

THE PERVERSE INCENTIVES OF THE NO CHILD LEFT BEHIND ACT

JAMES E. RYAN*

This Article examines the No Child Left Behind Act, which may be the most important federal education law in our nation's history. The Act is supposed to increase academic achievement in schools across the nation, raise the performance of disadvantaged students to the level of their more affluent counterparts, and attract qualified professionals to teach in every classroom. These goals are obviously laudable. As Professor Ryan explains, however, the Act creates incentives that actually work against their achievement. Specifically, the Act unintentionally encourages states to lower their academic standards, promotes school segregation and the pushing out of poor and minority students, and discourages good teachers from taking jobs in challenging classrooms. Should any or all of these effects occur, achieving the Act's goals will be more difficult, not less. Professor Ryan goes on to suggest a solution, albeit a partial one, to the problems created by the No Child Left Behind Act. Rather than focus on absolute achievement levels as the basis for school accountability, Ryan argues that the federal government and states should focus on rates of growth. Doing so would not only give a more accurate picture of school quality, and thus provide a fairer basis for school accountability; it would also diminish or eliminate the perverse incentives created by the No Child Left Behind Act. The Article concludes with a brief discussion of what the No Child Left Behind Act can teach us about the proper role of the federal government in education law and policy.

INTRODUCTION

The No Child Left Behind Act, perhaps the most important federal education law in our nation's history, is at war with itself. The chief goals of the Act are to boost academic achievement across the board and to eliminate the achievement gap among students from different backgrounds.¹ To accomplish these goals, the Act requires

* Copyright © 2004 by James E. Ryan, William L. Matheson and Robert M. Morgenthau Distinguished Professor, University of Virginia School of Law. Thanks to Dale Ballou, Jack Boger, Charlie Clotfelter, Anne Coughlin, Michael Heise, Rick Hess, John Jeffries, Mike Klarman, Daryl Levinson, Robert L. Linn, Liz Magill, Tom Nachbar, Tom Saunders, Elizabeth Scott, Scott Shapiro, Paul Stephan, Bill Stuntz, Lauren Wetzler, and the participants in workshops at George Washington University and the University of Virginia Law Schools for helpful discussions and comments on earlier drafts. Thanks as well to Monika Moore, Tim Doherty, and the librarians at Virginia and Yale Law Schools for excellent research assistance.

¹ See No Child Left Behind Act (NCLBA) of 2001 § 1001, 20 U.S.C. § 6301 (Statement of Purpose); see also CTR. ON EDUC. POL'Y, FROM THE CAPITAL TO THE CLASSROOM: STATE AND FEDERAL EFFORTS TO IMPLEMENT THE NO CHILD LEFT BEHIND ACT, at iii (2003) [hereinafter CTR. ON EDUC. POL'Y] (describing purpose of NCLBA), available at http://www.ctredpol.org/pubs/nclb_full_report_jan2003/nclb_full_report_jan2003.pdf.

states to establish “challenging” academic standards for all schools and to test all students regularly to ensure that they are meeting those standards.² The Act also requires states and school districts to employ teachers who are “highly qualified,” meaning that they have demonstrated some competence in the subjects they teach.³

Schools are expected to have all of their students scoring at the proficient level on state tests within twelve years of the Act’s passage. In the meantime, states must establish intermediate goals that require an ever-increasing percentage of students to demonstrate proficiency. The same intermediate achievement targets must be met both by schools as a whole and by various subgroups of students within each school, including those who are poor, racial and ethnic minorities, English-language learners, and those entitled to special education services.⁴ Schools that receive federal funding and fail to meet their targets face increasingly harsh sanctions for every year that they fail.⁵

The No Child Left Behind Act has been praised by some and condemned by others in the popular press and in education journals, although it has received surprisingly little attention in the legal literature.⁶ Those who favor the Act emphasize its laudable goals and celebrate its tough accountability measures.⁷ Those who criticize the Act lament the heavy emphasis on testing and the inevitable “teaching to the test” that will follow.⁸ They also chastise the federal government for interfering with state and local control over education while failing to fund all of the costs associated with the Act.⁹

² § 1111.

³ § 1119.

⁴ § 1111(b)(2)(C)(v).

⁵ § 1116.

⁶ Important exceptions include James S. Liebman & Charles F. Sabel, *The Federal No Child Left Behind Act and the Post-Desegregation Civil Rights Agenda*, 81 N.C. L. REV. 1703 (2003), and John Charles Boger, *Education’s “Perfect Storm”? Racial Resegregation, High Stakes Testing, and School Resource Inequities: The Case of North Carolina*, 81 N.C. L. REV. 1375, 1434–45 (2003).

⁷ See, e.g., Liebman & Sabel, *supra* note 6, at 1708–21 (arguing that Act can help spur race to top in educational achievement and provide “new and effective tools for significantly improving the educational outcomes of poor and minority children”); Michael Casserly, Commentary: *Can the Bush School Plan Work?*, EDUC. WK., Dec. 4, 2002, at 48 (arguing that law “could work if schools view it, as most major city school systems now do, as an opportunity to focus anew on student performance”).

⁸ See, e.g., Richard F. Elmore, *Testing Trap*, HARVARD MAG., Sept.–Oct. 2002, at 35, 97.

⁹ See, e.g., David J. Hoff, *Debate Grows on True Costs of School Law*, EDUC. WK., Feb. 4, 2004, at 1 (describing protests among state legislatures regarding costs of NCLBA); William J. Mathis, *No Child Left Behind: Costs and Benefits*, 84 PHI DELTA KAPPAN 679, 685 (2003) (arguing that federal government is asking for too much and giving too little in resources). Legislators in several states have considered rejecting federal funding in order to escape the requirements of the NCLBA. See, e.g., Pamela M. Prah, *Utah Considers*

My assessment is somewhat different. I agree that the Act's goals are laudable. In my view, however, the Act's fatal flaw is that it creates incentives that work against the Act's goals.

First, while the Act is supposed to raise achievement across all schools, it creates incentives for states to lower academic standards. Second, while the Act is supposed to close the achievement gap, it creates incentives to increase segregation by class and race and to push low-performing students out of school entirely, which will make it even more difficult for disadvantaged students to catch up to their more affluent peers. Finally, while the Act is supposed to bring talented teachers to every classroom, it may deter some from teaching altogether and divert others away from the most challenging classrooms, where they are needed the most. In short, although the Act is supposed to promote excellence and equity, it may work against both.

This Article explains how the Act creates these perverse incentives, and it offers a partial solution. The central problem is the Act's central feature—namely, the requirement that schools be sanctioned if their students fail to demonstrate an absolute level of achievement on tests within a relatively short time period. The requirement that an increasing percentage of students in every school achieve a certain test score each year is arbitrary and unrealistic, in that it establishes achievement goals without any reference to past achievement levels or rates of achievement growth. Many schools, including some that are considered effective, will be unable to meet these achievement targets. This will create pressure to make the targets easier to meet by dumbing down the tests or making scoring systems more generous. By this process, a law intended to raise academic standards may lower them.

