

## The Best-Laid Plans

And what is good, Phaedrus, and what is not good—  
need we ask anyone to tell us these things?

*Robert Pirsig, paraphrasing Plato's The Phaedrus*

It seemed like a good idea at the time. Samuel W. King, the newly appointed superintendent of the Portland, Oregon, schools, wanted results. Eager to improve his town's public education, King turned to standardized tests. He initiated a comprehensive system of test-based accountability, publishing the test results in the local newspapers alongside the names of the individual students, teachers, and schools. Uniform and administered annually, these tests were designed to determine student promotion and to serve as the basis for evaluations of teacher competency.

And so, in 1874, King administered the first round of his tests to Portland Oregon's fifty-four high school students.<sup>1</sup> Only eleven passed. King's policy of publicizing test results drew immediate opposition from parents as well as teachers. The superintendent, however, stood by his tests. "Next to a New England climate," he wrote, "these examinations necessitate industry, foster promptness and encourage pupils to do the right thing *at the right time*."<sup>2</sup>

The rest of the community did not agree, and a coalition of teachers and parents "forced King's resignation in 1877."<sup>3</sup> His successor,

Thomas Crawford, was quick to backtrack on his predecessor's reforms, keeping the test results from public consumption and criticizing their influence on the school, students, and curriculum. "It is not the excitement that is the source of much anxiety to the school authorities," Crawford argued, "but the useless and almost criminal squandering of precious time."<sup>4</sup> King's plan had other unintended consequences as well. The competitive spirit unleashed by these tests, according to the plan's detractors, had done "incalculable injury . . . both to the teachers and to the pupils of our free schools, resulting from a spirit of rivalry on the part of the teachers."<sup>5</sup> The incentive structure induced by the high-stakes nature of the program produced some highly undesirable bureaucratic behaviors, such as teachers encouraging suspensions of lower-performing students so that they would not drag down classroom averages.

We have traveled down the accountability road in education many times since King's experiment, and we have often learned—or failed to learn—the same lessons. Accurate, workable systems of test-based accountability are very hard to create and implement. They often produce apparent gains at first, but test scores inevitably reach a stubborn plateau. Testing often produces all sorts of unwanted and unforeseen changes in the behaviors of those in the school system. And they are usually abandoned or watered down in the face of organized opposition by those who stand to lose from their implementation.

This time, it might be different. Or it might not.

The No Child Left Behind Act of 2001 (NCLB), signed into law by President George W. Bush in January 2002, is at once more ambitious, more sweeping, and more consequential than any previous accountability initiative in American education. If properly implemented and sufficiently funded, NCLB holds the promise of being one of the great liberal reforms in the history of U.S. education.<sup>6</sup> This may sound surprising, given that the measure was passed under a Republican administration, but it is—or can be—true. The reason lies in the goals. Desegregation and the Americans with Disabilities Act were both about equality of opportunity; No Child Left Behind aims to provide equality of outcomes.<sup>7</sup> This is a very radical and ambitious goal. No longer content to provide access to education for tra-

ditionally excluded student populations, we are now demanding that these students receive equally good educations. In other words, we are now demanding equality of quality.

To reach this goal, No Child Left Behind places significant responsibilities on state educational agencies, school districts, principals, and teachers. It is high-stakes accountability with teeth, offering consequences for failing schools much more significant than public praise or embarrassment. Unfortunately, we have placed these liberal hopes on a large and powerful ship that is heading the wrong way.

My evidence suggests that any confidence placed in No Child Left Behind is probably misplaced. Its promise to close the achievement gap between the advantaged and the disadvantaged will go unfulfilled unless the law is significantly rethought and revised. The source of the problem lies in a failure to consider that the interface between the mind of the student and product of the school is a very difficult thing to observe and measure and may in fact be unmeasurable with complete certainty.

Unless reworked, NCLB will probably not close the achievement gap because it was not set up to distinguish between good and bad schools serving disadvantaged student populations. A fundamental reassessment of how we measure quality in education will be required to achieve the critical liberal purposes of No Child Left Behind and to avoid at least some of the unintended bureaucratic consequences that have plagued such initiatives in the past. NCLB can achieve these larger goals *if* we can get accountability right. My purpose in this book is to contribute to the exploration of *how* we can get accountability right.

#### How No Child Left Behind Is Supposed to Work

At 670 pages long, and with many of its key provisions modifying 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 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.”<sup>8</sup> It is not yet clear that the recent wave of

test-based accountability is closing this gap, as researchers have found the difference in test score outcomes “stagnating or rising in the 1990’s,”<sup>9</sup> a period of increased reliance on standards and accountability.

