The third academic year of life under No Child Left Behind (NCLB) in Minnesota began on an encouraging note. The number of schools failing to make the grade in 2005 under the law declined by nearly 50% from the year before, from 464 to 247 schools, according to data released by the Minnesota Department of Education in August 2005. Reading and mathematics scores on the Minnesota Comprehensive Assessments, the tests used to evaluate schools, were higher in all grades tested. “These results,” reported Governor Tim Pawlenty, “show that the hard work of Minnesota educators is paying off with real and measurable improvements in the classroom.”

The results of this study can neither confirm nor disconfirm that good things are happening in Minnesota’s schools, and that is precisely the point. Minnesota, like the rest of the country, is currently measuring educational quality in a very indirect way. What has been missing from much of the debate about No Child Left Behind is the realization that measuring educational quality is difficult, given the many factors that contribute to a given student’s performance on a particular test, on a particular day. Accepting this complexity challenges researchers and policy makers to be cautious and contemplative in assessing how well Minnesota’s schools are doing.

My purpose is to explore the relationship among NCLB’s evaluative tools, the determinants of student achievement (those that schools can and cannot control), and the behaviors and perspectives of those closest to the actual moment of educational production. My focus is on Minnesota’s public school principals, the relationship between their leadership and success and failure under No Child Left Behind, as well as their perceptions of the law’s effects on their leadership. I argue that the tools with which we measure achievement within Minnesota need to be rethought. By extracting what we really care about—excellence in educational leadership and teaching—from a set of test-based snapshots, we only capture part of what is really going on in schools and in classrooms. Moreover, the data that we do receive from these tests are probably biased against schools in high-need communities.

This study combines the results of a mailed survey of nearly 1,000 Minnesota public school principals with extensive state data on student and school characteristics, achievement test scores, and status under No Child Left Behind in 2003 and 2004. When a particular principal was responsible for more than one school, a survey was sent to only one of the schools. The response rate was very high, with just under 70% of the principals surveyed responding during the study period. I also conducted follow-up e-mail interviews with a small group of these principals. The survey was conducted in the fall of 2003 with support from a grant from CURA’s Faculty Interactive Research Program, in association with the Minnesota Center for Survey Research, and with assistance from the Minnesota Association of Secondary School Principals and the Minnesota Elementary School Principals’ Association.

Closing the Achievement Gap

At 670 pages long, and with many of its key provisions being modifications of parts of other laws, No Child Left Behind is as large and ponderous as it is ambitious. However, two basic goals define the law. The first goal is “closing the achievement gap between high- and low-performing children, especially the achievement gaps between minority and non-minority students, and between disadvantaged children and their more advantaged peers.” The second goal is to create and implement an assessment regime with significant consequences for those who fail, by “holding schools, local educational agencies, and States accountable for improving the academic achievement of all students” (20 U.S.C. § 6301).

The heart of the testing and sanctions regime under NCLB is something called adequate yearly progress (AYP). Making AYP is a defining quest for schools and districts under No Child Left Behind. Make AYP one year, and then you can start worrying about not making it next year. Fail to make AYP, especially for two or more years in a row, and the consequences become increasingly severe, eventually including reconstituting the persistently failing school as a charter school, replacing all or most of the school personnel, or contracting out for private management.

AYP means that either a sufficiently high percentage of the students in a school or district meet the state’s standards for academic proficiency or that the school or district demonstrates “continuous and substantial academic improvement for all students” (20 U.S.C. § 6301). Test results under No Child Left Behind are looked at in aggregate for all of the students in a grade level and for eight subgroups of students: five racial and ethnic identifiers (White, Black, Hispanic, American Indian, and Asian or Pacific Islander), 2 students who are eligible for free and reduced lunch, students with limited English proficiency, and students who qualify for special education services. The idea of subgroup test proficiency is absolutely central to NCLB, both in its goals and in its implementation. Schools are judged by the performance of all of their students and by the performance of each of these eight subgroups.

