable differences from most other states. In addition to its academic excellence, Minnesota ranks at or near the top of the nation in measures of all sorts of good, civic-minded behaviors and attitudes. To paraphrase the framing of social capital theorist Robert Putnam, Minnesotans bowl together.¹⁰ Does Minnesota's status as a state that features high social capital in any way narrow the broader utility of my findings?

I would answer no, although my data do not permit a definitive response to the question. The state's high levels of civic involvement seem likely to lead to more participation in education than in other states. If so, then the danger of reallocating principals' attention away from parents and toward test scores may be less destructive in Minnesota than in states starting out with lower levels of trust and involvement. Fundamentally, however, this book examines the effects of No Child Left Behind on individual principals confronting the challenges posed by uncertainty and limited time across all states, not how the details of state policies will condition these responses.

While, for example, the intensity of the responses of school principals to the threat of sanction may be shaped by how high a state sets the bar for achieving AYP, what method or combination of methods the state uses, or how vigorously the state pursues the sanctions for underperforming schools, the fundamental challenge of measuring educational quality is present in every school in each of the states. Given the considerable variation in states' definitions of proficiency, it may not make much sense to place too much faith in cross-state comparisons of No Child Left Behind's effects. With the stakes involved in NCLB's success or failure as high as they are, it is, therefore, it is not a bad idea to establish whether these themes matter at

all in one state before comparing the specific consequences of NCLB on the patterns of principals' leadership.

Like the rest of the country, Minnesota has witnessed a persistent and growing disparity in the test scores of students of majority and nonmajority ethnicities and from advantaged and disadvantaged homes. A 2002 study by the University of Minnesota's Office of Educational Accountability found some hopeful evidence that gaps between minority and nonminority students were narrowing a bit; nevertheless, these gaps remained quite large. Moreover, the gaps between economically advantaged and disadvantaged students appeared to be widening.¹¹ In the face of these inequalities, the Minnesota Department of Education framed its response to No Child Left Behind specifically in terms of remedying this problem: "This accountability system will help Minnesota solve the biggest challenge it faces in education: closing the achievement gaps between students of color and white students."¹²

The heart of Minnesota's implementation plan is a home-grown, previously existing test: the Minnesota Comprehensive Assessments. The Minnesota Comprehensive Assessments were not the only tests used in the state prior to 2002, nor were they originally designed to test students in grades three through eight and again in high school, as NCLB demands. Like many other states, Minnesota entered the No Child Left Behind era with multiple tests, shifting and politically contentious content standards (particularly in social studies), and the need to significantly ramp up testing and scoring capacity to meet the new federal demands.

As No Child Left Behind requires, Minnesota evaluates schools in four areas: test participation, test proficiency, average daily attendance, and (for high schools) graduation rates. Schools must demonstrate that 95 percent of their students took the test for all of their students and for any of the eight subgroups for which the school has a large enough population to trigger the evaluation (a number that the state generally sets at forty).¹³ These participation rules were designed to prevent schools from encouraging low-performing students to stay home on test day. While these rules discourage cheating, they raise their own problems, particularly with the possibility

that a school could be labeled as failing if one student in one subgroup stayed home on test day.¹⁴

Rather than looking at the total number of students attaining a proficiency threshold, Minnesota assigns points to schools based on the level of proficiency demonstrated according to a five-point scale. Schools receive no points for students achieving “significantly below grade level,” half a point for students “approaching grade level performance,” and one point for each student who demonstrates “grade level performance or above.”¹⁵ In addition, each elementary or middle school was required to demonstrate a 90 percent average daily attendance rate, not broken down by subgroups, and each high school had to demonstrate a 90 percent graduation rate, again not disaggregated. During the 2002–3 school year, Minnesota tested students in grades three and five in reading and math. Middle schools made or failed to make AYP on the basis of their attendance, while high schools were rated on the basis of graduation. In 2003–4, the state added math test results for grades seven and eleven and reading test results for grades seven and ten. For 2003–4, the Minnesota cut-off for subgroups in test proficiency was twenty students for all of the subgroups except special education, where it was forty students.

Not surprisingly, some glitches and concerns have arisen regarding the implementation of No Child Left Behind in Minnesota. Equally unsurprisingly, the state’s urban schools are not doing very well: said one observer, “It’s generally been just poor urban schools, with a smattering of others, that have been labeled by the state as low-performing because their average test scores on state tests have been too low.”¹⁶ While it is too early to tell if achievement scores have sustained across-the-board increases, several concerns are already being discussed. Most of these are related to the rising reading and math proficiency targets. By the 2013–14 school year, every school in Minnesota (and the rest of the country) will have to show that all of its students—and all of its students within qualifying subgroups—have achieved the state’s standards.

That deadline and the 100 percent passage requirement worry Minnesota’s educators and policymakers for several reasons. The first is cost. Though a 2004 report by the Minnesota Office of the Legisla-

tive Auditor—the most comprehensive analysis of NCLB implementation in Minnesota to date—noted that although the costs of implementing the law’s requirements had been “modest, so far,”¹⁷ the researchers could not even estimate the costs of compliance with the 100 percent proficiency requirement or of sanctioning all of the schools that failed to meet it.

