

1

The Politics and Practice of Accountability

MARTIN R. WEST AND PAUL E. PETERSON

As a path breaker, No Child Left Behind (NCLB), the federal legislation signed into law in January 2002, stands alongside the pioneering compensatory and special education laws enacted in 1965 and 1974. In the words of political analyst David S. Broder, NCLB “may well be the most important piece of federal education legislation in thirty-five years.”¹ The crucial aspect of all three pieces of legislation is not so much the money authorized as the policy framework imposed. Compensatory education and special education laws have never provided more than a fraction of the real cost of educating those they professed to serve. Similarly, NCLB increased the federal share of the country’s total school funding by barely 1 percentage point. The federal government’s fiscal role in education has always been small, in recent years hovering around 7 to 8 percent of all public funding of elementary and secondary education, with the balance being covered by local and, to an increasing extent, state revenues.

No, it is not the federal dollar contribution but the direction given to all school spending—whether federal, state, or local—that is key. Just as the 1965 compensatory education law sensitized the country to the needs of minority and low-income students, and just as the 1974 special education law guaranteed for the first time free, appropriate, education to disabled students, so the 2002 legislation redirects educational thinking along new channels. Under its terms, every state, to receive federal aid, must put into place a

set of standards together with a detailed testing plan designed to make sure the standards are being met. Students at schools that fail to measure up may leave for other schools in the same district, and, if a school persistently fails to make adequate progress toward full proficiency, it becomes subject to corrective action.

The new law will not transform American schools overnight. Just as it took decades before compensatory education and special education laws altered American educational practice, so NCLB will only gradually take hold in states and localities across the country. Although a few parents were almost immediately given the right to move their children from failing schools into another public school within the same school district, substantial delay is written into the law in many different ways. Its most strenuous provision—school reconstitution—does not take effect for at least five years, by which time the legislation will already need to be renewed.

The earlier laws also demonstrate how federal intervention can alter school practice in unexpected—and potentially unintended—ways. Compensatory education accelerated school desegregation in the South, enhancing the educational experiences of many African American students. Yet many compensatory education programs also left localities burdened with rules and regulations that may have diminished their educational effectiveness. Similarly, special education has had ambiguous consequences. On the one hand, it opened the door to educational and medical services previously denied to disabled children. On the other hand, the procedures used to define those in need of special education stigmatized many students, minorities in particular, as special when all they needed was a more appropriate classroom climate. In the same vein, NCLB will undoubtedly reshape the focus of public schools in ways yet unforeseen.

The law's arrival on the educational scene raises many questions. How did its passage come about? What were the educational, social, and political forces that gave it shape? What issues will arise in its implementation? What are its likely consequences? Full and complete answers to these questions cannot yet be given. But enough is known—both about the enactment of NCLB and about the workings of precursor accountability policies—that these topics can be explored in an informed and deliberative way.

In part 1 of the collection of essays that follows, four political scientists examine the process leading up to the passage of the legislation, the issues that can be expected to arise in the course of its implementation, and the nature of the political struggle that can be expected to ensue. In part 2, policy analysts explore the practice of school accountability, offering an early assessment of its effectiveness and valuable advice for state policymakers attempt-

ing to achieve compliance in the manner most beneficial for students. Although the accountability provisions of NCLB apply exclusively to schools, some existing rules instead hold individual students accountable. Part 3 examines the issue of student accountability through the lenses of the minimum competency testing and course graduation requirement policies of the 1970s and 1980s, the accountability system erected in Chicago in the 1990s, and curriculum-based exams administered before the completion of high school in many nations around the globe.

The Origins of Accountability

The accountability movement has its origins in long-standing efforts to measure cognitive aptitude and ability. It is premised on the notion that standardized tests can and do measure an important dimension of educational quality. Such a position is increasingly uncontroversial, as evidence mounts that student achievement as measured by standardized tests is strongly associated with both individual and aggregate economic success.² However, broad support for the use of standardized tests as measures of both student cognitive ability and school quality developed only after many years of academic research and trial-and-error application.

Major steps toward more precise measurement of cognitive ability performance occurred during World Wars I and II, when psychometricians, in their search for the best American soldiers, developed tests used to predict performance in a variety of military-related tasks. Testing instruments were further refined during peacetime in the expectation that they could help identify those best suited for further education and high-skilled employment. Over the course of the 1950s, the Scholastic Aptitude Test (SAT), a product of these efforts, was identified as a useful admissions tool by an ever-increasing number of colleges and universities. Soon, college-bound students in most regions of the country felt they had to take the exam.

Although the SAT was originally designed as a curriculum-free device for identifying talented individuals, it gradually came to be seen as a useful tool for measuring the quality of public schooling. When SAT scores began to fall during the 1960s and 1970s, critics of American schools had a quantitative measure to justify their concerns. Between 1967 and 1982, SAT test scores went down by no less than 0.3 standard deviations.