Critically, the proficiency test targets for adequate yearly progress apply to a specific subgroup of students only if a school or district has enough students to trigger AYP evaluation for a subgroup, in a specific subject, at a given grade level. The more qualifying subgroups that a school has, the more chances it has to fail, regardless of how it is doing in producing high-quality educational services or how successful it is with other subgroups of students. Large schools with diverse populations are, therefore, at a significant disadvantage.

Challenges to and critiques of No Child Left Behind are coming from many directions—from state governments worried about the costs of implementation, educators worried about the narrowing of curricula in response to extensive testing, school officials concerned about the lack of flexibility in the law, and policy makers worried about the possibility of massive closures of public schools as the requirements of the law gradually increase to mandating that every child demonstrate academic proficiency. What has received less attention, however, is a critical consideration of the core assumption underlying No Child Left Behind: That it is possible to assess the quality of education being delivered to the student by any “objective” measure, above and beyond all of the myriad factors that contribute to academic achievement during the lifetime of the student. Educational quality is a very difficult thing to observe, and its production involves many actors and many factors outside of what happens within the classroom.

Three risks arise from basing school quality assessments solely on the results of standardized tests, no matter how carefully constructed. The first risk is that success or failure to make adequate yearly progress may have more to do with the rules of assessment than with what schools are actually doing. The second risk is the possibility that our evaluations of successful and failing schools are based mostly on things that schools cannot control, including the ways in which these test scores are used. A closer look at patterns of success and failure in Minnesota’s public schools reveals that both of these risks are real, and that they should be a cause for concern and reflection. The third risk arises from the fact that the results of these tests may change the behaviors and relative influence of various actors in Minnesota’s educational system. This may be a good thing, if our evaluations of school quality are based on measures of what is really happening within those schools. If our evaluations are incorrect or unintentionally reflect the influence of other factors, then we run the risk of making these changes using faulty data.

Success and Failure Under No Child Left Behind

Based only on the number of schools making or failing to make adequate yearly progress during the past three years, it is not yet possible to say that NCLB is or is not working. In fact, a closer inspection of the trends in the number of schools failing to make adequate yearly progress reveals that the dramatic decrease in failure rates between 2004 and 2005 may be due mostly to changes in how No Child Left Behind is implemented.

Figure 1 presents the percentage of Minnesota’s public schools that failed to make adequate yearly progress in 2003, 2004, and 2005. Just as the state witnessed a drop in the number of failing schools between 2004 and 2005, it also witnessed a sharp increase in failure rates between 2003 and 2004. Rather than evidence of deteriorating educational quality, however, this early rise in failure rates is likely due primarily to changes in how NCLB was implemented in the second year. Adequate yearly progress status in 2003

---

2 Editor's note: Although inconsistent with CURA's house editorial style, the terms Black and Hispanic have been retained for consistency with the racial/ethnic categories used in the No Child Left Behind Act.
was based on the results of grades 3 and 5 reading and math tests for elementary schools, attendance for middle schools, and graduation rates for high schools. In 2004, however, adequate yearly progress calculations also included the results of grade 7 reading and math tests, along with the results of grade 10 reading and grade 11 math tests. Given the role of qualifying subgroup categories in assessing adequate yearly progress, including more grade levels makes it more likely that schools will fail, regardless of how well they are doing. It only takes failure in one group, in one subject, in one grade level to qualify a school as having failed to make adequate yearly progress for that year.

Just as the dramatic increase in failing schools that occurred between 2003 and 2004 likely reflected changes in the implementation of the law, the sharp decrease in failure rates between 2004 and 2005 may have been due mostly to two changes in the law made in 2004. First, schools were allowed to use one- to three-year averages of test scores rather than the current year only, to smooth out the inevitable bounces in test score results between years. Second, the academic achievement of students with limited English proficiency was examined separately only if the schools had at least 40 limited English proficiency students in the grade level, compared to a cutoff of 20 students in the previous year. A higher cutoff means that fewer schools qualified and, therefore, failed. These two changes alone resulted in 197 schools being removed from the list of failing schools in 2005, according to Minnesota Department of Education data. If one applies the 2004 rules to data for the 2005 school year, the failure rates of Minnesota’s public schools would decline only from 25% percent to 24%, making it very difficult to tell if genuine changes in school quality actually took place.