The second worry is one of perverse incentives. Given the increased probability of AYP failure in larger, more diverse schools, Minnesota’s educators and administrators fear that No Child Left Behind will make school principals and district officials reluctant to offer services to high-need students, since parents of such students would be likely, under Minnesota’s statewide school-choice program, to gravitate to such schools. The potential for resegregating the schools appears to be another plausible though unintended consequence of the law’s implementation.

District superintendents interviewed for the Minnesota Legislative Auditor’s study worried about the inevitable trade-offs that would occur: “The biggest challenge with NCLB is the need to reallocate existing resources (staff and operating) to meet requirements. This means that we don’t provide some of the other programs that have been in place.”¹⁸ The question of trade-offs is important, and I will deal with it at length in the next chapter from the point of view of the school principal. A principal’s scarcest and most valuable resource is time. School principals are, in the terminology of political scientists, “boundedly rational,”¹⁹ meaning that principals try to make the best decisions that they can given the competing demands on their time. Principals will make most decisions with a great deal of uncertainty, and they will never have enough time to obtain complete information. The worry, from the point of view of No Child Left Behind, is that principals will be forced to reallocate their time in ways that do not lead to better learning for the students but instead lead to an avoidance of the law’s punitive consequences.

Other problems and unintended consequences began to surface as Minnesota moved ahead with implementing its NCLB plan. The first list of underperforming schools, published in July 2003, identified 259 of roughly 1,800 schools as having failed to make AYP. A month later, the Minnesota Department of Education presented a

revised list that contained 93 fewer schools.²⁰ Officials claimed that clerical errors and data problems had resulted in the higher total, but the principals in those 93 schools had a lot of explaining to do during that month. In the winter of 2004, many Minneapolis parents were upset when their children brought home quarter-inch-thick homework packets to be completed over winter break as a means of helping the students perform better on the spring tests.²¹ And, in April 2005, several Minnesota districts were forced to ask parents not to participate in Take Your Child to Work Day, which fell during one of the test weeks, imperiling districts and schools under NCLB's test participation requirement.²²

In the long term, however, Minnesota district superintendents most fear that No Child Left Behind will lead to a wholesale condemnation of the state's public schools. Using simulation techniques based on three scenarios—that Minnesota's public schools will improve academic achievement at a high, medium, or low rate—the 2004 report by the Minnesota Office of the Legislative Auditor (hardly a partisan advocacy group) concluded that by 2014, "most Minnesota schools will not meet NCLB's goals"²³ and that "between 80 and 100 percent of Minnesota elementary schools will fail to make AYP."²⁴ A superintendent in an increasingly affluent and well-regarded suburban district worried that 90 percent of his district's schools would eventually fail and that his district's experiences would hardly be unique: "There's the fear that in a decade public schools will either disappear or look far, far different from what they do today."²⁵ If Minnesota's schools, which consistently rank at or near the top if the nation in test-score comparisons, are headed for closure, then the national picture could deteriorate even more dramatically. The country is accelerating very quickly down a road whose contours are not fully understood and whose potholes may be much larger than anyone has imagined.

Casting a Wide and Indiscriminate Net

The best analogy that I can think of when trying to envision the effects of No Child Left Behind on specific schools—in Minnesota or any other state—is that of throwing large nets into the water, letting

them drift for a while, and bringing in everything that gets stuck in their weave. The effect of NCLB is drift-net fishing for low quality schools.

If we compare schools based only on the percentages of their students meeting proficiency targets, then we can expect to catch some of what we are hoping to catch: schools that we objectively conclude are not doing well because they really are inefficient or otherwise underperforming. However, many schools wind up in our nets even though we do not want them there. Some of these schools may be doing an adequate job but wind up identified simply because they have diverse or high-need student populations. Given the correlation between a lifetime of resource inequalities and lower academic achievement and given the fact that more diverse schools have more ways to fail to make AYP under NCLB's subgroup rules, large diverse schools are much more likely to fail to make AYP, regardless of what their principals and teachers do. These one- and two-star schools will fail to make AYP only because of all of those factors in educational production that have nothing to do with teachers' and principals' actions and everything to do with the legacy of resource inequalities.

We can also expect to catch a few schools that are actively doing good things but that fail to make AYP only because of the relationship between the way that educational quality is measured and the student population. The teachers and principals in these unlucky schools are actively engaged in the kinds of behaviors that decades of effective schools research have shown matter to academic achievement: involved parents, safe and orderly environments, clear standards, and motivated teachers. But these schools are doing so while serving traditionally underperforming students, hence their misidentification under the current system. The very best of urban and rural public schools, charter schools, and alternative learning environments are particularly at risk for this kind of treatment. These high-quality schools are No Child Left Behind's dolphins.

Of course, this analogy only goes so far. After all, in fishing, the net is not designed to scare fish into changing their types or behaviors, as NCLB is supposed to do, though without much guidance on how this process will happen. The point is that large-scale indiscriminate measurement may be more haphazard or dangerous than is cur-

rently imagined. Another unfortunate fishing metaphor hopefully will not prove apt. “Ghost nets” are broken-off pieces of drift nets that float around the ocean, ensnaring creatures and dragging them down to the bottom of the sea. It is not difficult to imagine pieces of No Child Left Behind floating around causing damage long after the law has been scrapped, modified, or left to die from underfunding.