The second goal of No Child Left Behind 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.”<sup>10</sup> These accountability systems, however, are subservient to the primary objective of closing the achievement gap: “The purpose of this title is to ensure that all children have a fair, equal, and significant opportunity to obtain a high quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments.”<sup>11</sup>

Because the U.S. Constitution, unlike the state constitutions, contains no provision for or guarantee of an education, the federal government can directly order a state or its public schools only to do (or not do) very little.<sup>12</sup> Instead, the federal government uses money for leverage, much as it did when it tied transportation funds to setting a fifty-five mile-per-hour speed limit. Federal funding represents about 7–8 percent of total public school revenues—not determinative but hardly trivial. Any state that chooses to opt out of NCLB—and none have yet done so—stands to lose a significant amount of scarce money. It is not an easy bargain, however. By taking federal money, states commit themselves to implementing a massive and detailed testing and sanctions regime, resulting in annual assessments of most students. The results of these tests will have significant consequences for any school or district that does not meet the state’s standards of proficiency. As is the case with the tests themselves, states design many of the details about how these tests are used but must have their plans approved by the federal government.

No Child Left Behind is a reauthorization of the Elementary and Secondary Education Act of 1965, itself only one part of President Lyndon Johnson’s War on Poverty. That measure distributed federal funds to local school districts on the basis of how many low-income students they served. While states can (and under NCLB must) sanction poorly performing schools and districts in a variety of ways, the

withholding of Title I funds is the only direct consequence that the federal government can apply to education at the state and local levels under No Child Left Behind. While all public (and many private) schools in a state will participate in NCLB's testing regimen, not every state has chosen to apply sanctions to non-Title I schools.<sup>13</sup>

The heart of the testing and sanctions regime under NCLB is something called adequate yearly progress (AYP). Making AYP is the defining quest for every school and district under No Child Left Behind. If they make AYP one year, school officials can start worrying about not making it the next year. If a school fails to make AYP, especially for two or more years in a row, the consequences become increasingly severe. AYP is based on the results of students' scores on standardized tests administered once a year. Achieving AYP means either that a sufficiently high percentage of the students in a school or district meets the state's standards for academic proficiency or that the school or district is demonstrating "continuous and substantial academic improvement for all students."<sup>14</sup> If a sufficiently high percentage of students does not meet that year's proficiency targets, a school or district can still make AYP if the number of students not achieving the standards has declined by 10 percent since the previous year—the "safe harbor" option.

Test results under NCLB are looked at in aggregate for all of the students in a grade level and for eight subgroups of students: for five racial and ethnic identifiers (white, black, Hispanic, American Indian, and Asian or Pacific Islander)<sup>15</sup> and for students who are eligible for free or reduced-priced lunches, those with limited English proficiency, and those 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 apply to a specific subgroup of students only if that school has enough students in that subgroup to trigger AYP evaluation. States vary considerably in their standards for what qualifies as the minimum number of students within a given group in a given grade level to trigger AYP evaluation for that subgroup, though the number generally ranges from ten to fifty.<sup>16</sup> 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.<sup>17</sup>

These proficiency requirements are combined with test participation requirements to ensure that school principals do not encourage some students to stay home on test day. The subgroup rule also applies to the separate evaluation of a school's participation rate: 95 percent of all eligible students within each grade and subgroup must show up to take the test.<sup>18</sup> There are also attendance requirements and graduation requirements that depend on the grade levels offered within a school. All of these requirements mean that there are currently thirty-six "possible ways for an individual school or district to fail to make AYP" for each grade level in a given school—eighteen for reading (nine for proficiency and nine for participation) and eighteen for math, assuming that the school has enough students to qualify for each subgroup at that grade level.<sup>19</sup> There are thirty-seven ways if one includes attendance or graduation standards for the entire school. It only takes failure in one subgroup in one subject in one grade to trigger AYP identification for that school or district.

These test results are cross-sectional, meaning that they take a snapshot of students (in aggregate or within a subgroup) at one point in the year. NCLB, therefore, bases its sanctions not on measuring what a school is adding to the achievement of individual students but on the aggregate peer performance of students as a whole or within a group. Schools are measured and weighed by their students, not their services. Proficiency targets are set by the state, approved by the federal Department of Education, and derived from baseline proficiency levels obtained from tests given during the 2001–2 school year. The bar gets higher every year until the 2013–14 school year, when 100 percent of a school's and district's students must meet these standards.