A decline of this magnitude is no trivial matter. A standard deviation is a statistical unit that measures how much scores are spread around their average, facilitating comparisons from one test to another. In terms of test score performance, a difference of a full standard deviation is generally considered

very large. For example, the long-standing difference in average math scores of black and white students in American schools—the much noted black-white test score gap—is approximately 1.0 standard deviation, as is the performance difference between typical fourth and eighth graders. A full standard deviation is also the size of the difference in the math performance of students in the United States and Japan, a difference large enough to have led many scholars abroad to try their hand at explaining the difference. Although only about a third the size of these differences, the decline in SAT scores in the fifteen years preceding 1982 this period was large enough to cause many people to question the direction in which American education was headed.

Before SAT scores began their slide, most Americans thought their schools, which they usually regarded as the best in the world, were continually getting better. The United States was the first country to achieve universal elementary education, the leader in the expansion of secondary education, the earliest to create comprehensive schools that combined students from all backgrounds into a common institutional framework, and a trailblazer in the area of higher education. After World War II, the baby boom accelerated the demand for quality schools. New and attractive buildings were constructed, teacher salaries rose, the numbers of students per classroom steadily declined, and state and local commitments to education deepened. Not only were Americans proud of what their schools had already accomplished, but education also came to be seen as the solution to almost all the country's ailments. It was expected to solve problems associated with civil rights, hunger and malnutrition, immigration, crime, teenage drug use, economic inequality, and other issues too numerous to mention.

As early as the 1950s, a few elitist curmudgeons objected to the quality of instruction in America's schools. A small back-to-basics movement complained about progressive education, the growing number of life-experience courses, the paucity of attention paid to the "great books," and slipping academic standards. Hardly anyone had paid attention to these malcontents, however, until SAT scores began their slide. Some said the fall was simply the result of more students taking the SAT, itself a sign of progress, but others showed this factor could not account for the bulk of the decline.³

Because the SAT was being taken only by juniors and seniors bound for selective colleges, and therefore did not provide a comprehensive picture of overall student performance, the federal government, in the late sixties, funded a new test, the National Assessment of Educational Progress (NAEP), to be administered to a random sample of all students at ages nine, thirteen, and seventeen. By testing a sample of all students at these ages, NAEP results were expected to provide more complete information. Promoters of NAEP

overcame resistance from school officials by making it clear that the test was only a yardstick, not a program designed to hold local schools accountable. The sponsors of the test, the Education Commission of the States (an institution governed by representatives from state governments), foreswore providing any information about student performance for any particular state or locality. (NAEP was even designed so that no one student took the entire test, effectively making it impossible to calculate results for any particular individual or school.) Specific information about particular places would be misleading, it was argued, because student performance may reflect family background and other factors beyond state or local control. Instead, only national results, broken out by ethnic group, region, and community size, would be reported. But even though it was designed so as not to inform anyone about how individual schools were doing, NAEP, ironically, would prove to be a key mechanism in hastening the accountability movement forward.

Many expected that NAEP would prove the SAT wrong, that it would show that the country was progressing after all, just as the *P* in the acronym promised. Unfortunately, NAEP, by revealing more losses than gains in student test performance, only confirmed what the SAT had suggested. Between 1970 and 1982, the performance of seventeen-year-old students on the science examination fell by 0.4 standard deviations. In math, the drop was 0.2 standard deviations. Only the reading scores increased—and these only by a hairbreadth. The trends were not as disappointing for younger students (those age nine and thirteen), but even here NAEP revealed more stagnation than progress.

Concern intensified when Americans discovered further that the United States lagged behind many countries it thought it had left behind. Well-regarded international surveys of educational achievement routinely revealed that U.S. students trailed their peers abroad. The situation deteriorated the longer students remained in school. Among nine-year-olds, U.S. students performed in math and science among the top tier of nations, if not at the very highest levels attained by some Asian nations. By age thirteen, U.S. students had fallen below the international average in these subjects, and by age seventeen, they trailed all the other industrial countries in the world, remaining ahead of only such nations as Lithuania, Cypress, and South Africa.⁴

A decade or more passed before these trends had clear political consequences. But, in 1982, unexpectedly, the Reagan administration made education reform a top political priority. A national commission, appointed by Secretary of Education Terrel H. Bell, issued a report claiming that the quality of America's schools was leaving the country endangered by foreign competition. Released in 1983, *A Nation at Risk* called for a wide range of

reforms that it hoped would reverse the downward trend. Students needed to be given more challenging tasks; teachers needed to be better paid and better trained in the subject matter they taught; states needed to extend the school day and the school year; parents needed to expect more of their children; and a commitment to quality needed to be affirmed by all those responsible for training the young.⁵

That all this should happen during the Reagan years ran contrary to any reasonable expectation. After all, as a candidate for the presidency in 1980, Reagan had called for the disestablishment of the Department of Education that the Carter administration had successfully urged upon Congress just two years earlier. In the view of the early Reagan administration, most of the department's functions should have been turned back to states and localities. Consistent with this perspective, Reagan originally refused to appoint a presidential commission on education, thereby limiting Bell's commission to mere departmental status. Yet when the same commission's report was welcomed with great applause, the Reagan administration reversed course. It dropped all plans to shut down the Department of Education and even increased the level of federal funding. By 1985 the president had appointed William Bennett, an articulate and outspoken educational reformer, as his secretary of education. Still, the call for reform was not backed up by any clearly defined accountability scheme. It would take nearly two decades before another Republican administration would move beyond rhetoric and prompt a real intervention.