The importance of the relationship between a school’s diversity and its prospects under NCLB served as the basis for John R. Novak and Bruce Fuller’s conclusion that schools are currently being “dinged for diversity.”²⁶ In their study of school success and failure in California, the authors found that “even when students display almost identical average test scores, schools with more subgroups are more likely to miss their growth targets under federal rules set by the No Child Left behind Act.”²⁷

Having diverse student populations increases the likelihood of AYP failure for two reasons. The first, as I discussed at length in chapter 2, is that diversity in student populations typically also means resource inequalities, with all of the attendant negative consequences for test score results. The second reason is more prosaic but just as important. As discussed in chapter 1, test score results under NCLB are disaggregated by eight racial, ethnic, and needs subgroups, with thirty-six different ways to fail at each grade level tested if that grade level contains enough students in that subgroup to trigger AYP. The larger and more diverse a school’s population, therefore, the higher the chances of failure.

The question of where to set the minimum subgroup size for proficiency and, to a lesser extent, test participation is critical to both the number of schools that fail to make AYP and to a state’s commitment to closing the achievement gap between students of minority and nonminority ethnicity—in other words, to whether individual states and the country as a whole live up to NCLB’s promises. States vary considerably in setting the minimum number required to trigger academic performance evaluation for groups of minority, limited English proficient (LEP), or special needs students. Cutoff levels range between ten and fifty students, with many states choosing higher levels for special education and LEP students.

State decisions regarding the subgroup cutoff number matter

when they—in their desire to reduce the number of sanctioned schools—attempt to modify these and other provisions of the law. The political and economic costs of large-scale school failures are quite high. As strategic political actors, therefore, state legislatures and departments of education are already trying to modify the rules of assessment to produce a smaller list of failing schools. By 2004, forty-seven states had applied to the U.S. Department of Education “to approve changes to their accountability plans—changes that in many cases make it easier for schools and school districts to show adequate yearly progress.”²⁸ These proposed changes included increases in the minimum subgroup size, exemptions for small schools, and averaging proficiency data over two or more years.

In 2005 in Minnesota, for example, schools were allowed to use one- to three-year averages of test scores rather than the current year only as a means of smoothing out the inevitable bounces in test score results between years. In addition, the academic achievement of LEP students was examined separately only if the schools had at least forty such students in the grade level, twice the cutoff used a year earlier. These two changes alone resulted in the removal of 197 schools from the list of those failing.²⁹ If the 2004 rules had remained in effect, then what at first appeared to be a tremendous improvement in school quality—a 50 percent drop in the number of failing schools—disappears entirely, making it very difficult to tell whether genuine changes in school quality took place. These kinds of rules changes, with all of their attendant uncertainties, will increase the difficulty of determining whether No Child Left Behind is working.

In spite of the potential importance of minimum subgroup size to AYP compliance in a given school, subgroup cutoff levels alone do not appear to determine what percentage of a state’s schools make or fail to make AYP. As figure 4 shows, no clear relationship appears to exist between the percentages of schools in a state that failed to make AYP in 2004 and that state’s choice of a minimum number of students for separate AYP evaluation, possibly in part because not all states set the cutoff levels uniformly across groups. However, the primary reason for this finding is that states differ on many other crucial characteristics, including the underlying rigor of the tests and the percentages of minority and low-income students that the state serves.

Figure 5 presents the relationship between a state’s percentage of

The Relationship between Subgroup Size and State AYP Failure Rates in the United States, 2004

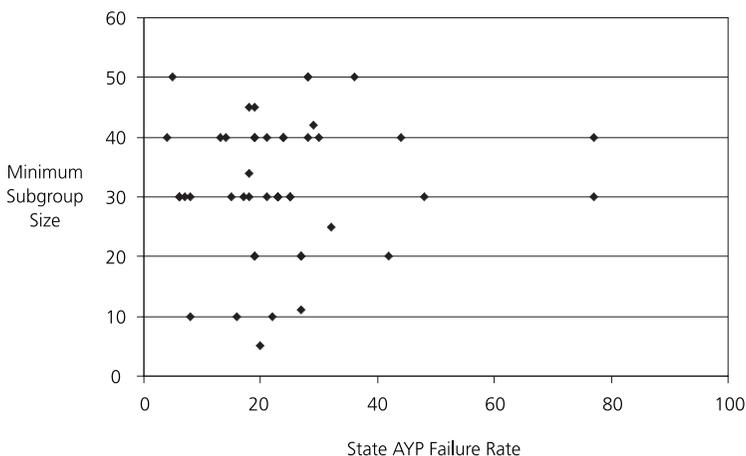


Fig. 4. Subgroup cutoff rules and state AYP failure rates. (AYP data from National Education Association 2006; minimum subgroup size from Pierce 2003.)

minority students and the percentage of its schools that failed to make AYP in 2004. This relationship is positive and statistically significant. States with more diverse student populations are more likely to have higher percentages of failing schools. Of course, none of these basic analyses account for state-to-state differences in the underlying rigor of the tests. The analyses do suggest, however, that getting at the heart of the opportunities and challenges of No Child Left Behind may require a deep rather than broad look at the many details of assessment under the law.