Since the law was first implemented, states have gradually added grade levels to the group of students tested. By the 2005–6 school year, all students in grades three through eight will have been tested in reading and mathematics and all high schools will test each subject

once.<sup>20</sup> Science tests will be added by 2007–8, and President Bush has proposed expanding grade-level testing requirements in high schools. No Child Left Behind “applies the same high standards of academic achievement to all public elementary school and secondary school students,” including students with disabilities and those with limited English proficiency.<sup>21</sup> Schools may make “reasonable adaptations and accommodations for students with disabilities”;<sup>22</sup> however, the achievement of students with disabilities is still measured against the same state standards as the achievement of students without disabilities.<sup>23</sup> The law provides that any student who has attended school in the United States (except for Puerto Rico) for at least three consecutive years must be measured by the same proficiency tests, whether or not English is the student’s native language. Schools are free to include, on a case-by-case basis, students who have been in the United States for less than three years. Students who have not been in U.S. schools for three consecutive years may be assessed using native-language instruments for two more consecutive years, provided that the “student has not yet reached a level of English language proficiency sufficient to yield valid and reliable information on what such student knows and can do on tests.”<sup>24</sup>

Every year, each state educational agency must produce and distribute school report cards, which are required to provide concise and understandable information about student performance in each school (disaggregated by each of the eight subgroups that apply), graduation rates (if applicable), and the “professional qualifications” of the teachers disaggregated by the top and bottom quartiles of poverty in the state.<sup>25</sup> Parents of students in non-AYP schools can use these report cards to help them choose more successful schools when exercising their rights under the law.

For schools that fail to achieve their state’s proficiency and participation targets, as a whole or for any one group with enough students (within a given group at grade level) to hit the subgroup trigger, or that fail to meet attendance or graduation requirements, the consequences become progressively more severe (table 1). While no sanctions are associated with the first year of AYP failure (other than the public identification of that school’s status, usually in the local newspapers), being identified as a school that failed is a label that sticks,

thereby setting the stage for the much more consequential sanctions that follow.

If a school fails to make AYP for a second consecutive year in the same subject area and grade level, it is identified as a “school in need of improvement.” Those schools so identified that also receive Title I money must spend at least 10 percent of that money on professional development for teachers and principals but only on services that will improve academic achievement and remedy the deficiency that triggered the “in need of improvement” status. These schools must also notify their students’ parents of this failure and allow them to transfer their children to public and charter schools within the district that are making the grade, giving “priority to the lowest achieving children from low-income families.”<sup>26</sup> Districts cannot use lack of

**TABLE 1. Increasingly Severe Consequences for Schools That Fail to Make Adequate Yearly Progress**

Consecutive Years of AYP failure (1)	Consequences for Individual Schools That Receive Monies for Disadvantaged Children under Title I (2)
2 years	Identified as “in need of improvement” School officials must . . . Develop a school improvement plan Spend at least 10% of Title I funds on professional development Allow parents to transfer their children to successful schools in the district Notify parents of their options under this plan
3 years	All consequences from previous years <sup>a</sup> School officials must . . . Implement school improvement plan Provide supplemental educational services for students
4 years	Corrective action This may include replacing staff, overhauling the curriculum, reducing management authority at the school level, hiring outside experts, or lengthening the school day and/or year.
5 years	Plan for restructuring Either by reconstituting school as a charter school, replacing all or most of the school personnel, contracting out for private management, state intervention, or other restructuring efforts
6 years	Initiate restructuring

<sup>a</sup>Each subsequent year continues consequences from all previous years

space or programs as an excuse, and they must spend 20 percent of their Title I money on transportation and other related services.<sup>27</sup> Districts must notify in writing all parents of students in the failing schools what being identified as a failing school means, what the school is trying to do in response, and what transfer options exist under the law.<sup>28</sup>

Schools that fail to make AYP for a third consecutive year must continue to provide parental choice and dedicate part of their Title I funds to professional development. These schools must also, however, begin to offer supplemental tutoring, remedial, and other academic services to their students, with providers chosen from a list of those approved by the state. These approved providers may include “faith-based” nonprofit entities as well as for-profit private-sector providers.<sup>29</sup> Again, these services are funded with a school’s Title I money. A district receives no additional funds to comply with these requirements.

After a fourth year of failure, schools must take “corrective action” to meet the state’s proficiency and participation goals for all applicable groups of students.<sup>30</sup> These schools must replace the school personnel “who are relevant to the failure to make adequate yearly progress,” overhaul the school’s curriculum, “decrease management authority at the school level,” appoint “an outside expert to advise the school on its progress toward making adequate yearly progress,” lengthen the school day or year, or restructure the “internal organizational structure of the school.”<sup>31</sup>

If corrective action fails and the school finds itself on the failure list the next year, then the school must produce a plan for restructuring to be implemented in the next consecutive year of AYP failure. Restructuring is no small thing. It involves five basic options: reconstituting the failing school as a charter school, replacing “all or most of the school staff (which may include the principal) who are relevant to the failure to make adequate yearly progress,” signing “a contract with an entity, such as a private management company,” turning the operation of the school over to the state government, or any “other major restructuring of the school’s governance arrangement that makes fundamental reforms.”<sup>32</sup> Schools that have failed to make AYP in one or more years but make AYP the next year do not move

up the consequence ladder. However, they are only taken off the list of failing schools if they make AYP for two consecutive years. After entering the school improvement phase and making AYP for one year, a school that fails to make AYP the next year moves on to the next level of consequences.