Even so, *A Nation at Risk* pushed the nation further toward accountability, principally by raising educational issues higher on state political agendas. In many states, especially in the South, governors saw political profit in making school reform a key plank in their governing platform. Voting rights legislation had made African Americans a significant bloc within the southern electorate, forcing gubernatorial candidates to find ways of garnering support from black and white voters alike. Increased school spending, coupled with accountability requirements, proved useful in this regard. Governors could call for more educational spending to upgrade predominantly black schools. At the same time, they could balance their liberal, pro-spending proposals with a more conservative insistence that stringent requirements accompany the new money, thus ensuring the support of business leaders concerned about the quality of the work force. Pioneering efforts were initiated by governors in Tennessee (under future Republican secretary of education Lamar Alexander), in South Carolina (under future Democratic secretary of education Richard W. Riley), in Arkansas (under Bill Clinton), and, most comprehensively, in North Carolina (under another Democratic presidential hope-

ful, James B. Hunt Jr.). That these governors found accountability an issue that helped elevate them to the national scene was not lost on their peers in other states.

The most significant story was unfolding in Texas. Businessman and future presidential candidate H. Ross Perot, as head of a state education commission, first called for tough requirements that would hold schools and students accountable. Perot captured national attention by requiring that athletes earn C's in class to be eligible to play on varsity teams. More important, in 1993 the Texas legislature heeded his call for testing procedures that would monitor the annual progress of students in each school. Perot's proposals had broad public support and bipartisan appeal. Both Democratic governor Ann Richards and her Republican successor, George W. Bush, embraced the idea, ensuring continuity down to the present time. And when a high-profile study of state NAEP results suggested that, as a result, test scores in Texas and North Carolina were rising faster than scores in other states, the finding helped set the stage for a national intervention.⁶

Initial federal efforts to promote accountability—both those promulgated by the George H. W. Bush administration and by the Clinton administration during its first days in office—had relied on voluntary cooperation from state and local officials. But in 1994, at the prodding of the Clinton administration, Congress imposed the first accountability mandate on the states. It ostensibly required local schools to show, by means of tests, annual student progress toward a state-designated standard of educational proficiency. In short, the core idea underlying NCLB had been conceived.

The pregnancy would prove elephantine. The 1994 law was vague, federal enforcement was lax, many of the state accountability plans were poorly designed, and progress was uneven. Yet the passage of the 1994 law signaled bipartisan support for school accountability. Just as governors found that accountability, coupled with increased expenditure, had an appeal across the political spectrum, so members of Congress—and presidential candidates—found it attractive to hoist their flag on a similar podium. Both George W. Bush and Al Gore embraced accountability in the 2000 presidential campaign.

Despite bipartisan support for the concept, it remained unclear whether the idea would reach infancy. Congress had been unable to agree on new federal education legislation during the waning days of the Clinton administration, and the disputed conclusion to the presidential 2000 election left both Democrats and Republicans embittered. After control of the Senate shifted to the Democrats, continued stalemate seemed as likely as not. Yet politicians often find ways of overcoming their grievances when failure to do so puts them on the wrong side of a popular issue. George W. Bush needed concrete

evidence that he was a compassionate conservative, and Democrats on Capitol Hill could not afford to obstruct passage of a law addressing an issue listed among the top items of voter concern, especially after a coming together of Americans in the wake of the September 11, 2001, terrorist attacks. Predictably, centrist Democrats played a key role in shaping compromises. More surprisingly, President Bush and Sen. Edward M. Kennedy, D-Mass., found it relatively easy to work with one another. For their efforts, they were rewarded with a school accountability law passed by a large, bipartisan majority, which the president signed in January 2002.

The key provisions of the law are well described in Andrew Rudalevige's thorough account of the legislation's origins and of the political process that led to its passage. Briefly, the law requires states to assess the performance of all students in grades three through eight in math and reading each year, with an additional test administered at some point during grades ten to twelve. Test results are to be released to the public. Each year, every school will need to show that students (as well as students within each ethnic subgroup of significant size) are making, on average, adequate progress toward full educational proficiency. Schools that do not measure up to standard will be identified as "in need of improvement," and their parents will have the option to place their child in another public school within the same district. Schools that fail to improve after five years will be "restructured" by the district, with new personnel in charge. States must take an analogous approach with persistently underperforming districts.