When assigning quality ratings to schools based only on their aggregate test score data—even when taking the community characteristics partially into account—policymakers run the risk of singling out schools for reasons that have nothing to do with events within the school and classroom walls, a troubling pattern that researchers have found in other states. The concern here, therefore, is the degree to which policymakers can extract the true quality of educational production from a set of aggregate test scores, no matter how carefully obtained.

The Relationship between Percentage Minority Enrollment and AYP Failure Rates in the United States, 2004

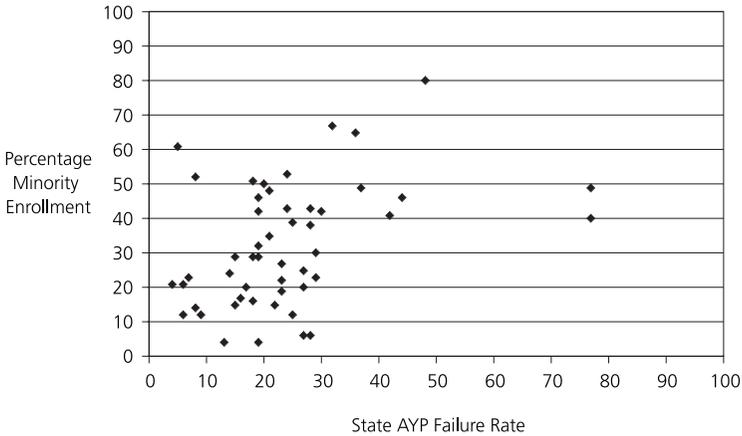


Fig. 5. Minority student enrollment and state AYP failure rates. (AYP data from National Education Association 2006; demographic data from National School Boards Association 2006.)

Comparing Minnesota's five-star traditional public schools based on their student and teacher characteristics raises concerns similar to what has been observed in California and elsewhere (see table 2, which presents the characteristics of Minnesota's traditional public schools that received five-star ratings in either reading or math in 2004).³⁰ In 2004, Minnesota's five-star schools typically enrolled half as many minority students as the rest of the schools in the state, 60 percent as many students who qualified for free or reduced-price lunch, and less than a third as many LEP students. Teachers in five-star schools had higher salaries and were more likely than those in the rest of Minnesota's public schools to have master's degrees.

Comparing the student populations of public schools that made AYP in 2004 to those schools that did not make AYP reveals a similarly troubling pattern (table 3). Minnesota's diverse public schools do not fare particularly well under No Child Left Behind. These schools, however, also disproportionately serve students whose demographic characteristics are associated with fewer lifetime resource

inputs—which are critical to true quality in educational production—and lower academic performance. The question, then, is whether Minnesota’s diverse public schools are being singled out because of their student populations or because of their product’s true quality. Nothing in the current implementation of the No Child Left Behind legislation allows one to distinguish between schools that fail to serve their students and a policy that fails to help the schools.

Most Minnesota public elementary schools qualified in two or fewer subgroup targets in 2003, and most of these schools made AYP. Larger schools and those with more qualifying categories of students; however, were significantly more likely to fail to make AYP in both

TABLE 2. School Characteristics and Minnesota’s Five-Star Public Schools, 2004

	Five Star (1)	All Others (2)
Percentage of minority students	9%	19%
Percentage of students eligible for free or reduced-price lunch	15%	25%
Percentage of special education students	11%	13%
Percentage of LEP students	2%	7%
Percentage of teachers with master’s degrees	45%	39%
Average teacher salary	\$45,043	\$44,253
Number of schools (percentage of total)	181 (12%)	1,307 (88%)

Source: Author’s analysis based on data from Minnesota Department of Education 2003c, 2004b.

TABLE 3. Drift-Net Fishing for Low-Quality Schools: No Child Left Behind’s Minnesota Catch, 2004

	Public Schools That Made AYP (1)	Public Schools That Failed to Make AYP (2)
Percentage of enrolled students classified as . . .		
Black	4%	15%
Hispanic	4%	8%
American Indian	2%	5%
Asian or Pacific Islander	3%	9%
Eligible for free or reduced-price lunch	21%	35%
LEP	4%	13%
Special education	13%	14%

Source: Author’s analysis based on data from Minnesota Department of Education 2003c, 2003d, 2004a.

reading and mathematics proficiency (table A1). Schools that qualified in five or six subgroups failed to make AYP at least 40 percent of the time, while those schools that qualified in only one category (typically for white, non-Hispanic students) failed less than 1 percent of the time. It would be very difficult to defend the assertion that the teachers and principals in affluent communities are better than their counterparts in poorer communities by a factor of forty or more.

Figure 6 presents these relationships graphically by comparing the percentages of Minnesota's public schools that failed to make AYP in 2004 based on the characteristics of their student populations. The patterns of failure in Minnesota's high-need schools are striking. Nearly half of Minnesota's public schools that qualified in the top quarter of schools on the basis of high minority, free-lunch-eligible, and LEP populations failed to make AYP in 2004, more than twice the state's average failure rate for that year. For schools that qualified in both high minority and high free-lunch categories, the failure rate reached 54 percent. As Minnesota and the rest of the nation expand the number of grades tested, even more of the most diverse schools are likely to fail. Across the country as in Minnesota, minority ethnicity and low income tend to go hand in hand. The percentage of students classified as minority and the percentage of students eligible for free or reduced-price lunch are highly correlated.³¹ Add to this the fact that only those public schools eligible to receive Title I funds (given for aid to low-income students) are subject to NCLB's increasingly harsh sanctions, and the playing field looks anything but even.