The law includes many other provisions, including requirements for ensuring that all academic teachers are “highly qualified,” that all schools are safe and drug-free, consequences for entire school districts that fail to make AYP (based on the cumulative scores of all students in the district, a proviso that significantly improves the odds of targeting), along with technology-, reading-, and language-instruction initiatives. The law contains language to protect teachers, principals, and other educational professionals from “frivolous litigation” when trying to comply with the state’s safe schools policies.

Although I focus on No Child Left Behind’s effects on individual schools, entire school districts can also fail to make AYP based on the aggregate grade-level scores of all students in the district (also broken down by the eight subgroups). The most significant district-level consequence is the requirement to implement corrective action at the end of the fourth consecutive year of district AYP failure: districts can choose to defer program funds, implement curricular reforms, arrange for student transfer to another school district, overhaul the administration, or abolish the school district entirely.

#### How Is It Working?

Among the disagreements and controversies that surround the No Child Left Behind Act of 2001, researchers and policymakers agree on two things. First, the act’s goals are laudable, even necessary. Ensuring that our public schools demonstrate improved performance for all students and for those students who have traditionally underperformed—including students of minority ethnicity, those from low-income households, and those with disabilities or limited English proficiency—is absolutely essential on the grounds of fairness, national economic interest, and fulfillment of the American dream.<sup>33</sup> Second, the law’s effects are likely to be far-reaching. The

sanctions to be imposed on schools that fail to make AYP clearly are significant, since they include restructuring and reconstituting failing public schools, perhaps under private management.

Beyond these points of agreement, however, very little is known about how the law will play out, not only because we remain in the early stages of implementing NCLB but also because the problem is both undertheorized and lacking in enough solid, nonpartisan, empirical research. Though recent work has examined the politics surrounding the law's passage and how and why it came to be,<sup>34</sup> much less work, at least by political scientists, has focused on the political and policy consequences of the law's implementation.

The little empirical evidence that exists on the achievement effects of No Child Left Behind is mixed. A March 2005 national survey of state and district education officials found that 72 percent of school districts reported that academic achievement was improving on the state-designed tests.<sup>35</sup> School district personnel uniformly reported that they were "aligning curriculum and instruction with standards and assessment (99%), and providing extra or more intensive instruction to low-achieving students (99%)."<sup>36</sup> The authors of an analysis of 320,000 student test scores from twenty-three states also found some evidence of improvement of mathematics and reading scores between the 2001–2 and 2003–4 school years. However, they observed that recent year-by-year growth in student test scores had declined since NCLB was put into place.<sup>37</sup> More alarmingly, the authors found that although growth in achievement levels has been declining for all ethnic groups, Hispanic students appeared to be falling behind white non-Hispanic students.<sup>38</sup>

In contrast to the relative lack of findings on the achievement effects of No Child Left Behind, an already large and still growing literature documents what is wrong with the law and what incremental fixes might be applied to it. Attacks on and critiques of No Child Left Behind are proliferating and are coming from many different sources focusing on many different consequences, intended and unintended. These worries center on a variety of issues that can be grouped into seven categories: the use of standardized tests; the use of cross-sectional test score data; the requirement for 100 percent

proficiency; implementation, incentives, and uneven playing fields; unfunded mandates; lack of flexibility in implementation; and secondary provisions.

*The use of standardized tests.* One group of critics focuses on the challenges of measuring anything as complex as student achievement with any set of standardized tests—no matter how thoroughly or thoughtfully implemented—and the consequences of these tests on the curricula and learning environments of schools. The danger is that in their single-minded desire to improve test scores, schools and teachers will damage the breadth and quality of the curriculum. This issue has been a topic of discussion ever since these tests were first implemented, as Superintendent Crawford noted.<sup>39</sup> Decisions about what to include on these tests are themselves highly political<sup>40</sup> and often result in a watered-down consensus curriculum that fails to make any real cognitive or evaluative demands on the students.