Focusing on absolute achievement levels rather than achievement gains also will generate incentives for parents, teachers, and administrators to shun disadvantaged children and the schools that educate them. The reason is fairly simple: Disadvantaged students tend to do worse on standardized tests than do their more affluent counterparts. An accountability system that rewards and punishes schools based on absolute achievement levels will thus reward relatively affluent schools and punish relatively poor ones. Moreover, given that minorities are disproportionately poor, and that all schools are held responsible for the performance of their minority and poor students, this accountability system will tend to punish those schools that are racially and economically diverse. All of this will make racial and

Opting Out of No Child Left Behind, STATELINE.ORG, Dec. 30, 2003, at <http://www.stateline.org/stateline/?pa=story&sa=showStoryInfo&id=341320>.

socioeconomic integration even more difficult to achieve than it is already, and it will provide even more incentives for good teachers to choose relatively affluent schools. These trends, in turn, make it possible that a law designed to narrow the achievement gap will help widen it.

Given that the main problem with the Act is its focus on absolute achievement levels, a partial solution is to focus instead on rates of growth. As explained below, this would provide a more accurate measure of the value that schools add to a student's knowledge and abilities, and thus a more accurate picture of school quality. Focusing on rates of growth would also level the playing field among schools with different student populations, making it less likely that judgments about school quality will simply track the socioeconomic status of the students who attend them. This, in turn, would reduce the incentives of good teachers to avoid disadvantaged schools.

Looking to growth rates is not an untested idea. On the contrary, it formed the basis of so-called "value-added" accountability systems, which were increasingly popular prior to the passage of the No Child Left Behind Act. While focusing on rates of growth is thus not a radical and untried proposal, it is also not a panacea. Rewarding schools for meeting growth targets that are realistic may not provide sufficient incentives for schools to do any more than improve marginally on the status quo. Focusing on growth, moreover, necessarily tolerates different levels of absolute achievement for different students. A school whose fifth graders begin the year reading at the third-grade level and end reading at the fourth-grade level certainly has added value and done a decent job; but those students would still be a year behind in reading.

Despite these shortcomings, a value-added approach seems clearly superior to the approach embodied in the No Child Left Behind Act. The question then becomes whether it is politically plausible. At first glance, one might be skeptical. Considerable tension exists between a value-added accountability system and the core rhetorical commitment of the No Child Left Behind Act, which is to bring *all* students to the same level of achievement in a relatively short period of time. Nevertheless, the Department of Education has given states some flexibility in implementing the Act, offering them the opportunity to incorporate some focus on growth rates in their accountability systems, at least in the short run. States should take advantage of this flexibility. All states, moreover, should recognize that it is highly unlikely that the No Child Left Behind Act will remain in force, unchanged, for the next ten years. States thus should concen-

trate on short-term attempts to combat the perverse incentives created by the Act.

This Article proceeds in three Parts. Part I places the No Child Left Behind Act in historical context and describes its central features. Part II describes how the Act may push states to lower their academic standards and goals, increase pressure on schools to segregate students by class and race and exclude low-performing students altogether, and deter talented teachers from entering the profession and taking jobs in challenging schools. Part III explores the benefits and limitations of using rates of growth rather than absolute achievement levels as a basis for school accountability. While recognizing the inherent problems of this alternative, I argue that it is preferable to the approach of the No Child Left Behind Act and that states should do what they can to incorporate elements of such an approach into their accountability systems. I conclude with some tentative observations as to what the No Child Left Behind Act reveals about the proper role of the federal government in education law and policy.

One final introductory note is in order. The topic of standards and testing is a very large one, which could be approached from a number of angles. This Article concentrates on the mechanics and incentives of test-based accountability systems, focusing primarily on the regime established by the No Child Left Behind Act. I leave to one side some basic questions about the wisdom of standards and testing generally. There are serious disagreements as to whether relying solely or primarily on tests to hold students and schools accountable is wise policy, especially in light of the generally poor quality of current tests and the pervasive inequality of school funding. Those arguments already have been rehearsed thoroughly by others.¹⁰ What has received less attention is the issue that I address here—namely, the incentives created by the implementation of a test-based accountability system. Given that testing is ubiquitous and likely to continue for some time, even if the No Child Left Behind Act is modified or repealed, understanding the incentives created by test-based accountability is crucial.¹¹

¹⁰ See, e.g., ALFIE KOHN, *THE CASE AGAINST STANDARDIZED TESTING: RAISING THE SCORES, RUINING THE SCHOOLS* (2000) (describing ways in which standardized testing interferes with quality education); SUSAN OHANIAN, *ONE SIZE FITS FEW: THE FOLLY OF EDUCATIONAL STANDARDS* (1999) (same); PETER SACKS, *STANDARDIZED MINDS: THE HIGH PRICE OF AMERICA'S TESTING CULTURE AND WHAT WE CAN DO TO CHANGE IT* (1999) (same).

¹¹ One sign of how entrenched testing has become can be found in two recent editorials, one in the *New York Times* and the other in the *Washington Post*. The former discusses recent problems with testing in New York, which required results of a statewide math test to be discarded. Rather than use this as an opportunity to criticize testing gener-

I

THE NO CHILD LEFT BEHIND ACT

A. *Background*

Supported by an overwhelming majority in Congress and signed into law by President Bush in 2002, the No Child Left Behind Act (NCLBA) is remarkably ambitious and unusually intrusive.¹² The NCLBA revises the Elementary and Secondary Education Act,¹³ which was first enacted in 1965 and has been reauthorized periodically ever since.¹⁴ The most important and well-known component of the Elementary and Secondary Education Act is Title I, which is the federal government's single largest educational aid program and ostensibly is designed to assist disadvantaged students. In exchange for federal funding, which all states receive, states and local school districts must comply with various federal directives.

From its passage until fairly recently, Title I received more criticism than praise. Empirical studies generally concluded that Title I fell far short of its goal of closing the achievement gap between poorer and more affluent students.¹⁵ One problem was the way federal money was used. Title I funding mostly supported the hiring of teachers' aides and the creation of remedial classes for disadvantaged students, who typically were pulled out of regular classrooms and

ally, however, the *Times* instead argued essentially that testing should be improved, not discontinued. Editorial, *Why Testing Can't Fail*, N.Y. TIMES, June 30, 2003, at A20. The editorial in the *Post* chastised states for failing to follow through with promises to withhold diplomas from students who do not pass the tests. Editorial, *High-Stakes Games*, WASH. POST, Aug. 1, 2003, at A18. These editorials obviously are not reliable proxies for how "America" thinks, but the fact that two politically liberal editorial pages endorsed testing is nonetheless a telling sign of its potential staying power.

¹² See, e.g., Sam Dillon, *Thousands of Schools May Run Afoul of New Law*, N.Y. TIMES, Feb. 16, 2003, § 1 (National), at 33 (quoting Paul Houston, Executive Director of American Association of School Administrators, who called law "the largest federal intrusion into the educational affairs of the states in the history of this country"); Elmore, *supra* note 8, at 35 (calling NCLBA "the single largest—and the single most damaging—expansion of federal power over the nation's education system").

¹³ NCLBA § 101.

¹⁴ Prior to the NCLBA, the most recent reauthorization of the Elementary and Secondary Education Act occurred in 1994. See *Improving America's Schools Act (IASA) of 1994*, Pub. L. No. 103-382, 108 Stat. 3518.

¹⁵ See, e.g., MICHAEL J. PUMA ET AL., U.S. DEP'T OF EDUC., PROSPECTS: FINAL REPORT ON STUDENT OUTCOMES 9 (Abt Assoc., Inc. ed., 1997) ("Chapter 1 assistance was, on average, insufficient to close the gap in academic achievement between advantaged and disadvantaged students."); Gary Natriello & Edward L. McDill, *Title I: From Funding Mechanism to Educational Program*, in *HARD WORK FOR GOOD SCHOOLS: FACTS NOT FADS IN TITLE I REFORM* 31, 33-34 (Harvard University 1999) (detailing impact of Title I spending on academic achievement), available at <http://www.columbia.edu/~gjn6/id001.html>.

exposed to a watered-down curriculum.¹⁶ Not surprisingly, this strategy did little to bridge the achievement gap.