Figure 7 presents the relationship between the percentages of Minnesota public schools that achieved five-star status or failed to make AYP in 2004 and the racial and ethnic diversity of their student populations. Schools are ordered from left to right, divided into equal increments of seventy-four schools (with roughly 5 percent of schools in each group). The *x*-axis depicts the percentages of minority students in each of these twenty groups of schools. The schools on the right, therefore, contain higher percentages of minority students. The *y*-axis presents average percentages of public schools in these groups that failed to make AYP or were labeled as five-star schools in 2004.

The Percentages of Minnesota’s Regular Public Schools Failing to Make AYP, 2004

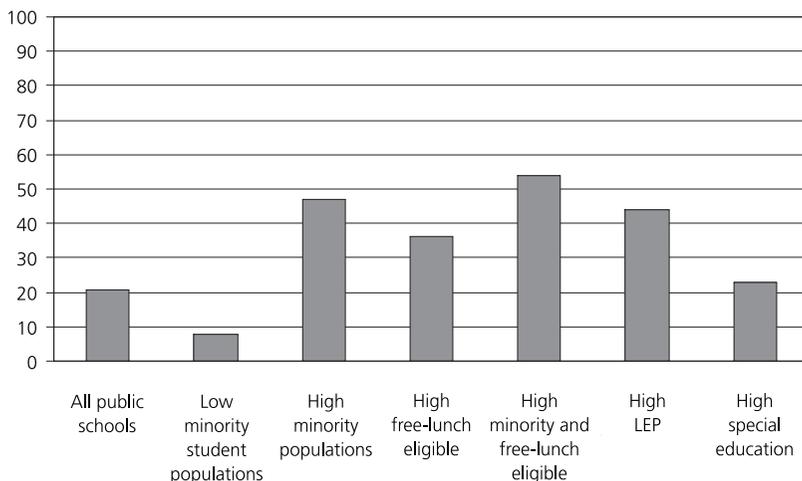


Fig. 6. Student composition and AYP failure. (Data are for regular public schools at all grade levels. Quartiles represent between 369 and 372 of the 1,489 regular public schools in the data set [depending on ties in the quartile cutoff]. Data from Minnesota Department of Education 2003c, 2003d, 2004a.)

The figure demonstrates that the percentage of minority students does not increase evenly in Minnesota’s schools, rising slowly at first and then more sharply at the higher end, concentrating high minority populations in a relatively small number of urban and rural schools. The AYP failure rates are much higher for schools with high minority populations, which also have no chance of being labeled as schools of excellence. The same patterns appear when examining the percentages of students eligible for free or reduced-price lunch (figure 8).

The specifics of the patterns of success and failure under No Child Left Behind may be unique to Minnesota and a few other states where diverse student populations are generally confined to urban centers and a few rural communities with large Native American populations. However, the overall picture confirms what other

Minority Student Enrollment, Five-Star Status, and AYP Failure in Minnesota, 2004

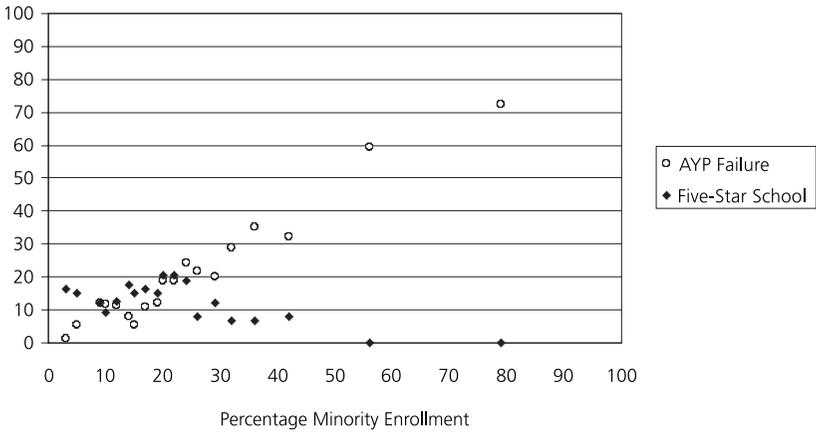


Fig. 7. Student ethnicity, success, and failure under NCLB. (Data from Minnesota Department of Education 2003c, 2003d, 2004a.)

Free-Lunch Eligible Student Population, Five-Star Status, and AYP Failure in Minnesota, 2004

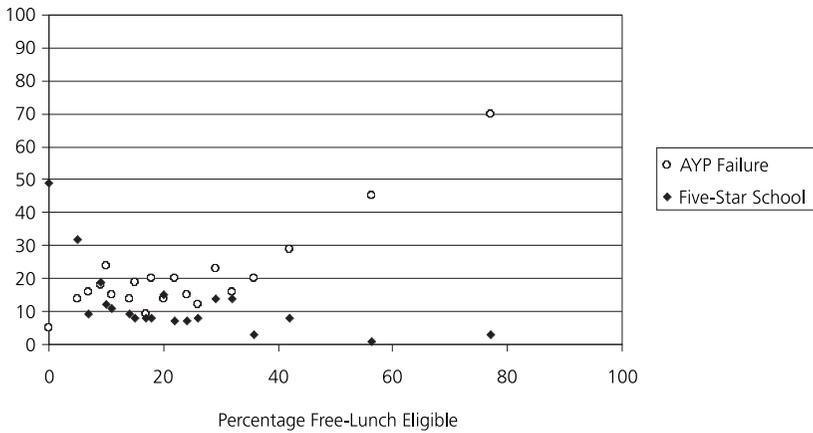


Fig. 8. Student socioeconomic status, success, and failure under NCLB. (Data from Minnesota Department of Education 2003c, 2003d, 2004a.)