The Politics of Accountability

These sweeping new requirements imposed tough mandates on states, localities, and schools. Still, as the Rudalevige essay reveals, the need to achieve a legislative consensus ensured that many aspects of the accountability regime remained unspecified, leaving them to be resolved by the federal Department of Education responsible for its implementation. As a result, some have wondered whether the law would prove to be as influential as many analysts initially expected. It is still too soon to tell, but signs already are evident that the law will not have the force its passage seemed to foreshadow. Here are some of the law's chief limitations:

—Congress left to the states the precise standards to be set, the specific design of their testing instruments, and the administration of their accountability systems. Although all states must administer the NAEP, this national test need not be used as a standard of performance. As a result, standards actually have been lowered in some states.

—If a school fails, parents have the right to send their children only to those nonfailing public schools located within the same school district. States may leave the administration of this requirement to districts, which have scant incentive to ensure that they provide parents with meaningful choice. Evidence is already mounting of districts making it difficult for parents to find an alternative.⁷

—Although annual progress toward full proficiency is required, schools have twelve years to reach this target, and the specific amount of progress required each year is not stated. At least two states have already exploited this loophole by submitting accountability plans that postpone the bulk of the necessary improvements to the end of this twelve-year period.

—The toughest requirements in the legislation do not take effect for several years, opening up the possibility that a subsequent Congress will revise them before they are enforced. For example, a school must fail to make progress for each of five years before the restructuring requirement comes into play.

—Students themselves face neither sanctions nor rewards based on their performance. States need not establish high school graduation requirements—nor standards that govern promotion from one grade to the next. While schools are held strictly accountable, students are not.

In short, the legislation's impact will be highly dependent upon the way it is administered by the states and on the specific strategies they devise to promote improvement. If states establish and maintain high standards, if they develop precise measurement tools that accurately identify both excellent schools and those in need of improvement, if they ask students to pass tests to reach the next grade level and to graduate from high school, if they take action when signs of low performance are evident, if they dismiss principals and teachers who are low-performing, and if these rules are put into place promptly and decisively, then NCLB may have dramatic consequences. But if standards are low, measurement weak, students exempted, few consequences imposed, and implementation postponed, then its influence will be more limited.

Perhaps the best guide to future state behavior is the way states have approached accountability in the past. In this regard, Frederick M. Hess's essay on the politics of accountability at the state level is particularly instructive. He distinguishes between tough, coercive high-stakes accountability, on the one hand, and soft, nice, low-stakes accountability, on the other. The former sets high standards, imposes rigorous testing, and specifies clear consequences. The latter has low standards and few penalties. In his review of state practices thus far, Hess finds a tendency for state accountability systems to

drift from a tough approach toward a softer one. As popular as tough accountability is when first announced, it encounters political opposition as time goes by. Tough accountability has vague, general support from broad constituencies, but, as its coercive teeth begin to bite, the individuals and groups most directly affected complain bitterly. To ease political opposition, standards are lowered, exceptions granted, and penalties postponed.

The pattern is the same for students, teachers, and schools alike. Student accountability is particularly challenging, because rigorous standards will keep underperforming students from advancing to the next grade level with their peers and may discourage some from completing their secondary education. The appearance of punishing students let down by a school system in which few have confidence makes strict accountability for students difficult to sustain.

It would seem easier to hold teachers accountable, if their students underperform. But it is unlikely that many teachers will lose their jobs simply because students in their classes are failing to learn. As Terry M. Moe explains in his penetrating analysis of the role that teacher unions can be expected to play in the implementation of high-stakes accountability arrangements, the unions have a duty to protect the job security of their members, even the weakest ones. In pursuit of this objective, unions have negotiated contracts with school boards that require extensive grievance procedures before an employee can be dismissed. Unions are likely to find mitigating circumstances whenever students' performance falters. Attributing results to the work of a specific teacher will be virtually impossible.

Admittedly, teacher unions have, on occasion, embraced laws requiring high standards. As Jennifer Hochschild recounts in her essay on accountability politics, former president of the American Federation of Teachers Albert Shanker was a particularly forceful proponent of such standards at a time when the accountability movement was just beginning. Similarly, Robert Chase, former president of the National Education Association (NEA), called for reform-minded unions as a supplement to the "bread-and-butter unionism" that focused solely on wages and job protection. But, Hochschild also notes, with the passage of NCLB, the union mood has altered. Those campaigning for NEA leadership in 2002 denounced the new law, calling its title "another empty phrase." Even Chase, in his farewell address, complained that accountability tests "have little or no use in pinpointing the learning problems of students."

Teacher opposition threatens the viability of any accountability scheme, simply because their unions are well poised to shape educational policy, especially at state and local levels. Whereas unions are only one of many partici-

pants in national education debates, they are often the dominant players at the state and local level. They are heavy contributors in school board and state elections. Their expertise is taken seriously in public debates. They employ many full-time group representatives in state capitols. Their members vote more frequently than the average citizen. According to some analysts, they are the single most influential lobby in state government. In many states, all of this political power is backed up with the right to strike, a power that comes into play any time a teacher's job is on the line.