By the time Title I was scheduled for reauthorization in 1994, many in and outside of the federal government agreed that the program needed alteration. Congress and President Clinton turned to standards-based reform for inspiration and direction.¹⁷ Standards-based reform centers on the simple idea that states should set ambitious academic standards and periodically assess students to gauge their progress toward meeting those standards.¹⁸ The reform traces back to the 1983 publication of *A Nation At Risk*, a highly critical and widely publicized report on public schools, which argued in dramatic terms that America's schools set their sights too low.¹⁹ Standards-based reform promised to raise the academic bar by requiring all schools within a state to meet uniform, challenging standards. In addition to promoting excellence, standards-based reform also promised to promote equity by requiring all students, not just those in privileged suburban schools, to meet the same rigorous standards.²⁰

In reauthorizing Title I in 1994 through the passage of the Improving America's Schools Act (IASA),²¹ Congress and President Clinton incorporated the core ideas of standards-based reform.²² In doing so, they fundamentally changed the nature of Title I. Instead of providing funds to support remedial instruction for disadvantaged students, Title I funds now had to be used to create standards for all

¹⁶ Liebman & Sabel, *supra* note 6, at 1721 n.59; Paul Weckstein, *School Reform and Enforceable Rights to Quality Education*, in *LAW & SCHOOL REFORM: SIX STRATEGIES FOR PROMOTING EDUCATIONAL EQUITY* 306, 328–29 (Jay P. Heubert ed., 1999).

¹⁷ See Elmore, *supra* note 8, at 36 (describing history of 1994 Title I reauthorization).

¹⁸ For helpful discussions of standards-based reform, see DIANE RAVITCH, *NATIONAL STANDARDS IN AMERICAN EDUCATION: A CITIZEN'S GUIDE* (1995); David K. Cohen, *Standards-Based School Reform: Policy, Practice, and Performance*, in *HOLDING SCHOOLS ACCOUNTABLE: PERFORMANCE-BASED REFORM IN EDUCATION* 99, 99–127 (Helen F. Ladd ed., 1996) [hereinafter *HOLDING SCHOOLS ACCOUNTABLE*]; Lawrence B. Joseph, *Introduction to EDUCATION POLICY FOR THE 21ST CENTURY: CHALLENGES AND OPPORTUNITIES IN STANDARDS-BASED REFORM* 1, 1–21 (Lawrence B. Joseph ed., 2001) [hereinafter *EDUCATION POLICY FOR THE 21ST CENTURY*]; and Diane Massell, *Standards-Based Reform in the States: Progress and Challenges*, in *EDUCATION POLICY FOR THE 21ST CENTURY*, *supra*, at 135, 135–68.

¹⁹ See NAT'L COMM'N ON EXCELLENCE IN EDUC., U.S. DEP'T OF EDUC., *A NATION AT RISK* 5 (1983) (arguing that America's educational foundations were being eroded by "rising tide of mediocrity that threatens our very future as a Nation and a people"). On the connection between *A Nation At Risk* and standards-based reform, see, for example, RAVITCH, *supra* note 18, at 52–56, and Massell, *supra* note 18, at 136.

²⁰ William L. Taylor, *Assessment as a Means to a Quality Education*, 8 *GEO. J. ON POVERTY L. & POL'Y* 311, 312–13 (2001) (arguing that standards-based reform is promising strategy to equalize educational opportunities).

²¹ IASA § 6311(a)(1).

²² Elmore, *supra* note 8, at 36.

students.²³ In order to receive Title I funds, states had to create “challenging” content and performance standards in at least reading and math, develop assessments that were aligned with those standards, and formulate plans to assist and ultimately sanction failing schools.²⁴ Importantly, standards and assessments for Title I schools had to be the same as those established for all other schools within a state.²⁵ In this way, the federal government hoped to ensure that states would hold all students to the same high expectations and hold all schools, regardless of their student population, accountable for failure.

B. Key Provisions

The No Child Left Behind Act follows the same basic approach as the IASA, but it establishes more ambitious goals and places greater constraints on the states. States must still develop “challenging” content and performance standards, now not only in reading and math, but also in science.²⁶ States must still use assessments that are aligned with those standards, and must hold schools and school districts accountable for failing to meet ambitious achievement goals.²⁷

The most significant changes have to do with teachers, testing, and accountability. As for teachers, the NCLBA requires that Title I schools hire only “highly qualified” teachers for all subjects²⁸ and that veteran teachers in such schools demonstrate that they are “highly qualified” by 2005–6.²⁹ The Act also reaches beyond Title I schools and requires that all teachers of “core academic subjects”³⁰ in non–Title I schools must be “highly qualified” by 2005–6.³¹ Pursuant to the Act and accompanying regulations, teachers are considered “highly qualified” if they are fully certified and have demonstrated competency in the subjects they teach. Competence is assumed if the teacher majored in the subject in college; alternatively, it can be

²³ Taylor, *supra* note 20, at 312–15; Weckstein, *supra* note 16, at 328–29.

²⁴ § 6311(b)(1)(A), (C).

²⁵ § 6311(B)(2)(A). In the same year that it passed the IASA, the federal government enacted Goals 2000, which provided money to states to assist them in developing academic standards for all students. *See* Goals 2000: Educate America Act, Pub. L. No. 103-227, 108 Stat. 130.

²⁶ NCLBA § 1111(b)(1)(A)–(C).

²⁷ §§ 1111(b)(3), 1116.

²⁸ § 1119(a)(1).

²⁹ § 1119(a)(3).

³⁰ “Core” subjects actually make up a long list, including English, reading, language arts, math, science, foreign languages, civics and government, economics, arts, history, and geography—essentially every subject except for vocational training. *See* Improving the Academic Achievement of the Disadvantaged, 34 C.F.R. § 200.55 (2003).

³¹ NCLBA § 1119(a)(2).

demonstrated by passing a state test or, for existing teachers, by convincing state evaluators that they know their subject areas.³²

As for testing and accountability, whereas the IASA required testing in math and reading at three points in a student's school career, the NCLBA requires annual testing in reading and math in grades three through eight. At least one more test in reading and math must be given in grades ten through twelve.³³ Beginning in 2007–8, students must also be tested in science at least three times between grades three through twelve.³⁴

Test scores are the fuel that makes the NCLBA run. Scores are tabulated for schools in the aggregate and must be disaggregated for a number of subgroups, including migrant students, disabled students, English-language learners, and students from all major racial, ethnic, and income groups.³⁵ All of these scores are then used to determine whether schools are making "adequate yearly progress." Adequate yearly progress (AYP), in turn, is the linchpin of the NCLBA.

Adequate yearly progress is tied to whether a sufficient percentage of students are performing proficiently on state tests.³⁶ The NCLBA requires states to bring *all* students to the proficient level within twelve years of the Act's passage (i.e., by 2014), and states must ensure that their definitions of adequate yearly progress will enable the ultimate twelve-year goal to be met.³⁷ To accomplish this, states must set a proficiency goal each year, and that percentage must rise periodically so that by 2014, it hits 100%. For a school to make adequate yearly progress, the student population as a whole, *as well as each identified subgroup of students*, must meet the same proficiency goal.³⁸ For example, if in the year 2004–5, the state determines that 65% of students must be "proficient" on the tests, 65% of all the students within a school *and* 65% of the students within each subgroup (e.g., disabled students, poor students, minority students) must be per-

³² § 9101(23); 34 C.F.R. § 200.56.