**Adequate Yearly Progress and Student Characteristics**

The second risk of relying on No Child Left Behind’s method of assessments for evaluations of school quality is that these objective measures may be capturing many contributors to educational quality that principals and teachers cannot control, particularly resource inequalities between students, schools, and communities.

Having diverse student populations increases the likelihood of AYP failure for two reasons. The first 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. Because test score results are disaggregated by eight racial, ethnic, and need subgroups, larger and more diverse schools have a higher probability of failure, even when test scores between schools are exactly the same. The test results currently used under NCLB are cross-sectional, meaning that they take a snapshot of students (in aggregate or within a subgroup) at one point in the year. Consequently, sanctions under NCLB are based not on measuring what a school is adding to the achievement of individual students, but rather on the aggregate peer performance of students as a whole or within a group. Basing sanctions on these snapshot assessments risks sanctioning schools based mostly on the characteristics of the students, rather than the school’s contribution to students’ academic achievement.
Basing sanctions on these snapshot assessments runs the considerable risk of identifying and sanctioning schools based mostly on the characteristics of the students, rather than the school’s contribution to students’ academic achievement, because many things outside of a school’s control show up in the cross-sectional aggregation of student test scores. Determining a school’s success or failure based on these kinds of cross-sectional tests results incorrectly singles out schools with high-need populations, a consequence that researchers John Novak and Bruce Fuller—in their 2003 policy brief, “Penalizing Diverse Schools?”—call being “dinged for diversity.”

Figure 2 presents the relationship between the percentages of Minnesota public schools that failed to make adequate yearly progress in 2004 and the racial and ethnic diversity of their student populations. Schools are ordered from left to right, divided into equal increments of 74 schools (roughly 5% of schools in each group). The x-axis depicts the percentages of minority students in each of these 20 groups of schools. As one moves to the right on the graph, schools have increasingly higher percentages of minority students. The y-axis shows average percentages of Minnesota public schools in these groups that failed to make AYP in 2004.

The first thing to note is that the percentage of minority students is not evenly distributed among Minnesota’s public schools, but concentrated instead in a relatively small number of schools with high percentages of students of minority ethnicity. Within these schools, AYP failure rates are much higher. These patterns are also evident when one examines the AYP probabilities for schools based on the percentage of students at the schools who are eligible for free lunch (Figure 3).

Unless one assumes that education is uniformly and progressively worse in Minnesota’s most diverse public schools, then it appears that these schools are being disproportionately identified and, eventually, sanctioned. The problem is that it is not possible to say if this identification is misplaced or unfair because the system of measurement is not designed to distinguish between schools of varying quality that serve similarly diverse student populations. After all, if we simply look at failure rates among very different schools, we cannot ascertain if a school has failed to make adequate yearly progress because of the number of categories for which it qualifies, the resource inequalities of the school and community, or the performance of the teachers and principals. The question, then, is whether diverse and low-income schools, in Minnesota or anywhere else, are doomed to sanction under No Child Left Behind because of their demographic composition, or if their teachers and principals can make any difference at all.

Leadership, Success, and Failure Under No Child Left Behind

Given the strong correlation between student characteristics and AYP failure, can principals make any difference under NCLB’s snapshot method of quality evaluations? My goal in this section is to see if principals’ leadership patterns can make it more or less likely that a school will make or fail to make adequate yearly progress in the following spring, above and beyond the myriad student characteristics that we know factor into achievement test scores. In other words, I am curious if the best schools and worst schools...
in this system differ in terms of the primary focus of their leaders or only in the composition of their student bodies.

I used a measure of how Minnesota school principals spent their time at work to see if the allocation of time on core or peripheral tasks affects a school's progress under No Child Left Behind. In the survey, I asked the principals, “During the past month, about how much of your time was spent on the following activities?” Using a five-point scale, the principals reported the relative time they spent on the following tasks: facilitating the school’s mission, supervising faculty, guiding curriculum development, building relationships with parents, maintaining the physical security of students and staff, managing facilities, and completing administrative tasks.