But if teachers cannot be held accountable, can schools be? According to NCLB, a school is to be reconstituted if students do not make annual progress in any one of five years. It is not clear whether this provision will ever be invoked. Depending on how this clause is interpreted, even the poorest of schools could slip through its loopholes. As Thomas J. Kane and his colleagues have shown, test scores can fluctuate randomly in ways that have nothing to do with student progress.⁸ For any year with a bounce downward, chances are good the next year will see a bounce upward. If so, it will be difficult for a school to avoid showing progress in at least one of five years. It remains to be seen whether state accountability machinery will deal adequately with meaningless annual fluctuation in test scores. Moreover, even if persistently failing schools are reconstituted, nothing in the law prevents the same personnel from being reassigned to other schools. NCLB does nothing to relieve school districts of union-negotiated constraints on the dismissal of teachers. Shifting ineffective personnel from one school to the next will not enhance overall educational quality.

Finally, many will argue that schools cannot be held accountable unless they are given more resources. Many of the first state accountability programs were a product of political bargains that gave educators additional resources in exchange for increased accountability for results. As NCLB is being implemented, however, states and localities are experiencing major budgetary shortfalls, which have already forced cutbacks in school expenditures. Teacher organizations and local officials will argue that resource constraints limit their capacity to meet the law's objectives. Indeed, in July 2003 the NEA announced plans to file a lawsuit claiming that NCLB imposed unfunded federal mandates on local schools.

The argument that additional resources are needed may seem odd, given that the cost of designing and implementing a comprehensive accountability system is tiny, dwarfed by the huge cost of other reform strategies, such as class-size reduction or increases in teacher salaries.⁹ Moreover, the 15 percent increase in school expenditures during the 1990s—per pupil public expenditure in real dollars increased in the United States from around \$8,000 in

1990 to more than \$9,200 in 2000—was not accompanied by any noticeable improvement in student performance.¹⁰ Nor have most of the more closely calibrated studies of school expenditure found much connection between dollars spent and learning gained, a sign that greater accountability may be just what is needed to ensure that resources are used effectively.¹¹ But if studies show little effect of resources on school performance, most voters think otherwise. Thus the lack of resources will be invoked as still another reason to slow the application of accountability provisions.

Fortunately, the more persuasively this argument is made, the greater the likelihood that existing resource disparities among schools will decline. Especially where schools are low-performing, states and localities will be asked to see whether performance levels are a consequence of resource limitations. In this regard, Julian R. Betts and Anne Danenberg's analysis of the impact of California's accountability program on resource distribution within the state is encouraging. In contrast to the state's class-size reduction policy, which seems to have adversely affected the quality of the resources available to low-performing school districts, California's accountability law had no such negative impact.

In short, student, teacher, and school accountability all pose major challenges. Keeping intact the necessary political will over the long run is likely to be highly problematic. Hess, Moe, and Hochschild all emphasize that high standards and tough enforcement depend upon the continuous involvement of political leaders responsive to the broad constituencies that support such policies. If authentic accountability is to be established, presidents, governors, and mayors, backed by a well-organized business community, need to remain committed to the effort. Yet such leaders, with their numerous responsibilities, are easily distracted. Fighting wars, preventing terrorism, maintaining economic growth, balancing budgets, and many other issues, too unpredictable to anticipate, can easily shift educational accountability to the back burner. When that happens, well-organized, narrow interests gain the upper hand. All in all, there is every reason to believe that tough, coercive accountability will gradually evolve into something softer, nicer, more acceptable to those directly affected.

The Practice of School Accountability

If No Child Left Behind will be implemented less vigorously than some might wish, it may still constitute a landmark piece of legislation. If a highly coercive accountability system is politically infeasible, a softer version may be enough to prod the American education system forward, though perhaps only gradu-

ally and to a lesser extent than otherwise. Soft accountability may be conceived of as something akin to transparency, the simple reporting of facts about student performance in specific grades at individual schools. Increasingly, this kind of soft accountability is becoming commonplace to American education. Under NCLB, with its requirement that all students in grades three through eight be tested annually, many more parents will have more detailed information about what children are learning at their child's school.

Two consequences flow from this new transparency. First, parents can be expected to express concerns about low or falling test scores to school officials. The bland satisfaction most parents currently express concerning their child's school may already be in jeopardy. Still, it is difficult for parents to organize effectively vis-à-vis well-entrenched bureaucrats, who can bring the full weight of their official authority and educational expertise to bear on the deliberations. For transparency to translate into effective change, it will take more than just political action.

Fortunately, in many parts of the country, parents can influence school policy through a second channel: They can leave communities whose schools are ineffective. Even without well-defined testing systems, parents have long been using available information to distinguish good schools from bad. The better the school, the larger the number of parents who want their child attending that school. As demand grows, property values in the area served by the school increase. When property values increase, the whole community—parents and other property owners alike—prosper accordingly. If the local property tax finances the school, more money becomes available to the school board. On the tide of rising test scores floats the entire community. Conversely, a communitywide ebb tide often accompanies falling test scores, making more than just individual parents unhappy. When property values are adversely affected, transparency can pressure schools to improve even in the absence of explicitly coercive accountability policies. Accountability via transparency works best in suburban areas with multiple school districts, especially where most parents have the financial resources to purchase homes in neighborhoods with first-rate schools. Unfortunately, it is less effective in holding schools accountable in urban districts attended by large numbers of low-income, minority families.