³³ NCLBA § 1111(b)(3). The annual testing requirement in grades three through eight does not begin until the 2005–6 school year. Until that time, students must be tested at least once in grades three through five, once in grades six through nine, and once in grades ten through twelve. § 1111(b)(3)(C)(v)(I).

³⁴ § 1111(b)(3)(C)(v)(II).

³⁵ § 1111(b)(3)(C)(xiii).

³⁶ § 1111(b)(2)(C). To determine adequate yearly progress (AYP), states must also look to graduation rates and at least one additional academic indicator, such as attendance rates. But these additional indicators need not be set at any particular level, nor need they increase over time. *See infra* note 168 and accompanying text. Moreover, the additional indicators can only be used *against* schools; failure to post certain test scores, for example, cannot be excused by a high graduation rate. § 1111(b)(2)(C)(vi)–(D).

³⁷ § 1111(b)(2)(F).

³⁸ § 1111(b)(2)(C).

forming proficiently for a school to be making adequate yearly progress.

Adequate yearly progress is thus less about yearly achievement gains than it is about hitting uniform benchmarks. All states must set a uniform bar for achievement for all schools and all subgroups of students within a school. The first benchmarks were based on test scores from 2001–2. Using these test scores, states had to establish a starting point for AYP that was the higher of the following two values: (1) the percentage of students in the lowest-achieving subgroup, state-wide, who were performing proficiently; or (2) the threshold percentage of students performing proficiently in the lowest-performing quintile of schools statewide.³⁹ If 30% of a state's poor students, for example, scored at the proficient level in 2001–2, while 40% of all students in the school at the twentieth percentile of achievement scored at the proficient level, the initial AYP bar must be at least 40% for all schools *and* all subgroups of students. According to the language of the Act,⁴⁰ the percentage of students performing proficiently must rise every two or three years, like stair steps, until the 2013–14 school year, when all students must be scoring at the proficient level.⁴¹

Although the Act is quite strict in defining AYP, it is remarkably loose with regard to state standards and tests. (The basic reason for this structure is the continued resistance to national standards and tests.⁴²) States are free to determine their own standards, to create their own tests, and to determine for themselves the scores that indi-

³⁹ § 1111(b)(2)(E).

⁴⁰ As explained below, the Department of Education has approved state plans that do not conform to the language of the Act, insofar as they do not require steady growth in performance levels among schools. See *infra* Part III.B.

⁴¹ § 1111(b)(2)(H). The Act contains a so-called “safe harbor” provision, which in theory is supposed to ease the burden of schools in meeting AYP. If students in the aggregate, or any subgroup within a school, fail to meet the target for AYP, but at least ten percent more of the relevant group score at the proficient mark than did the prior year, the school will be deemed to have made AYP. § 1111(b)(2)(I)(i). The safe-harbor provision is unlikely to provide much relief, however, because the annual achievement gains required are still quite ambitious and, indeed, arbitrary. The important point to recognize is that the provision applies not to the same group of students, but to different groups. Each year, schools face a new set of students who may be stronger or weaker than the preceding set. If, for example, a set of students in one year begins even further behind than did the preceding year's students, it may be totally unrealistic to expect this new group of students to perform even better on tests than the preceding group. For these reasons, it is unlikely that the safe-harbor provision will preclude large numbers of schools from being deemed “in need of improvement” under the NCLBA.

⁴² For descriptions of earlier, failed attempts to enact national standards and tests, see, for example, Chester E. Finn, Jr., *Who's Afraid of the Big Bad Test?*, in *DEBATING THE FUTURE OF AMERICAN EDUCATION: DO WE NEED NATIONAL STANDARDS AND ASSESSMENTS?* 120, 133–43 (Diane Ravitch ed., 1995) [hereinafter *DEBATING THE FUTURE OF AMERICAN EDUCATION*], David Nather, *Education: High Expectations, Limited Means*,

vidual students must receive in order to be deemed "proficient." The harder the tests or the higher the scores needed to be deemed proficient, the harder it will be for schools to meet the NCLBA's definition of adequate yearly progress. For the same reasons, some states have much farther to travel than others in order to meet the goal of 100% proficiency. The starting percentages in Massachusetts, for example, were roughly 40% proficiency in reading and 20% proficiency in math. In Colorado, the starting percentages ranged, depending on the grade level, from roughly 75%–90% in reading and 50%–80% in math.⁴³

The Act requires all schools within a state, regardless of whether they receive Title I funding, to make adequate yearly progress. It also requires states and districts to disseminate information about each school's AYP status.⁴⁴ The stricter accountability mechanisms, however, are reserved for schools receiving Title I funding.⁴⁵ Although this is obviously a subset of all schools, it is a surprisingly large one. Over half of all schools in the nation receive Title I funds.⁴⁶ Not all of these schools, moreover, are in predominantly poor districts. Because Title I funds are widely distributed, it is not uncommon for relatively poor schools in predominantly middle class suburbs to receive Title I funding.⁴⁷

Those schools that receive federal funding and fail to make adequate yearly progress are identified as in need of improvement.⁴⁸ They are also subject to a range of progressively more serious actions. After two consecutive years of failure, schools must develop a plan for

58 CQ WKLY. 2394 (2000), and Diana Jean Schemo, *Schoolbook Balancing Act*, N.Y. TIMES, Jan. 24, 2001, at A1.

⁴³ ROBERT L. LINN, ACCOUNTABILITY: RESPONSIBILITY AND REASONABLE EXPECTATIONS 15 (Regents of the University of California 2003), available at <http://cresst96.cse.ucla.edu/reports/R601.pdf>.

⁴⁴ § 1111(b)(2)(C)(v), (G)(iii), (I).

⁴⁵ See § 1116(b)(1)(A), (5), (7), (8) (describing remedial measures mandated by NCLBA for schools receiving Title I funds that fail to make AYP).

⁴⁶ See DEP'T OF EDUC., FACT SHEET ON TITLE I, PART A (Aug. 2002), at <http://www.ed.gov/rschstat/eval/disadv/title1-factsheet.doc> (indicating that fifty-eight percent of all public schools are Title I schools); see also NAT'L CTR. FOR EDUC. STATISTICS, NCES STATISTICAL ANALYSIS REPORT: OVERVIEW OF PUBLIC AND ELEMENTARY SCHOOLS AND DISTRICTS, SCHOOL YEAR 2001–02 (2003) (listing Title I schools and percentage of students within those schools), http://nces.ed.gov/pubs2003/overview03/table_01.asp.

⁴⁷ All school districts with at least a two percent child poverty rate are entitled to Title I funding, and all schools within Title I districts are eligible for funds if the poverty rate of the school's attendance area is equal to or greater than the district's poverty rate. § 1113(a)(2)(B). As a result, schools with low poverty rates in districts with even lower poverty rates are eligible for Title I funds. Indeed, such schools may be more likely to receive Title I funding than schools with higher poverty rates that are located in high-poverty districts.