Deborah Meier, a former teacher and principal in Boston and New York and one of the strongest critics of No Child Left Behind, calls the current push for high-stakes standardized testing “fundamentally misguided,” primarily because of the damage it will cause education as students experience it.<sup>41</sup> Rather than encourage the same kinds of qualities of leadership and governance that we would like to see in our students—the support and encouragement of local decision-making and citizen involvement—NCLB takes all power away from those closest to the creation of education and places it in the hands of distant “experts.” As Meier observes, the use of standardized tests to improve education is based on a list of assumptions, none of which has been established beyond doubt:

There is a definition of what we mean by a good education.  
 Experts can know what this definition is.  
 Experts can measure it.  
 We can trust the experts more than we can trust the teachers  
 and principals.  
 The system will produce equity in education, without major  
 resource shifts.  
 The system will work.<sup>42</sup>

To which I would add:

We will know whether all this is working.

*The use of cross-sectional test score data.* Even if one believes that standardized tests can accurately and benignly measure student achievement, extracting what the school is doing for that student's achievement is much trickier than simply asserting how well one group of students is doing. No Child Left Behind currently relies on what is often called a "status model" of educational achievement that bases its assessments on a one-time snapshot of all of the students in a grade and in the grade-level subgroups.<sup>43</sup>

Using status models to measure the quality of education within a school has three main drawbacks. The first is the likelihood that schools will be designated as failing even though they are making gains: schools starting from extremely low levels of average proficiency cannot possibly expect to hit the targets before identification and sanction kick in. Theoretically, the "safe harbor" provision should protect such schools, but they must still reduce the number of students failing by 10 percent every year. If only one subgroup realizes smaller gains than other groups or schools as a whole, a school can still fail.<sup>44</sup>

A further complication is the issue of where to set the cutoff for triggering the inclusion of scores from specific subgroups in assessing whether a school is making AYP. Setting the trigger low (for example, ten students minimum per grade in any of the eight subgroups) increases the number of schools that qualify and, therefore, fail. Setting the trigger higher (fifty students in any given subgroup) runs the risk of ignoring the needs of minority students, since their results may not be directly looked at in schools with too few minority or high-need students to trigger the AYP evaluation for that subgroup. If the needs of specific subgroups of students are ignored, the basic premise of No Child Left Behind is lost.

The most important risk, however, is that basing sanctions on a status model also runs the considerable risk of identifying and sanctioning schools based mostly on the characteristics of the students rather than on the school's contribution to students' academic

achievement, since many things outside of a school's control show up in the cross-sectional aggregation of student test scores.<sup>45</sup> I will take up this issue in considerable detail later in this volume. For now, however, it is useful to note that determining a school's success or failure based on these kinds of cross-sectional tests appears to result in incorrectly singling out schools with high-need populations.<sup>46</sup>

*One hundred percent proficiency.* Closely related are potential problems resulting from the law's insistence on full proficiency as of 2013–14, when 100 percent of each school's and each district's students must be proficient in reading and mathematics. Critics have argued that this goal at best is overly ambitious and at worst is completely unattainable, leading to the prospect of failure for nearly all of the nation's public schools: "In 2003, no state or large district had anything close to 100% of their students performing at the *basic* level, much less the *proficient* level."<sup>47</sup> To attain 100 percent proficiency by 2013–14 would require a rate of improvement many times that which has been observed in recent years in even the most promising schools.<sup>48</sup>

The prospect of inevitable and widespread failure has led some observers and researchers to assert that the real goal of No Child Left Behind is the destruction of the public schools.<sup>49</sup> The challenge of 100 percent proficiency is especially problematic when one considers the prospect of full proficiency for special education students. Researchers have also pointed out that many special education students are so labeled because their disabilities prevent them from attaining grade-level proficiency, producing an inherent contradiction between definitions and goals.

*Implementation, incentives, and uneven playing fields.* All policies have the potential for unforeseen and unwanted consequences, and No Child Left Behind is no different.<sup>50</sup> Critics have focused mostly on the skewed incentive structures that state, district, and school officials will face—that is, states face powerful incentives to set the proficiency bar as low as possible, given the costs involved in instituting the sanctions regime.<sup>51</sup> Teachers have incentives to cheat, while their leaders have incentives to look the other way if they feel that they will not be

caught.<sup>52</sup> Districts have few incentives to accept poor and minority students from other districts (unless they are forced to do so), and schools might be encouraged to move their lower-performing students out of regular education and into alternative learning centers or high school equivalency programs.<sup>53</sup> Given the malleability of and tremendous variance in how states define academic proficiency, many potential problems arise with the utility and reliability of data. A report by the U.S. General Accounting Office cautioned that “more than half of the state and school district officials . . . interviewed reported by being hampered by poor and unreliable student data.”<sup>54</sup>