researchers have found in other states—that is, that schools that fail to make AYP are concentrated in racially and ethnically diverse and low-income communities.³²

Although striking, none of these results allows me to get at my core concern in this chapter: the degree to which these concentrations cause failure or the degree to which they are symptomatic of other underlying factors and processes. In other words, are high-minority and low-income schools, in Minnesota or anywhere else, doomed to one or two stars and AYP identification and sanction under No Child Left Behind, or can teachers and principals make any difference? After all, if we simply look at differences in failure rates between very different schools, we cannot ascertain if a school has failed to make AYP because of the number of categories for which it qualifies, its resource inequalities, or the performance of its teachers and principals. Examining this question requires more complicated techniques.

Leadership, Success, and Failure under No Child Left Behind

The composition of a school's student body clearly matters to success and failure under No Child Left Behind and its snapshot method of evaluating schools. This is hardly a novel finding; however, the magnitude of its influence on outcomes under NCLB is both striking and troubling. If differences in academic achievement result only from student characteristics, then the link between the threat of sanction and improvement in educational quality becomes almost meaningless. As discussed in chapter 2, while achievement tests do appear primarily to capture factors that schools cannot control, several decades of research into schools that work suggest that leadership can matter, at least at the margins.

I will use a measure of how Minnesota school principals spent their time at work to see whether the allocation of time on core or peripheral tasks affects a school's progress under No Child Left Behind. In the fall of 2003, principals were asked, "During the past month, about how much of your time was spent on the following activities?" The activities listed were (1) facilitating the school's mission, (2) supervising faculty, (3) guiding curriculum development,

(4) building relationships with the parent community, (5) maintaining the physical security of students and staff, (6) managing facilities, and (7) completing administrative tasks. The principals reported their answers on a five-point scale, with one the lowest response and five indicating that they spent the most time on that task.

Because the principals self-reported their activities, their responses may have reflected their perceptions of what I wanted to hear. In this case, the emphasis is not on whether principals spend more or less time on one activity or another but on the differences in these reports between schools based on their status under NCLB.

The patterns of principals' allocation of time raise some concerns. The completion of administrative tasks appears to consume far more time than other activities (table A2). Despite the importance of guiding curriculum development, this area appears to receive the least amount of time. These results agree with the findings of those who critique the negative educational consequences of the effects of public bureaucracy.³³ To the extent that No Child Left Behind increases the administrative burdens on principals, it might reduce students' academic achievement or at least counteract any other benefits that the law might produce.

Of the seven areas in which the principals reported their allocation of time, the strongest positive correlations existed between managing facilities and completing administrative tasks on the one hand and focusing on mission, curriculum, and parents on the other. These two groups of variables are negatively correlated with each other, indicating a pattern that makes intuitive sense in how principals cluster their time. These self-reports are included in the analysis to assess whether any association exists between how principals allocate their time during the academic year and the probability that their schools will be identified as failing to make AYP (based on tests administered in the spring) while controlling for students' demographic characteristics.

To disentangle the controllable and uncontrollable factors of achievement and sanction, I conducted a series of multivariate regressions. With roots in the ruminations of observers of eighteenth-century French gamblers, the theory underlying these techniques seeks mostly to tell us how much we do not know and how much faith

we can place in what we think we do know. Regression techniques provide some purchase on the magnitude of our uncertainty. In addition, they allow us to disentangle various possible explanations for a phenomenon and to explore other possible causes for observed patterns (see the appendix for tables with these statistical results).

The regression techniques used here have two limitations. First, these or any other statistical techniques at best allow us only to gauge the probability that we are being fooled by our data. We can never know the “truth” but only the probability that the relationships that we see in the data resulted from chance and the randomness inherent in this kind of work. Second, these techniques do not permit me to determine the causal relationship between two factors but only whether they are related. Determining causation requires randomized experiments, which are extraordinarily difficult in education. Few parents would sign up for an experiment that offered a good chance that their children would be placed in low-quality schools.

These techniques can, however, show that one factor is probably related to another factor and is independent of third factor or set of factors. These analyses have as their objective determining whether principals’ leadership patterns can make it more or less likely that a school will receive a five-star rating or fail to make AYP based on the test results obtained in the following spring, above and beyond the myriad of student characteristics that we know factor into achievement test scores. In other words, do the best and worst schools in this system differ in their leaders’ focus or only in the composition of their student bodies?

I will begin with the mark of success under No Child Left Behind in Minnesota: the five-star school. Five-star status is treated as a function of the demographic characteristics of the student population (including percentage of minority students, percentage of students eligible for free or reduced-price lunch, percentage of LEP students, and percentage of special education students). The model also includes a variable indicating the principal’s administrative experience, his or her teaching experience, the school’s status as a rural school, the number of students, the highest grade offered in the school, the average teacher salary, the percentage of teachers with master’s degrees, and whether the school is a charter school. These

same demographic and school variables will be used throughout the analyses.