⁴⁸ § 1116(a)(1).

improvement and are supposed to receive “technical” assistance.⁴⁹ Students in those schools are also allowed to choose another public school, including a charter school, within the same district.⁵⁰ After three years, students who have not already departed for greener pastures must be provided with tutoring services from an outside provider, public or private.⁵¹ Those schools that fail to make AYP for four consecutive years must take one of several measures, including replacing school staff or instituting a new curriculum, and those that fail for five years in a row must essentially surrender control to the state government, which can reopen the school as a charter school, turn over management to a private company, or take over the school itself.⁵²

One final testing requirement should be mentioned, because it is important to the argument below. The NCLBA requires that the National Assessment of Educational Progress (NAEP) reading and math tests be administered every two years to fourth and eighth graders.⁵³ The NAEP is an extensive testing program that has been used for over thirty years to collect data about student achievement.⁵⁴ The test is essentially national, in that it is not aligned with any state standards. Instead, the NAEP attempts to measure content and skills thought common to all state educational systems. Prior to the NCLBA, participation in the NAEP was voluntary, but now all states must participate. Nonetheless, only a random sample of students within each state must take the test, and scores are not reported for individual students or individual schools.⁵⁵ The NCLBA does not indicate what is supposed to be done with the results of the NAEP, but supporters of the Act suggest that results on the NAEP will ensure the rigor of standards and tests used in each state.⁵⁶ Whether this use

⁴⁹ § 1116(b)(3), (4).

⁵⁰ § 1116(b)(1)(A), (E).

⁵¹ § 1116(b)(1)(E), (5).

⁵² § 1116(b)(7)–(8).

⁵³ § 1111(c)(2).

⁵⁴ See *BD. ON TESTING AND ASSESSMENT, NAT’L RESEARCH COUNCIL, HIGH STAKES: TESTING FOR TRACKING, PROMOTION, AND GRADUATION* 34 (Jay P. Heubert & Robert M. Hauser eds., 1999).

⁵⁵ Mark D. Reckase, *Using NAEP to Confirm State Test Results: An Analysis of Issues*, in *NO CHILD LEFT BEHIND: WHAT WILL IT TAKE?* 11, 12 (Thomas B. Fordham Found. ed., 2002), available at <http://www.edexcellence.net/doc/NCLBreport.pdf>.

⁵⁶ See, e.g., Liebman & Sabel, *supra* note 6, at 1729–30 (suggesting that NAEP results will push states to make their tests more rigorous). See generally Lynn Olson, *Want to Confirm State Test Scores? It’s Complex, But NAEP Can Do It*, *EDUC. WK.*, Mar. 13, 2002, at 1 (explaining hope of NCLBA advocates that NAEP will be used to judge rigor of state standards and tests).

of the NAEP will be successful in keeping state standards and tests rigorous is subject to serious question, as the next Part explains.

II

PERVERSE INCENTIVES AND UNINTENDED CONSEQUENCES

One can certainly disparage the NCLBA's singular focus on testing, as many have. But it is difficult to criticize the overarching goal of the Act, which is to ensure that all students are academically proficient in the not-so-distant future. The problem with the NCLBA is that it creates incentives that work against this central goal.

The Act creates counterproductive incentives by establishing overly ambitious achievement goals and imposing significant sanctions for failing to meet those goals. It allows states to act on these incentives by leaving them free to create their own tests and scoring systems. This odd combination of regulatory stringency and laxity, as described below, could well prove disastrous. It will encourage states to lower their standards, make their tests easier, or lower the scores needed to be deemed proficient. It will promote greater segregation by class and race. And, finally, it will help push talented teachers away from schools likely to be deemed failing, or from teaching altogether.

A. *Driving a Race to the Bottom*

The NCLBA, less than two years old, is already causing consternation and alarm in most, if not all, states.⁵⁷ Some of the complaints by state and local officials concern the costs associated with the Act and the failure of the federal government to cover those costs.⁵⁸ Many complaints, however, relate to the fact that a large number of schools in every state are likely to be deemed "failing" because of the Act. Based on test scores from 2001–2, many predict that a large percentage of schools within every state will fail to make adequate yearly progress.⁵⁹ Some schools will fail because aggregate test scores are too low; others will fail because one or more subgroups do not meet the AYP target. Although precise predictions vary, most state officials suggest that at least half of their public schools will fail to meet the

⁵⁷ See Michael Winerip, *A Pervasive Dismay On a Bush School Law*, N.Y. TIMES, Mar. 19, 2003, at B7 ("In all the world, the loneliest people must be that handful of men and women of the Department of Education dispatched by the Bush administration to wander the country, defending the No Child Left Behind Act. Talk about friendless.").

⁵⁸ See *supra* note 9 and accompanying text.

⁵⁹ See CTR. ON EDUC. POL'Y, *supra* note 1, at 32 (discussing various estimates made in different reports).

benchmark.⁶⁰ Some estimates are much higher, with state and federal officials predicting that in some states, close to 90% of the schools will fail to meet their targets.⁶¹ Early testing returns are somewhat mixed but generally support these dire predictions.⁶²

If the tests and scoring systems remain the same, moreover, the number of schools that fail to meet their marks will likely rise over time. The reason is two-fold. First, the Act requires a rate of improvement on test scores that will be incredibly difficult, if not impossible, to sustain.⁶³ As the percentage of students who must demonstrate proficiency rises, it will be increasingly difficult for schools to meet the mark. Second, even if schools initially keep pace, as the benchmark rises, those students whose performance must improve will necessarily be those most difficult to educate. It is perhaps not surprising, then, that commentators consider the ultimate goal of achieving 100% proficiency in twelve years to be utterly unrealistic.⁶⁴

Schools that fail to make AYP will likely be deemed failures, which in turn will generate pressure on state and local officials to do something to avoid that label. To be clear, schools that do not make AYP are not labeled failures by the Act itself. Title I schools that fail to make AYP are deemed in need of improvement,⁶⁵ while schools that do not receive federal funding are not labeled or otherwise sanctioned at all. Nonetheless, the media have translated “in need of improvement” to mean “failing,” fueling the popular perception that any school that does not make AYP—regardless of whether it receives

⁶⁰ See Dillon, *supra* note 12, at 33; Lynn Olson, *All States Get Federal Nod on Key Plans*, EDUC. WK., June 18, 2003, at 1.

⁶¹ See Diana Jean Schemo, *Sidestepping of New School Standards is Seen*, N.Y. TIMES, Oct. 15, 2002, at A21; Winerip, *supra* note 57.

⁶² See Alexandra Marks, *As Schools “Fail,” Parents Talk Transfers*, CHRISTIAN SCI. MONITOR, July 21, 2003, at 1 (reporting that more than 25% of New York City schools; 60% of Charlotte, North Carolina schools; and 70% of all schools in New Mexico failed to meet AYP benchmarks in 2003); Kate N. Grossman & Rosalind Rossi, *More Schools Not Making Grade*, CHI. SUN-TIMES, July 24, 2003, at 6 (reporting that 630 schools in 90 districts in Illinois failed to make AYP).

⁶³ See LINN, *supra* note 43, at 9 (explaining that growth rates required of most schools greatly exceed past growth rates).

⁶⁴ Massachusetts Board of Education member Abigail Thernstrom, for one, calls the goal of one hundred percent proficiency in twelve years “ludicrous” and suggests that it can only be accomplished, at least in Massachusetts, by defining “‘proficiency way down . . . way, way down.’” Lynn Olson, *A ‘Proficient’ Score Depends on Geography*, EDUC. WK., Feb. 20, 2002, at 1; see also LINN, *supra* note 43, at 9 (suggesting that achieving one hundred percent proficiency in twelve years would be “miraculous”).

⁶⁵ NCLBA § 1116(a)(1).