Self-enforcing accountability of this kind may already have arrived in many states. In their essay opening part 2 of this collection, Eric A. Hanushek and Margaret E. Raymond report that by 2001 thirty-one states had accountability arrangements of one form or another. It is true that many of them were of the soft variety. The measuring sticks used by many states were flawed, their enforcement mechanisms weak, and their standards not

particularly high. Some merely reported test results, depending exclusively on transparency as the mechanism for improvement. But despite these apparent limitations, Hanushek and Raymond find that NAEP scores rose more rapidly during the 1990s in the states that had accountability arrangements than in states that did not. Hanushek and Raymond are the first to say their results are preliminary. Nevertheless early signals suggest that soft accountability is moving many schools in the right direction.

Other essays in this section explore specific aspects of school accountability in greater depth. Thomas J. Kane and Douglas O. Staiger consider an oddity of accountability plans that ask schools to ensure that students from all ethnic backgrounds make test score gains. Such rules appear to penalize integrated schools, even when they are doing as well as more segregated ones, because the multiple categories within which progress needs to be made increase the risk that random fluctuation will wrongly make them appear ineffectual. Kane and Staiger accordingly recommend that schools be judged by their performance as a whole, not by group-specific performances. Tom Loveless next points out that many state accountability schemes provide misleading information about the performance of charter schools, especially small ones and those that serve at-risk populations. He foresees adverse consequences of NCLB—unless consideration is given to the smaller size of charter schools, their relatively disadvantaged student populations, and the fact that many will have formed only recently.

If these essays suggest that much remains to be sorted out, another suggests that school accountability has already had positive effects. Julian Betts and Anne Danenberg evaluate a voluntary intervention program for underperforming schools in California in which the state gave schools additional resources if they agreed to stiff sanctions when students failed to improve. A larger number of underperforming schools applied than could be funded, so participants were selected by lottery, making it possible for the authors to conduct the first randomized field trial of an accountability scheme. When Betts and Danenberg compare the test scores of participating schools with those who applied for the program but were not selected, positive differences are observed. If these results hold elsewhere, school accountability may work after all.

Student Accountability: What Works?

If No Child Left Behind is designed to hold schools accountable, it places no direct burdens on students themselves. It does not require standards for high school graduation or levels of performance for passing from one grade

to the next. Although nothing in the legislation prevents states from instituting such standards on their own, they are under no federal mandate to do so. Yet the student is the learner, the one person whose engagement in the educational process is essential to the enterprise. If a student is attentive, curious, enthusiastic, committed, and hardworking, much can be accomplished—even with limited resources. Abraham Lincoln, raised by a stepmother and near illiterate father (of whom he seems not to have been particularly fond), managed to write the Gettysburg Address, arguably the finest piece of American prose ever written, even though he had less than a year's worth of formal schooling. Lincoln educated himself by reading—and rereading—Shakespeare, the Bible, and a few classics neighbors loaned him. Not every student can be as resourceful as “Honest Abe,” but systems that try to get teachers to work harder will not have much effect if students are unresponsive.

However essential student engagement may be, it is easy to see why NCLB dodged this issue. Consider the requirement that students pass an examination to graduate from high school. If the standard is set too high, many students will fail, provoking an outcry among parents and educators. As Frederick Hess points out in his essay on accountability politics, tough initial requirements are eventually relaxed—by stretching out the compliance period, lowering the passing standard, and permitting students to have a second chance. If the standard emerging from this process is too low, nearly every one passes easily, and the signal goes out that not much is required to obtain a high school diploma. The policy has an effect precisely the opposite of what is intended.

Such may have happened during the 1970s when many states set minimum competency standards for graduation. Precursors to the contemporary accountability movement, minimum competency laws were the first to mandate that students achieve a certain level of test performance, if they were to graduate. The tests proved a very soft form of accountability. States set passing grades at a level almost every high school graduate could achieve, allowing minimum competency to live up to its name. Nearly everyone who tried seems to have eventually passed. As Hess points out, the requirement became so bland it faded into the educational woodwork, eclipsed by the educational excellence approach endorsed by *A Nation at Risk*, an approach that rejected these minimalist reforms as hopelessly inadequate.

The first paper in this collection's third section, which focuses on student accountability, offers the first rigorous assessment of the long-term impacts of minimum competence tests. In 1990 the U.S. Bureau of the Census

asked 5 percent of all households a detailed set of questions about their members' education, income, place of residence, and many other matters. Using this resource, Thomas Dee has devised a sophisticated way of estimating the effect of state-level reform programs. He examines course taking in high school, high school graduation rates, college attendance rates, employment in 1990, and wages in 1990 for those affected by the reforms, as compared with those who completed their schooling before the reforms took effect. Controlling for other factors, he offers the best available estimate of the long-term impact of the early effort to hold students minimally accountable.