Although the federal Department of Education must approve states’ plans, the standards and assessments that states use are typically based on preexisting accountability systems, which may or may not have overlapped with NCLB’s goals and methods, creating a very messy patchwork of tests. This tremendous state variation in baseline levels of proficiency means that a successful school in one state could easily be classified as failing in another state that set its initial proficiency targets at a higher level.<sup>55</sup> The General Accounting Office also found a great deal of variance in how and how well states were measuring academic proficiency in these early stages of implementation, leading, for example, California to require that only 14 percent of its elementary students be proficient and Colorado to require 78 percent proficiency in the same year.<sup>56</sup> According to Minnesota’s Office of the Legislative Auditor, “Due to inter-state differences in proficiency standards, testing practices, and ‘adequate yearly progress’ calculations, there is no meaningful way to use AYP data to make multi-state comparisons of educational performance.”<sup>57</sup>

Other implementation worries focus not on the failure of the law, its goals, or its methods but on the school, district, and state officials who are supposed to be carrying out NCLB’s reforms. Of particular concern are levels of compliance and the pace of implementation of parental exit options in failing schools (which apply to schools that have failed to make AYP for two consecutive years). Very small percentages of parents who are eligible to send their children to successful schools appear to be doing so, and only a slightly larger percentage of parents whose children are eligible for supplemental

services take advantage of that option. Lack of capacity in successful schools, lack of information on the part of parents, and active district resistance may have combined to prevent large numbers of parents from exercising their exit options.<sup>58</sup>

*Unfunded mandates.* Claims that No Child Left Behind does not adequately fund the ambitious goals and requirements that it places on states and districts have come from all levels of the public educational system. At the district level, such costs include the requirement to comply with the fully trained teacher requirement, the provision of supplemental services, and the task of administering and grading the ever-increasing volume of tests. At the state level, these costs include implementing all of the corrective action and restructuring provisions as well as overseeing the testing regimen.<sup>59</sup> Ted Kennedy, who helped steer the bill through the Senate, later lamented, “The tragedy is that these long overdue reforms are finally in place, but the funds are not.”<sup>60</sup>

In April 2005, school districts in several states joined the National Education Association in suing Secretary of Education Margaret Spellings on the basis of insufficient funding of NCLB.<sup>61</sup> Though the law states, “Nothing in this Act shall be construed to . . . mandate a State or any subdivision thereof to spend any funds or incur any costs not paid for under this Act,”<sup>62</sup> the plaintiffs argued that the compliance requirements placed an unfunded and undue financial burden on local schools and districts. The U.S. General Accounting Office concluded that No Child Left Behind was not an unfunded mandate because participation in NCLB is technically voluntary and the conditions laid out in the law “were a condition of federal financial assistance,” thus disqualifying NCLB as an unfunded mandate as defined by law.<sup>63</sup> A February 2005 report by the National Council of State Legislatures, argued that No Child Left Behind’s ambiguous and coercive provisions fail to pass constitutional muster.<sup>64</sup> In addition, observers have expressed concern that rigidities in the provisions for schools that fail to make AYP will make academic improvement less rather than more likely as insufficient monies are spread too thinly and are used only in crisis mode.

*Lack of flexibility in implementation.* Reports from the field also indicate a great deal of frustration on the part of school, district, and state officials regarding their lack of flexibility in complying with the law's underlying goals. Of particular concern are rigid and unrealistic targets of academic success, particularly for students receiving special education services and those with limited English proficiency. In its 2005 task force report, the National Conference of State Legislatures concluded that "states should be allowed to develop any system they choose as long as it meets the spirit of NCLB."<sup>65</sup> The report also argued that a preexisting statutory basis for this much-needed flexibility lies in provisions allowing the U.S. Department of Education to offer statutory waivers to states, districts, and schools.<sup>66</sup>

*Secondary provisions.* Small but significant provisions continue to come to light, such as the withholding of all federal funds for school districts that fail to notify their state's educational administration that "no policy of the local educational agency prevents, or otherwise denies participation in, constitutionally protected prayer in public elementary schools and secondary schools."<sup>67</sup> Funds are also to be withheld from districts that "deny equal access [to] any group affiliated with the Boy Scouts of America" or any other group listed as a "patriotic society" as well as from districts that fail to "provide, on a request made by military recruiters or an institution of higher education access to secondary school students' names, addresses, and telephone listings."<sup>68</sup>

In response to these concerns, researchers and policymakers have proposed a series of modifications to No Child Left Behind.

1. Fully funding the law

No Child Left Behind's passage was accompanied by promises of significant increases in funding. Officials in several states are now complaining that the additional monies allocated by the federal government do not even cover the costs of compliance, much less those necessary to achieve meaningful progress in closing the achievement

gap.<sup>69</sup> The combined concerns of unfunded mandates and the increased federal intrusion into state policy-making have sparked legal and political opposition among a very diverse group of stakeholders that runs the gamut from teachers' unions to Republican governors and legislators.