Title I funding—is a failing school.⁶⁶ In Title I schools, this popular perception will be reinforced by the sanctions imposed for failure to make AYP. After two consecutive years of failure, for example, parents will be told that their children have the right to attend another public school that has not been deemed in need of improvement.⁶⁷ Regardless of the intended message, telling parents that their children can flee their neighborhood school undoubtedly will convey the impression that the school is failing.⁶⁸

Needless to say, state and local officials are not pleased by the prospect that half or more of their schools will be considered failing by the media and the public. Indeed, such a result is politically unacceptable. The only real question is what states will do to avoid it.

1. *Avoiding Failure*

States have four options. First, they could direct their energy and resources in an earnest effort to improve achievement, hoping against hope that they can pull off a miracle and meet the Act's goals. Given the slim chances that such a strategy would be successful, it is doubtful that many states will follow this approach, or will follow *just* this approach.

Second, states could stall by setting annual yearly progress goals in such a way as to postpone the need for large increases until later in the twelve-year period. Rather than requiring the same rate of progress throughout the twelve-year period, states could set up a system akin to a balloon mortgage, with the largest gains required in the last

⁶⁶ See, e.g., Tess Nacelewicz, *Maine May Resist Federal School Law*, PORTLAND PRESS HERALD (Maine), May 29, 2003, at 1A (noting that “all of Maine’s schools could be deemed failing under the No Child Left Behind law by 2009”); Diane Jean Schemo, *New Federal Rule Tightens Demands on Failing Schools*, N.Y. TIMES, Nov. 27, 2002, at A1 (noting that NCLBA “is expected to result in vast numbers of schools being designated as failing”); Michael Winerip, *Defining Success in Narrow Terms*, N.Y. TIMES, Feb. 19, 2003, at B7 (describing school that next fall “will be labeled ‘failing’ under” the NCLBA). The use of the “failing” label has become sufficiently widespread that advocates of the NCLBA have acted to debunk it, as evidenced by a recent paper authored by Kati Haycock and Ross Wiener of the Education Trust, which supports the NCLBA. Haycock and Wiener go out of their way to dispel the “[m]yth” that “[t]he AYP process will stigmatize schools as ‘failing.’” They emphasize that the NCLBA identifies schools as in need of improvement and that “[n]eeds improvement” does not mean ‘failing.’” Kati Haycock & Ross Wiener, *Adequate Yearly Progress Under NCLB*, in IMPLEMENTING THE NO CHILD LEFT BEHIND ACT 3 (National Center on Education and the Economy et al. eds., 2003), available at http://www.pponline.org/documents/Ed_NCLB_0403.pdf.

⁶⁷ § 1116(b)(1)(E), (5).

⁶⁸ Marks, *supra* note 62 (describing how letters notifying parents of right to transfer send message, whether deserved or not, that there is something wrong with their children’s current school); see also William L. Taylor, *Title I as an Instrument for Achieving Desegregation and Equal Educational Opportunity*, 81 N.C. L. REV. 1751, 1759 (2003) (suggesting that allowing students to choose another school is “a public confession of failure”).

few years.⁶⁹ Roughly twenty states have already proposed such an approach. Somewhat surprisingly, the Department of Education has approved these proposals,⁷⁰ despite language in the Act that requires steady progress toward 100% proficiency.⁷¹ This strategy might help states delay the day of reckoning, and below I argue that states should make constructive use of the Department's willingness to tolerate this sort of postponement.⁷² If the Act is not modified or repealed, however, postponement will not be enough.

A third strategy would be to ignore the federal mandate or to decline Title I funding. Federal enforcement of Title I has been notoriously lax, and some states might bet on continued laxity.⁷³ Similarly, some states might turn down Title I funds in order to escape the mandates of the NCLBA.⁷⁴ Until now, however, states have neither ignored the Act nor turned down federal funds. Although large-scale defection may indeed occur over time, at the moment it appears that all states are attempting to comply with the letter of the law.⁷⁵

This leaves one last option: make the tests easier or lower the score needed to be considered proficient. Of all the options, this seems like the one most likely to succeed in "raising" achievement to acceptable levels both in the short and long runs. It is also a strategy that is not only tolerated by the Act, but implicitly encouraged by it.⁷⁶ Recall that the Act requires states to set "challenging" standards and test students regularly, and that annual yearly progress requires a cer-

⁶⁹ See Lynn Olson, *States' Plans Likely to Test ESEA Pliancy*, EDUC. WK., Feb. 19, 2003, at 1 (describing this strategy).

⁷⁰ Lynn Olson, *Education Department Accepts Variety of Strategies*, EDUC. WK., June 18, 2003, at 21.

⁷¹ § 1111(b)(2)(H) (stating that AYP targets "shall . . . increase in equal increments" over twelve-year period).

⁷² See *infra* Part III.B.

⁷³ See Michael Cohen, *Implementing Title I Standards, Assessments and Accountability*, in *NO CHILD LEFT BEHIND: WHAT WILL IT TAKE?* 75, 83–84 (Thomas B. Fordham Found. ed., 2002) (describing enforcement history of Title I), available at <http://www.edexcellence.net/doc/NCLBreport.pdf>.

⁷⁴ Michelle R. Davis, *Paige Stresses Flexibility of Education Law*, EDUC. WK., Mar. 17, 2004, at 40 (describing Utah's plan, since put on hold, to turn down federal aid in order to avoid NCLBA's requirements).

⁷⁵ See, e.g., Olson, *supra* note 60 (noting that all fifty states submitted plans describing steps that they have taken, and will take, to comply with NCLBA); CTR. ON EDUC. POL'Y, *supra* note 1, at 19 (reporting that "[s]tates are making a good faith effort to comply with the new law").

⁷⁶ As one expert on testing and accountability observed, the "severe sanctions" of the NCLBA "implicitly encourage states to water down their content and performance standards in order to reduce the risk of sanctions." Sam Dillon, *States are Relaxing Education Standards to Avoid Sanctions from Federal Law*, N.Y. TIMES, May 22, 2003, at A29 (quoting Professor Robert L. Linn, immediate past president of American Educational Research Association).

tain percentage of students to score at the “proficient” level. At the same time, the Act leaves states free to establish their own standards and tests and to determine the score needed to be considered proficient. The harder the test or the higher the proficiency score, the more schools that will be deemed in need of improvement. The converse is obviously true, and it will not take long for states to figure this out.

Indeed, some already have. Louisiana, Colorado, Connecticut, and Texas have all tinkered with their scoring systems in order to increase the number of students who will be deemed proficient for purposes of the NCLBA. In Louisiana, for example, passing scores had been divided into three categories: basic, proficient, and advanced. Last year, only 17% of eighth graders scored at the proficient or advanced level on an English test, while 31% scored at the basic category; in math, only 5% were advanced or proficient while 37% scored at the basic level. So what did Louisiana do? It deemed those who scored at the basic level “proficient” for purposes of the NCLBA. Similarly, Colorado and Connecticut have redefined categories of scores, making it easier for students to reach the newly dubbed “proficient” level. And the Texas State Board of Education, after a field trial of state tests, lowered the number of questions students must answer correctly in order to be considered proficient on the third-grade reading test.⁷⁷

2. *A Puzzle*

In one respect, this reaction seems completely predictable, and we should not be surprised to see more states following the lead of Louisiana, Colorado, Connecticut, and Texas. Other states may alter not just the scoring system but the tests themselves, making them easier to pass. That said, there is something slightly puzzling about this reaction. More precisely, there is something puzzling about why the NCLBA should alter a state’s incentives vis-à-vis standards and testing. The puzzle is worth exploring because it calls into question the need for the NCLBA and sheds some light on the larger issue of the proper role of the federal government in education law and policy.