In addition to measuring how principals reported their time, the regression analyses that underlie the results reported here also accounted for the demographic characteristics of the student population (including the percentage of students that were of minority ethnicity, eligible for free lunch, of limited English proficiency, and enrolled in special education). The statistical models also took into account 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, and the percentage of teachers with master’s degrees.

Figure 4 presents the results of a simulation analysis. The simulation incorporated both principals’ self-reported time allocation to various tasks and the characteristics of the student populations with whom they work. This produced a set of predicted probabilities that a principal’s allocation of time to specific tasks would impact the likelihood of a school making adequate yearly progress. These data are estimated probabilities, rather than actual percentages. This approach was necessary to try to control for factors that are highly predictive of success on NCLB’s tests but that are unrelated to the actual behaviors of public school principals, such as enrollment, grades offered, and student and community composition. The gray black bars represent the simulated probabilities that a given school will make AYP if that school’s principal spends more time on the specific activity than 75% of his or her colleagues. Conversely, the maroon bars represent the predicted probabilities that a given school will make AYP if that school’s principal spends less time on the specific activity than 75% of his or her colleagues at other schools.

Although I observed earlier in relation to Figures 2 and 3 that student population matters critically to AYP and failure, the results in Figure 4 indicate that there is also an important and significant association between the allocation of a principal’s time, and whether or not his or her school passes or fails the spring round of tests. In short, principals can matter. Mission-oriented principals and those who spend relatively more of their time guiding the curriculum are less likely to have their schools labeled as failing to make AYP the following year. These relationships are statistically significant.

It is important to note that, although spending more time on security and facilities is associated with higher AYP failure rates, this does not mean that principals should or can choose to devote less time to these efforts. Time is a very finite commodity in the principalship, and energy devoted to one area must be taken away from another. Principals’ time is not always theirs to spend. Although few principals would choose to spend a great deal of their time managing facilities, many have to, particularly those in older facilities often found in Minnesota’s urban areas.

What these results do suggest, however, is that there is an important correlation between principals’ activities and the performance of their students, after controlling for all other likely influences on test-score performance. This relationship, however, may be masked by the correlation between test score outcomes and the social and economic conditions under which the schools operate.

The Effects of AYP Failure on Leadership

The third risk identified earlier is that the results of these tests may have real consequences for the behaviors of principals and teachers. There has been much discussion about whether or not teachers teach to the tests. Much less discussion has been devoted to whether or not principals lead to the tests, and whether or not this would be a desirable outcome. Although the underlying study I conducted focused on many aspects of the connection between NCLB and leadership, here I focus on the connection between principals’ perceptions of their own influence and their status under AYP at the beginning of the school year.

Figure 5 presents the results of a simulation based on a different, but related, set of statistical analyses. This figure shows the predicted percentages of principals who report that they have “a great deal of influence” on the specific policy area identified, broken down by whether or not they were the
Hiring

2003

Source: Author’s analysis based on data from the Minnesota Schools Survey (2003).

Figure 5. The Effects of Adequate Yearly Progress (AYP) Success or Failure on Principals’ Perception of Influence on Specific Tasks, 2003

principal of a school that made or failed to make AYP the previous August. Principals were asked to rate their own influence on setting performance standards in their schools, guiding the curriculum, and implementing discipline policy, as well as hiring and evaluating teachers within their schools.

Based on the results of these simulations, principals of schools that have failed to make AYP are 63% less likely to report that they have a great deal of influence in setting performance standards in the school and 59% less likely to report that they have a great deal of influence in guiding the curriculum within the school. Both of these relationships are statistically significant.

During the follow-up interviews, principals—both those who felt that there were positive aspects to No Child Left Behind and those who were more critical—repeatedly mentioned that the law was altering their behavior and influence. Refocusing efforts invariably means time taken away from other tasks, and having a bigger stick with which to motivate teachers has the very real potential of eroding the consensus-building aspects of the public school principalship. As one principal stated:

NCLB is just one small aspect of my leadership. However, it has challenged me to move mountains in short order. It has challenged me to become an expert in best practices in curriculum and instruction. It has challenged me to provide learning opportunities for all staff to increase student achievement. These challenges are very exciting . . . however, they are also very time consuming and draining. . . . The day-to-day operations, being in the hallways with kids, etc., suffer.