On April 20, 2005, the Utah legislature voted to instruct state school leaders to focus on Utah's goals and priorities and to ignore provisions of NCLB that conflicted with the state's plan; Utah's governor signed the measure the following month. The bill did not directly withdraw Utah from NCLB but increased the likelihood of a showdown over the state's autonomy and put \$76 million in federal education funding at risk.<sup>70</sup> Republican U.S. Representative Steve Mascaro was quite blunt: "I'd just as soon they take the stinking money and go back to Washington with it. Let us resolve our education problems by ourselves. I will not be threatened by Washington over \$76 million."<sup>71</sup> The following August, Connecticut sued the federal government to reclaim \$50 million in state money spent on compliance. Though the amount was small, observers cited the symbolic importance of the suit, arguing that the "real battle is over who will make key education decisions—local officials or policymakers in Washington."<sup>72</sup> These political rifts are likely only to deepen as NCLB's sanctions sweep up more and more schools, thereby forcing states to cope with the attendant costs.

## 2. Moving away from status models of measurement

Many reform proposals focus on the way in which achievement is measured. One proposed change is to look at the year-by-year changes in percentages of students who achieve academic proficiency rather than the percentages of students who are or are not meeting the targets. This approach is usually referred to as a "growth model."<sup>73</sup> A second approach is to use what is called a "value-added" model of achievement, which attempts to extract the value a school adds to a given student's achievement by tracking the changes in test scores over time for individual students.<sup>74</sup> I will discuss both of these modifications in some detail later in this volume. Neither is per-

fect, and each raises its own challenges, but either might be preferable to the status model currently embodied in the law.

### 3. Providing more flexibility and more realism

Some observers have argued that the requirement for 100 percent proficiency should be amended or scaled back to account for the almost impossible nature of its attainment: “there should be evidence that the goal does not exceed that [which] has previously been achieved by the highest performing schools.”<sup>75</sup> Others have argued for a more realistic and flexible approach to the testing of and achievement standards for students with limited English proficiency and those with disabilities.<sup>76</sup> Critics have also declared that the 10 percent growth requirement necessary to achieve safe harbor cannot be attained by all but a very small number of schools not making AYP.<sup>77</sup>

### 4. Increasing coherence and consistency between states, particularly in the definition of academic proficiency

More standardization of accountability systems would not only be fairer but also provide researchers with much more reliable data on which to base any evaluations of NCLB’s effects on achievement.<sup>78</sup>

## What Is Still Missing from the Debate

Of all the problems with No Child Left Behind, the ultimate success or failure of its ambitious agenda will likely come down to how the law interacts with two unavoidable realities in American education.

The first reality is that the United States has a very unequal and class-stratified society.<sup>79</sup> Most children go to school with children from families with similar incomes. However ambitious, No Child Left Behind does nothing to address these long-standing resource inequalities in the United States and its educational system. To the extent that these resource inequalities—rather than failures by administrators, principals, and/or teachers—are responsible for

unequal academic achievement, No Child Left Behind is destined to fail to live up to its liberal promises.

The second reality is that education involves a great deal of uncertainty on the part of those who would evaluate it and a great deal of discretion on the part of those who perform its tasks.<sup>80</sup> In other words, it is very hard to know what is actually happening at the moment of “education.” In all of the debate about No Child Left Behind, policymakers have failed to ask the most basic and most important question: Can we ever really know if a child’s education is good?

So many factors contribute to and confound what ultimately happens at the interface between a student’s mind and the school’s products that policymakers should think very carefully about how to measure “educational quality.” Doing so will require a deeper and more thorough investigation of the underlying problem of producing and measuring quality in education. It may seem like an obscure task, but, as I hope will become apparent, it represents the first step on an interesting and necessary journey to achieve a vision of equality of educational outcomes for all students, especially for those students whose dreams have been continually deferred.

This book argues primarily that policymakers must begin with the goal of producing quality in education and move from there rather than assume that any test can accurately measure “good” education. This approach is radical in the literal sense of getting to the roots of the challenge at hand. We cannot succeed without doing so. If—and this is a big if—we accept the limitations of our evaluative capabilities, and if we get beyond the magical thinking of NCLB’s punishment-driven philosophy, then we can make the law live up to its liberal promises.

Chapter 2, “The Problem of Quality,” lays out the foundational arguments of the book with an exploration of two questions:

What determines educational quality?

How, if at all, can we observe and identify it?

To answer these questions, I explore the ephemeral nature of assessing quality in education. I take a detour into what economists and

scholars of organizations can tell us about designing systems of measurements and rewards to lay the groundwork for a more effective way of going about things under No Child Left Behind.