⁷⁷ Sources for this paragraph are Dillon, *supra* note 76 (discussing lowering of state testing standards in Texas, Michigan, and Colorado); David J. Hoff, *States Revise the Meaning of ‘Proficient,’* EDUC. WK., Oct. 9, 2002, at 1 (describing changes in Louisiana, Colorado, and Connecticut); and Schemo, *supra* note 61 (describing efforts by officials in Connecticut, Colorado, and Michigan to revisit their standards for academic proficiency in light of NCLBA). See also Richard Rothstein, *How U.S. Punishes States that Set Higher Standards*, N.Y. TIMES, Sept. 18, 2002, at B8 (noting that officials in Massachusetts have considered lowering standards in order to avoid federal sanctions).

Prior to the NCLBA, states had already established testing regimes and instituted some accountability systems. Some states held students accountable for failing the tests; others employed various measures to hold schools accountable, which ranged from issuing report cards to dismissing administrators to taking over perpetually failing schools. The difficulty of state tests and the stringency of scoring systems varied considerably from state to state.⁷⁸ Some states—such as Massachusetts, Michigan, and New York—established ambitious achievement goals, which many students failed to meet, while others established lower standards and goals.⁷⁹

The question is: Why would any state establish ambitious goals that are quite hard to meet? The obvious consequence of doing so is to make it seem like a large number of students and schools are not performing well. Why not devise tests that make students and schools within a state look good? Put differently, why would the NCLBA cause a race to the bottom? Why would states not already engage in this race on their own?

To understand this puzzle and potential solutions, it is useful to begin with some of the arguments made within the economics and political science literature on interjurisdictional competition. A running debate within that literature concerns the extent to which competition among states and localities for firms and individuals is likely to be productive or destructive.⁸⁰ Some commentators argue that such competition tends to promote efficiency by acting as a check on excessive taxation and as a spur to productive spending on infrastructure and public services.⁸¹ Others contend that, at least under certain cir-

⁷⁸ *CTR. ON EDUC. POL'Y*, *supra* note 1, at 32-33.

⁷⁹ See Dillon, *supra* note 12 (discussing Michigan's standards); Olson, *supra* note 64 (describing how rigor of tests varies from state to state); Rothstein, *supra* note 77 (same). See generally THOMAS B. FORDHAM FOUND., *THE STATE OF STATE STANDARDS* (Chester E. Finn, Jr. & Michael J. Petrilli eds., 2000) (describing standards and accountability systems in fifty states), available at <http://www.edexcellence.net/library/sooss2000/Standards2000.pdf> (Jan. 2000).

⁸⁰ See Daphne A. Kenyon & John Kincaid, *Introduction to COMPETITION AMONG STATES AND LOCAL GOVERNMENTS: EFFICIENCY AND EQUITY IN AMERICAN FEDERALISM* 1, 1-33 (Daphne A. Kenyon & John Kincaid eds., 1991) [hereinafter *COMPETITION AMONG STATES AND LOCAL GOVERNMENTS*] (reviewing literature); see also Richard L. Revesz, *Rehabilitating Interstate Competition: Rethinking the 'Race-to-the-Bottom' Rationale for Federal Environmental Regulation*, 67 *N.Y.U. L. REV.* 1210, 1210-12 (1992) (same).

⁸¹ See, e.g., Wallace E. Oates & Robert M. Schwab, *The Allocative and Distributive Implications of Local Fiscal Competition*, in *COMPETITION AMONG STATES AND LOCAL GOVERNMENTS*, *supra* note 80, at 127, 127-45; Robert W. Rafuse, Jr., *Some Observations on Interjurisdictional Competition and Federal Grant Policy*, in *COMPETITION AMONG STATES AND LOCAL GOVERNMENTS*, *supra* note 80, at 167, 167-68; John Shannon, *Federalism's "Invisible Regulator"—Interjurisdictional Competition*, in *COMPETITION AMONG STATES AND LOCAL GOVERNMENTS*, *supra* note 80, at 117, 117-25. While arguing

cumstances, such competition is likely to promote a race to the bottom, with states softening or abandoning worthwhile regulations in order to attract business.⁸² Still others point out, correctly, that in trying to understand or predict policy decisions, it is myopic to focus solely on the effects and incentives of competition to the exclusion of internal local and state politics.⁸³

Public education plays a fascinating, if under-explored, role in interjurisdictional competition. Survey and anecdotal evidence suggest that the perceived quality of public schools influences decisions by firms and individuals as to where to locate.⁸⁴ Regardless of the precise strength of this influence, state and local officials seem to

that interjurisdictional competition may foster efficiency, all of the authors cited recognize that it may also impede equity, insofar as it constrains the redistribution of income.

⁸² See, e.g., Therese J. McGuire, *Federal Aid to States and Localities and the Appropriate Competitive Framework*, in COMPETITION AMONG STATES AND LOCAL GOVERNMENTS, *supra* note 80, at 153, 153–66 (arguing that competition among states, but not among localities, is likely to be destructive and result in “the provision of inefficiently low levels of public goods and services”).

⁸³ John E. Chubb, *How Relevant is Competition to Government Policymaking?*, in COMPETITION AMONG STATES AND LOCAL GOVERNMENTS, *supra* note 80, at 57, 60–62.

⁸⁴ See, e.g., THOMAS M. SMITH ET AL., U.S. DEP’T OF EDUC., *THE CONDITION OF EDUCATION 1997*, at 23 (1997) (reporting results of National Household Education Survey, in which approximately fifty percent of parents whose children attended neighborhood schools claimed that “their choice of residence was influenced by where their children would go to school”); Kenneth N. Bickers & Robert M. Stein, *The Microfoundations of the Tiebout Model*, 34 *URB. AFF. REV.* 76, 88 (1998) (reporting survey results which indicated that ninety-two percent of respondents with school-age children stated that “the quality of schools is important or very important in influencing their decision” of where to live); F.J. Calzonetti & Robert T. Walker, *Factors Affecting Industrial Location Decisions: A Survey Approach*, in *INDUSTRY LOCATION AND PUBLIC POLICY* 221, 232–34 (Henry W. Herzog, Jr. & Alan M. Schlottmann eds., 1991) (reporting survey results which indicate that manufacturing plants rank education as important to their locational choice among regions); Henry W. Herzog, Jr. & Alan M. Schlottmann, *Metropolitan Dimensions of High-Technology Location in the U.S.: Worker Mobility and Residence Choice*, in *INDUSTRY LOCATION AND PUBLIC POLICY*, *supra*, at 169, 176–77 (reporting survey results which indicate that high-technology firms rank academic institutions as important to their locational choice among regions and proximity to good schools as important to their choice within regions); Caroline M. Hoxby, *The Effects of School Choice on Curriculum and Atmosphere*, in *EARNING AND LEARNING: HOW SCHOOLS MATTER* 281, 299–312 (Susan E. Mayer & Paul E. Peterson eds., 1999) (finding that parents consider academic quality and discipline when choosing among public schools in metropolitan areas); Michael Wasylenko, *Empirical Evidence on Interregional Business Location Decisions and the Role of Fiscal Incentives in Economic Development*, in *INDUSTRY LOCATION AND PUBLIC POLICY*, *supra*, at 13, 29 (reviewing empirical literature on firm location and concluding that there is some evidence “that public expenditures on education and other functions enhance the attractiveness of a state [to firms] and lead to employment growth”). See generally Jeffrey R. Henig & Stephen D. Sugarman, *The Nature and Extent of School Choice*, in *SCHOOL CHOICE AND SOCIAL CONTROVERSY: POLITICS, POLICY, AND LAW* 13, 14–17 (Stephen D. Sugarman & Frank R. Kemerer eds., 1999) (discussing parents’ residential decisions as means of school choice).

believe that the perceived quality of public schools matters.⁸⁵ These officials typically act as if “good” public schools attract businesses and are thus useful in the competition among states and localities for firms.⁸⁶ These officials also apparently believe, as do local homeowners, that good public schools boost property values.⁸⁷ These beliefs provide a strong incentive for state and local officials to maintain public schools that are perceived to be of high quality.