For high school students taken as a whole, the tests had little impact, one way or another. Apparently the standards were so minimal they did not deter high school graduation, but neither did they stimulate more learning that would pay off in higher levels of college attendance, dependable employment, or higher wages. Their only significant effect seems to have been to reduce the number of math and science courses students took, perhaps because the undemanding nature of the tests implied that such courses were unnecessary. But for two groups, the minimum competency movement had broader effects. African American male students, who, on average, score much lower on these kinds of tests, had modestly lower high school graduation rates after the tests were introduced. Despite these lower graduation rates, the tests had a positive impact on future black male employment. For black males as a whole, then, very soft accountability seems to have had a modestly positive long-term impact. White females had slightly lower employment rates, a result difficult to interpret, inasmuch as no other effects were detected for this subgroup.

Dee also considers the impact of increasing the number of academically oriented courses required of students, a central component of the states' response to *A Nation at Risk*. In many states, students were expected to take additional courses that offered at least nominally academic instruction in English, science, math, and so forth. Once again, the reform was softened by the fact that principals, department heads, and teachers could decide what was to be taught under the academic label. Dee nonetheless finds that the new course requirements for graduation had noticeable effects. Not surprisingly, the clearest impact was on the type of courses a student took. After the reforms, many more students took courses that were at least apparently academic in content. Turning to long-term outcomes, Dee finds more diverse impacts. On the one hand, the reforms depressed high school graduation rates, presumably because some students found such courses not to their liking. On the other hand, they had positive effects on future employment,

both overall and especially for black males. Academic courses in high school seem to make for more productive workers later in life.

More recently, student accountability has been given a serious trial in Chicago. In 1995 Mayor Richard Daley appointed Paul Vallas as superintendent of Chicago's schools and gave full backing to the superintendent's efforts to raise student performance by imposing a new set of requirements on students. Tougher high school graduation requirements were put into place, and students in grades three, six, eight, and nine were expected to pass a test, if they were to advance to the next grade. Students who failed to pass the test could get a second chance if they went to summer school. Although Chicago's plan also included school accountability measures similar to those mandated by NCLB, as of spring 2003 the district had yet to subject any elementary or middle school to this penalty.

Two essays in part 3 assess the impact of this, the longest running student accountability scheme to have been closely examined by the scholarly community. Their analyses reveal just how hard it is in the near term to draw firm conclusions about the impact of accountability policies, especially on the basis of the aggregate evidence typically available. At first glance the reform seems to have boosted test scores dramatically, by as much as half a standard deviation (approximately half the black-white test score gap). But at least some of this gain is more apparent than real. As Anthony S. Bryk points out, more students were being retained in their previous class for a year, more were assigned to special and bilingual education programs (exempting them from testing), and the test day was shifted back a month, allowing for additional instruction. All of these moves helped lift the test score average, even without any real improvement in the quality of instruction.

Fortunately, in the case of Chicago, the availability of rich student-level data allows statistical adjustments to be made for most of these factors. When that is done, the gains are less remarkable. But both Bryk and Brian A. Jacob, in a second analysis, identify a noticeable improvement in performance after the city's reform plan, with its heavy emphasis on student accountability, was put into place. Jacob also analyzes performance on individual exam questions to evaluate the meaningfulness of changes in test scores, showing that they reflect a genuine increase in skills. Less clear is whether these underlying gains constitute a one-time impact or whether they are evidence of a more productive school system.

Economist Ludger Wößmann considers an older, more enduring accountability system that has been practiced worldwide and remains intact in many nations: curriculum-based examinations offered to students in their final year of high school, examinations whose results are given great weight by universi-

ties and employers when students leave secondary school. These exams are generally offered in a variety of subject areas, and students can pass them at different levels. As such, they can challenge even the most able students, while remaining within reach of less talented ones. Although virtually unheard of in the United States, such exams are a prominent instrument of educational accountability in numerous school systems around the world. Well-known examples include the General Certificate of Secondary Education in Britain, the Baccalaureate in France, the Abitur in some German Länder, and the Hesanchi in Japan.

Wößmann's magisterial paper provides the best estimates to date of the effects of curriculum-based graduation examinations on student achievement. Because these exams are generally administered on a nationwide basis, estimating their effects requires data on student performance from many countries. Wößmann examines results from the math and science tests administered by the International Education Association to seventh- and eighth-grade students in thirty-nine countries in 1995 and, again, in thirty-eight countries in 1999. Using state-of-the-art econometric techniques, he adjusts for other differences between countries to isolate the independent effect of curriculum-based graduation examinations. Results are consistent for the two, separate administrations of these tests. Overall, he provides convincing evidence that students in middle school perform at a higher level in math and science when the prospect of a demanding examination at the close of high school awaits them.