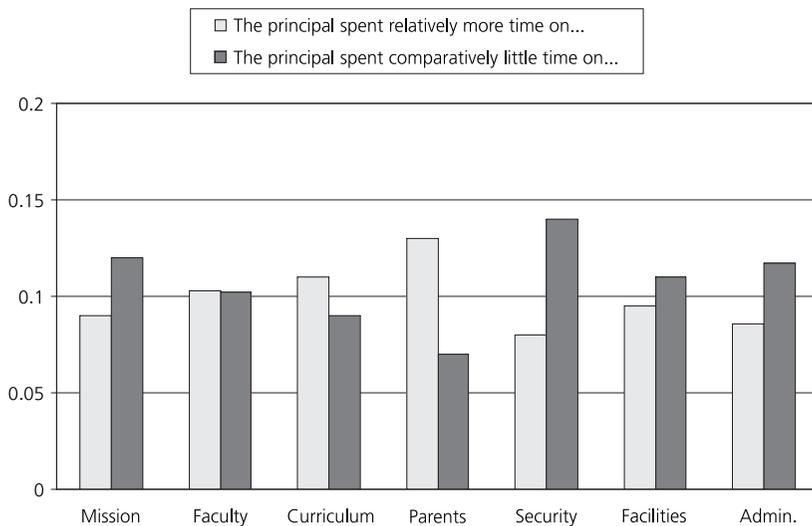
Most important, the model includes principals' responses to how they spent their time during the previous month.³⁴ This is the key variable, allowing me to examine whether the ways in which principals allocate their time have any relationship to success under the five-star schools program, independent of all of the other factors that likely affect whether a school is a top scorer on achievement tests.

Other than the negative relationship between percentages of special education students and five-star status, schools are not more or less likely to receive the five-star badge because of their student populations. This makes intuitive sense when one remembers that under Minnesota's plan, schools are compared to those with similar student populations that made AYP. Given the high correlation between student characteristics and AYP failure, student variables would likely be highly significant in a model predicting the probability of being labeled as a one- or two-star school. Schools with higher teachers' salaries appear to be more likely to receive five stars. I will not comment extensively on the teacher variables, though they are included in the models because they are probably important and may be useful for those who would follow up this study.³⁵

Figure 9 presents the results of a simulation that uses the coefficient estimates from the underlying probit regression of being labeled a five-star school to produce a set of predicted probabilities that incorporates both principals' self-reported time allocation and the student populations with whom they work.³⁶ The data portrayed in figure 9 are estimated probabilities rather than actual percentages because the underlying regression techniques are necessary to try to control for those factors that are highly predictive of success on NCLB tests but are unrelated to the behaviors of public school principals. The bars with the dots represent the simulated probabilities that a school will be labeled as a five-star school if that school's principal spent more time on the specific school policy area than 75 percent of his or her colleagues. The solid bars represent the predicted probabilities for those principals who spent less time on the activity than 75 percent of their colleagues at other schools.

These simulations allow me to interpret whether the statistically

Leadership and the Predicted Probability of Being a Five-Star School, 2004



Question:

"During the past month, about how much of your time was spent on the following activities?" Coded: 1 = "None or almost none," 2 = "Slightly less time than on other activities," 3 = "About as much time as other activities," 4 = "Slightly more time than other activities," 5 = "A great deal of time."

Fig. 9. Principals' leadership patterns and success under NCLB. (Probabilities obtained using Clarify [see King, Tomz, and Wittenberg 2000]. Source: Minnesota Schools Survey 2003. Demographic, star, and AYP data from Minnesota Department of Education 2003b, 2003c, 2003d, 2004a.)

significant differences that I observed in the regressions are substantively meaningful. They do not provide exact point estimates, as in "Principals who spend a great deal of time reaching out to parents will have a 13 percent chance of having their school receive five stars." Rather, these simulations show the rough magnitude of the effect of principals' behaviors in a way that controls for all of the other things that are known to factor into success on standardized tests.

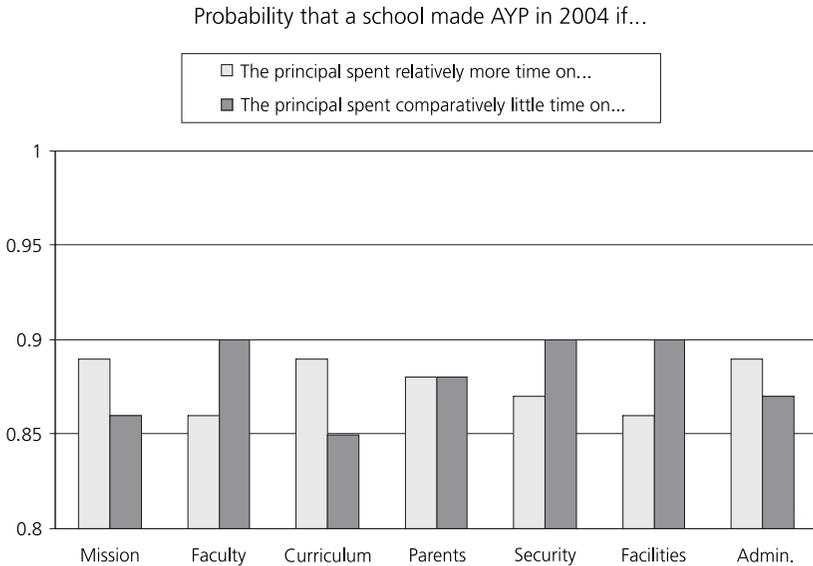
According to these simulations, principals who focus on reaching out to their parent communities are nearly twice as likely to have