Chapter 3, “Making the Grade (or Not): Success and Failure in NCLB’s World,” begins the book’s empirical analyses by taking a broad look at how schools fare under the law and how success and failure in No Child Left Behind’s peculiar world are determined by phenomena over which schools have absolutely no control. I combine a discussion of national developments with original data from a survey of Minnesota school principals as a way to explore the complex reality of school success and failure:

Why do schools succeed or fail to make adequate yearly progress?

Given the dominance of race, ethnicity, and inequality in determining achievement test scores, do school principals’ actions matter at all in measuring educational achievement?

Chapter 4, “Top-Down and Bottom-Up: NCLB, Charter Schools, and the Public School Principalship,” examines two sets of questions about the connection between NCLB and the behaviors and attitudes of the principals who are ultimately responsible for what happens inside the school. The first set of questions concerns the effects of No Child Left Behind on what principals do, how they spend their time, and what they think about their influence:

What do public school principals think about No Child Left Behind?

What effects does the law have on leadership within the public schools?

The second set of questions concerns the critical but mostly overlooked issue of how No Child Left Behind interacts with the other major reform effort already under way in the United States, which is the introduction of competition to public schooling, whether in the form of charter schools, public school choice, or vouchers for private

schools. I focus on the connection between the provision of parental choices in education and NCLB's top-down accountability approach, asking:

How does doing top-down accountability and bottom-up school choice at the same time affect the quality of the public school principalship? Do the two approaches reinforce each other, or do they get in each other's way? Is sanction better than competition?

Charter schools are in a unique position as both targets and solutions under No Child Left Behind. They are subject to the same tests and sanctions as their traditional public school counterparts; however, they also constitute one of the options for reconstituting a failing school in the later stages of NCLB implementation. In spite of charters' involvement with both top-down and bottom-up reforms, their placement as such has not received enough attention in the literature.

Chapter 5, "Rethinking Assessment," takes up only one question, but it is an important one:

If we wish to preserve the goals and achieve the purported benefits of NCLB, can we do a better job?

I examine two commonly discussed alternatives to the cross-sectional testing: comparing schools on the year-by-year changes in percentages of students making proficiency targets and comparing schools based on changes in individual student test scores over time. I also suggest a new alternative: using, in a meaningful way, the opinions of those closest to the production of the good and the bad in education in our assessments of schools.

In chapter 6, "Carrots, Sticks, and Unbroken Windows: Making NCLB Live Up to Its Promises," I propose that we take these lessons about measuring educational quality as a basis for trying to make No Child Left Behind fulfill its goals. I offer a series of increasingly substantial modifications to or revisions of No Child Left Behind that will get us closer to realizing the critical vision of academic equality

embodied in the law. We should directly measure leadership and quality within the institutions rather than just trying to extract this information from student test scores. We should also include meaningful rewards in the incentive structure that we create based on these assessments.

No Child Left Behind has received considerable attention, an unsurprising development given both the ambitious goals embodied in the law and the significant effects that its implementation has had and will continue to have. Because we remain in the early stages of the law's implementation, relatively little peer-reviewed scholarly work has been published on the topic.<sup>81</sup> This book seeks in part to help remedy this deficiency.

This book's value ultimately will be determined by practitioners and stakeholders in all levels of U.S. education, from parents to policymakers. I hope that the book will be relevant beyond its immediate subject. Many states, including those with Republican governors and legislatures, are growing increasingly uncomfortable with the federal government's reach into what has traditionally been the purview of state and local governments. NCLB will come up for reauthorization in 2007. It could be gutted, revoked, or starved, the preferred way of killing a policy that benefits populations that are seen as deserving. Regardless of what happens with No Child Left Behind, however, high-stakes accountability will remain a central part of the education reform debate for the foreseeable future.

Three conditions are likely to endure in U.S. educational policy: a desire to improve academic achievement among traditionally underperforming students, the struggle with the problem of how to do so, and the related challenge of assessing whether what we are doing works. This book seeks primarily to help create a framework that will allow us to respond to these challenges in closing the achievement gap.

Perhaps a course for policymakers—something along the lines of *Measuring and Producing Educational Quality 101*—would be useful. If I were to offer such a course, I would begin with the lessons learned and the lessons not learned since Superintendent King's unfortunate experiment. I would take my students down the long and winding road of understanding the importance of resource

inequalities in education; the challenge of measuring what is good and what is not good; and the importance of leadership, organization, and incentives. In attempting to find and support this ephemeral thing called a good education, I would end with the hopeful caution that we not leave our common sense behind while we try to move these mountains.