Wößmann further advances the literature on this subject by offering a comprehensive theoretical rationale for why curriculum-based examinations should prove effective. They provide a clear, external standard, outside the school, against which student performance can be measured. Performance has real consequences for a student's future. The student does not simply pass or fail; instead, the exams challenge students across the ability spectrum. As an external standard, it motivates peers to work together, teachers to coach students in their lessons, and parents to ensure their children are receiving the best possible education. Armed with reliable information about student performance, each of these stakeholders is well equipped to pursue these goals effectively. Consistent with this logic, Wößmann provides evidence that external exams are most beneficial when teachers have the most flexibility and autonomy. Apparently, teachers left to their own devices can find the best learning solutions when external exams hold their students accountable. The United States might well give further thought to an accountability system other countries have had in place for nearly a century.

Conclusions

Much can be learned from reading the essays that follow. Though *No Child Left Behind* is undoubtedly the most important piece of education legislation in thirty-five years, it does more to initiate a political process than to decide it. So much has been left to state and local governments, the most important political battles are more likely to be waged at these levels than in Washington. As much as political leaders may insist that they have established clear standards of performance and have ended what presidential candidate George W. Bush called “the soft bigotry of low expectations,” much remains to be decided.

If the past is any guide to what will happen in the next few years, softer forms of accountability are likely to be the norm. NCLB itself leans in this direction. Although it requires annual testing, annual progress, parental choice, and, in the extreme, school reorganization, close inspection of these requirements reveals many gaps. The annual testing is only for those in elementary and secondary school. No high school graduation test is required. Exactly how much annual progress toward what standard is left unspecified—until twelve years are up (a date far enough in the future that new legislation could easily supersede this rule). Parental choice is limited to nonfailing public schools within the same school district. School reorganization is not to be posed for five years, and the exact conditions that warrant this intervention—and just how draconian it will turn out to be—are left unstated.

Given the flexibility in the law, it will be up to states and localities to interpret its terms. The process will be political, heavily influenced by the teacher unions whose members are subject to its provisions. They will press for softer rules—weaker standards, postponed deadlines, and minimal consequences for teachers. Unless governors and mayors take strong stands on the other side, the union position will carry great weight.

But if soft accountability is to be expected, it may be sufficient to make a difference. If student performance is transparent to parents and community residents, this by itself will place new pressure on schools, which will then expect more of their students. Still, one element is missing in the national legislation. Too little attention has been given to holding students accountable, despite accumulating evidence that this is where the greatest immediate gains could be achieved. The lost opportunities that result may weigh most heavily in the nation’s distressed urban school districts, where transparency alone is least likely to stimulate improvement. States that take the initiative

to experiment with curriculum-based examinations administered before graduation may be positioning themselves as leaders in the next generation of accountability.

Notes

1. David S. Broder, "Long Road to Reform: Negotiators Forge Education Legislation," *Washington Post*, December 17, 2001, p. A1.

2. For a recent review of this evidence, see Eric A. Hanushek, "The Failure of Input-Based Schooling Policies," *Economic Journal*, vol. 113, no. 485 (February 2003), pp. F64–F66.

3. Diane Ravitch, *National Standards in American Education: A Citizen's Guide* (Brookings, 1995), pp. 64–70.

4. Erik A. Hanushek, "The Seeds of Growth," *Education Next*, vol. 2, no. 3 (Fall 2002), p. 15.

5. National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Department of Education, 1983).

6. David W. Grissmer and others, *Improving Student Achievement: What NAEP State Test Scores Tell Us* (Santa Monica, Calif.: RAND Corporation, 2000), pp. 55, 99–100.

7. Ron Brownstein, "Locked Down," *Education Next*, vol. 3, no. 4 (Spring 2004), pp. 40–46.

8. Thomas J. Kane and Douglas O. Staiger. "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives*, vol. 16, no. 4 (Fall 2002), pp. 91–114; and Thomas J. Kane, Douglas O. Staiger, and Jeffrey Geppert, "Randomly Accountable," *Education Next*, vol. 2, no. 1 (Spring 2002), pp. 57–61.

9. Caroline M. Hoxby, "The Cost of Accountability," NBER Working Paper (Cambridge, Mass.: National Bureau of Economic Research, March 2002). See also General Accounting Office, *Title I: Characteristics of Tests Will Influence Expenses; Information Sharing May Help States Realize Efficiencies* (May 2003).

10. On the increase in per pupil expenditures, see Caroline M. Hoxby, "What Has Changed and What Has Not," in Paul E. Peterson, ed., *Our Schools, Our Future: Are We Still at Risk?* (Stanford, Calif.: Hoover Institution Press, 2003), pp. 101–02. On performance, see Paul E. Peterson, "Stagnant Schools," in Peterson, ed., *Our Schools, Our Future: Are We Still at Risk?* pp. 39–72.

11. Hanushek, "The Failure of Input-Based Schooling Policies," pp. F64–F98.