their schools labeled as excellent under Minnesota's system. Those who focus instead on security are just under 60 percent as likely to have their schools succeed. The results of the regressions confirm the statistical significance of these differences. Spending more time on curriculum and, somewhat surprisingly, less time on mission are also associated with higher probabilities of success, but these findings are not statistically significant and may result only from the randomness inherent in this kind of research.

That principals in five-star schools spent less time on security is probably not surprising, but it raises an important complication in these analyses. Effective schools research has demonstrated that safe schools are good, but time spent maintaining security is time not spent reaching out to parents or developing and guiding the curriculum. Principals in urban schools likely have to spend more time on security, and they succeed at a lower rate. The danger would be to then assert that having principals spend less time on security causes a school to receive five-star status. This would be a ridiculous and unsupported assertion. Principals do not always control how they spend their time but must always react to events around them. A more critical question is the school's level of safety, orderliness, and security, a point that I will examine in detail in chapter 5.

Figure 10 presents the results of a simulation that examines the other side of Minnesota's compliance with No Child Left Behind: failure to make AYP. As before, the underlying regression estimates are presented in the appendix. Unlike the models of five-star school identification, however, student characteristics matter a great deal to whether a school made or failed to make AYP in 2004, confirming statistically the positive relationship between school diversity and AYP failure. Schools with large minority student populations and those with larger numbers of special education students are twice as likely to fail to make AYP as those with low populations of those two groups. Both of these differences are statistically significant.

An important and significant association also appears to exist between the allocation of principals' time and whether their schools pass or fail the spring round of tests. Mission-oriented principals and those who spent relatively more time guiding the curriculum were more likely to make AYP the following year, while those who were



Question:

"During the past month, about how much of your time was spent on the following activities?" Coded: 1 = "None or almost none," 2 = "Slightly less time than on other activities," 3 = "About as much time as other activities," 4 = "Slightly more time than other activities," 5 = "A great deal of time."

Fig. 10. Principals' leadership patterns and failure under NCLB. (Probabilities obtained using Clarify [see King, Tomz, and Wittenberg 2000]. Source: Minnesota Schools Survey 2003. Demographic, star, and AYP data from Minnesota Department of Education 2003b, 2003c, 2003d, 2004a.)

more focused on managing facilities were more likely to fail to make AYP. While devoting more time to building parent relationships does not appear to be meaningfully associated with a lower probability of AYP failure, spending more time guiding the curriculum and less time managing facilities does appear to pay off.³⁷ Both of these relationships are statistically significant. Of course, these kinds of survey results cannot establish causation or rule out the possibility that these patterns arise as a consequence of other unobserved variables. However, these results are consistent with what we know about effective schools and principal leadership.³⁸

One concern with these results is the possibility that Minnesota's public school principals in high-need schools are failing not because of the legacy of the resource inequalities that they face but because they do not or cannot allocate their time as efficiently. Comparing the allocation of time by principals in high-minority schools with those with smaller minority student populations, however, casts some doubt on this possibility (table A3). Principals in schools with the largest percentages of minority students were more likely to report that they spent a "great deal of time" focused on the school's mission and on reaching out to their parents than principals of low-minority schools. These principals were doing more of what we want them to do with a less active parent community and in the face of serious obstacles to their success under AYP's one-size-fits-all philosophy.³⁹

These principals appear, however, to spend more time supervising their faculty, suggesting the possibility that they have less confidence in their teachers' ability to achieve the school's goals without supervision, a long-standing criticism of urban education in this country. Perhaps these principals would have more favorable views of a modified version of No Child Left Behind that more closely based its assessments on actual leadership and that provided principals with more authority to hire and retain the best possible teachers.

Conclusions

In this chapter, I have confirmed in the specific case of No Child Left Behind and the public school principalship what others have observed about academic achievement in general: a lifetime of resource inequalities matters, but leadership can matter, too. With regard to the test scores by which schools are judged, public school principals appear to operate at the margins. However, as Stanley Kelley Jr. observed about elections and Walter Murphy noted about U.S. Supreme Court justices, being able to successfully operate at the margins, "as every student of politics—academic and practical—knows, is saying quite a bit."⁴⁰

I have looked at the effects of leadership on success and failure under No Child Left Behind's particular and possibly peculiar testing strategy. In the next chapter, I will turn to the effects of No Child Left

Behind on principals' influence and leadership decisions. This turn represents a bit of a detour, as I consider the mostly overlooked fact that No Child Left Behind is not the only game in town. Rather, we are simultaneously implementing a set of reforms that seek to control principals and teachers from the top down (in the form of NCLB) and from the bottom up (in the form of charter schools and other forms of school choice). The question is whether these two seemingly contradictory approaches can coexist within the same educational policy space.