Another principal agreed with the leadership consequences, but was much less optimistic:

I’ve been more directive—top-down decisions, even though I have a site team. The pressure is on me from my superintendent to produce higher achievement too. It trickles down to teachers and students and parents. I feel that it puts me in a position to “enforce” rather than “support” good teaching. The mandate of NCLB has become a “do or I’ll hurt you” model . . . make the progress or you’ll end up on the dreaded list.

If one believed that principals were not doing a good job setting performance standards or guiding the curriculum, then this would be exactly the result that one would want to see. The problem is, given that many factors outside the principals’ control affect the test scores under NCLB, we cannot say if this co-opting of principals’ powers is based on a true measure of the quality of services within the schools.

Recommendations

Based on the challenges inherent in extracting school quality from cross-sectional test results that I have discussed here, I suggest several actions and modifications to Minnesota’s implementation of No Child Left Behind.

First, Minnesota’s policy makers should continue with efforts to incorporate methods of assessment that track students over time. Two alternatives to the cross-sectional model currently in use have been proposed and discussed or implemented. The first model, usually referred to as a “growth” model, involves looking at the year-over-year changes in percentages of students who achieve academic proficiency rather than the percentages of students who are or are not meeting the targets. The second model, called the “value added model,” attempts to extract the value that a school is adding to a given student’s achievement by tracking the changes in test scores over time for individual students.

Considerable efforts to move down these alternative paths are already under way in Minnesota and other states, and these efforts should continue. Neither alternative model is perfect, and each raises its own challenges. However, either is preferable to the cross-sectional model typically used. The main issue is whether one of these alternatives will be used alone or in combination with NCLB’s current methods in determining AYP. Most likely, it will be in combination with existing methods. If we were to adopt only a growth model or value-added model of assessment, then wealthy schools would be sanctioned in roughly the same disproportion as poorer and high-minority schools under the current model, given that these students would probably be starting from a relatively high level of achievement. More attention needs to be paid to the actual details of how alternative test-based models will assist or conflict with the snapshot models that will still be in place.

Second, policy makers need to consider adding more direct measures of principal (and school) quality to No Child Left Behind. No matter how well-designed the test, test scores are indirect measures of the quality of leadership and teaching in a school which is, after all, what policy makers are trying to encourage. The dominance of student
demographics in adequate yearly progress calculations attests to this problem. Perhaps policy makers would be better served by trying to measure the quality of leadership and teaching as directly as possible, rather than trying to extract it from the results of tests given to a group of students on a given day. Parent, teacher, staff, and student surveys might all be incorporated into such a system. The goal would be to supplement No Child Left Behind’s exclusive use of test score results with more experience-based data, in a way that allows one to identify those schools with excellent leaders confronting monumental tasks. Unfortunately, and in spite of decades of research into what kind of principalship produces quality education, currently these options are not on the table. Identifying excellent schools in disadvantaged communities and providing them with more autonomy and resources as rewards for their excellence might allow policy makers to get more accurate readings of school quality, distinguish the best schools in the communities that need them the most, and help No Child Left Behind live up to its ambitious promises.

Scott F. Abernathy is assistant professor of political science at the University of Minnesota. Prior to earning his Ph.D. from Princeton University in 2002, he worked as a public school teacher in Wisconsin and as an outreach counselor with homeless adolescents in Boston. His research interests include education politics and policy, public administration, and political development.

This article is based on a larger research project that became the basis for a book titled No Child Left Behind and the Public Schools, which is currently under review with the University of Michigan Press. This research was supported by a grant from CURA’s Faculty Interactive Research Program. The program was created to encourage University faculty to carry out research projects that involve significant issues of public policy for the state and that include interaction with community groups, agencies, or organizations in Minnesota. These grants are available to regular faculty members at the University of Minnesota, and are awarded annually on a competitive basis.

The author would like to thank Robert J. Schmidt, executive director of the Minnesota Association of Secondary School Principals, and P. Fred Stortl, executive director of the Minnesota Elementary School Principals’ Association. In addition, the author would like to thank Rossana Armson, Pam Jones, and Marc Wagoner from the Minnesota Center for Survey Research.