I emphasize perceived rather than actual quality because the latter is always going to be contested. Even with perfect information about a given set of schools, there will be room for disagreement about what makes one school better than another,⁸⁸ simply because there is room for disagreement about the goals of education. It thus makes more sense to think in terms of qualities that are considered by many to make a school good or bad. More importantly, perceived quality matters precisely because firms and individuals will usually possess limited information when assessing public schools. They will of necessity rely on heuristics or proxies for quality.⁸⁹

Whether interjurisdictional competition will cause a race to the bottom or the top in the school context is thus a fairly complicated question. It depends, first, on some agreement on the goals of public education, such as imparting a particular set of skills or instilling a certain body of knowledge, preparing students for college, the workplace, and citizenry, and inculcating healthy attitudes and strong values. It then depends, crucially, on how closely proxies for quality relate to these goals or, put differently, how closely proxies relate to “actual” quality. One obvious proxy for quality, for example, is school funding, which is used by parents and firms to judge particular schools. In the past, interjurisdictional competition has prompted some low-spending states to spend more on schools in order to match

⁸⁵ The ubiquity of gubernatorial candidates campaigning to be the “education governor” and urban mayors vowing to reform school systems are two obvious manifestations of this perception.

⁸⁶ See Shannon, *supra* note 81, at 120–21; William Raspberry, *Mississippi Moving Up*, WASH. POST, May 30, 1988, at A27; see also Diane Ravitch, *Introduction to DEBATING THE FUTURE OF AMERICAN EDUCATION*, *supra* note 42, at 2 (noting that state governors, believing that “the future health of their state . . . depended on improving education,” were among early supporters of education improvement efforts).

⁸⁷ See WILLIAM A. FISCHEL, *THE HOMEVOTER HYPOTHESIS: HOW HOME VALUES INFLUENCE LOCAL GOVERNMENT TAXATION, SCHOOL FINANCE, AND LAND-USE POLICIES* 4–18 (2001) (arguing that homeowners want to promote local school quality in order to protect their housing investments).

⁸⁸ See, e.g., Bickers & Stein, *supra* note 84, at 83 (recognizing that parents may use different criteria—such as test scores, athletic programs, and safety—in selecting schools).

⁸⁹ See Bickers & Stein, *supra* note 84, at 78 (arguing that citizens “use informational heuristics and proxies . . . to locate in jurisdictions that provide them with desired level of collective goods”).

or at least approach their higher-spending neighbors.⁹⁰ But whether more school funding actually correlates with a better education is a notoriously difficult question.⁹¹ Unless the proxy of school funding relates to something we really care about or some “genuine” quality, it is hard to say whether competition that leads to higher funding is truly productive.

To return to the focus of this Article, another obvious proxy for school quality is test scores.⁹² And, indeed, studies confirm that a school’s test scores influence public perceptions of the school’s quality.⁹³ The theory and limited empirical evidence regarding interjurisdictional competition suggest that state and local officials have a strong incentive to make their schools “look good” by focusing on such proxies for quality. This means that they have a strong incentive to boost test scores, which means that states on their own might engage in a race to the bottom by creating easy tests that are not accurate proxies for quality but nonetheless give the impression that their schools are good.⁹⁴ Hence the puzzle: Why would a state ever use a test or scoring system that does not make most schools look good? Why wouldn’t states instead try to manipulate tests or scoring systems to give the impression that their schools are excellent? And why should we expect the NCLBA to alter a state’s existing incentives regarding testing and scoring systems?

⁹⁰ Shannon, *supra* note 81, at 119–21.

⁹¹ See James E. Ryan, *Schools, Race, and Money*, 109 *YALE L.J.* 249, 268–70 (1999) (discussing literature on this topic).

⁹² Cf. Donald R. Haurin & David Brasington, *School Quality and Real House Prices: Inter- and Intrametropolitan Effects*, 5 *J. HOUSING ECON.* 351, 358–59 (1996) (arguing that test scores are good measure of school quality because they are easily observed by parents, given that scores are published in newspapers).

⁹³ See, e.g., Sandra E. Black, *Do Better Schools Matter? Parental Valuation of Elementary Education*, 114 *Q.J. ECON.* 577, 578 (1999) (finding positive relationship between test scores and housing prices within same school district); Haurin & Brasington, *supra* note 92, at 363 (finding positive relationship between test scores and housing prices in 134 jurisdictions in six metropolitan areas). Other studies suggest that parents do not have very accurate information about test scores but nonetheless are able to locate high-scoring districts through heuristics. See FISCHER, *supra* note 87, at 60–61, 154–55 (discussing studies). Fischer himself believes that Black’s study, which controls for neighborhood effects, “shows without much doubt that homebuyers do notice differences in test scores, or some school quality closely related to test scores, and are willing to pay a premium for them.” *Id.* at 155.

⁹⁴ Think here of the pressure among universities and graduate schools to inflate student grades. Cf. James W. Guthrie, *American Education Reform: What is Needed is “National” Not Federal*, 17 *ST. LOUIS U. PUB. L. REV.* 125, 129 (1997) (reporting findings of scandal and corruption in locally and state-administered exams); Hoxby, *supra* note 84, at 312 (finding that “greater parental choice [among public schools] appears to cause grade inflation”).

3. *A Solution*

At first blush, one might think that some states established challenging testing regimes because they could not get away with using easier tests or more generous scoring systems. This seems unlikely. It is doubtful that the general public pays enough attention to state tests and scoring systems to know that a particular testing regime is relatively easy or challenging. Moreover, the main and most powerful interest groups—teachers' unions—are unlikely to clamor for more difficult tests and stringent scoring systems. It would seem to follow that states have a good deal of political discretion to determine the level of difficulty of their testing regimes, a point that is partially borne out by the variation among state regimes.⁹⁵

Some states established strict testing regimes, I contend, because of three related factors: the ambiguity of test scores, the lack of significant sanctions, and the interaction of interjurisdictional competition and internal politics. As for the first factor, although a state could get away with creating easy tests, it could also defend a decision to create a more stringent testing scheme. Prior to the NCLBA, a state might have profitably claimed that more students and schools failed their tests because the academic goals in that state were relatively ambitious as compared to those in other states. In Massachusetts, for example, which has a very challenging battery of state tests, state officials defended the standards and tests as a way to establish a "world class" public school system. The goals, according to the chairman of the state school board, were "something to stretch for rather than something that simply validated the existing curriculum."⁹⁶ Thus, prior to the NCLBA, test scores were somewhat ambiguous proxies for quality, because generally lower scores in one state could be interpreted as a sign of a truly rigorous test.

Lower test scores generated by harder tests were also tolerable because sanctions were limited and relatively mild, at least for schools. Although states adopted accountability measures prior to the NCLBA, they reserved the most severe sanctions for students. In eighteen states, students had to (and still must) pass exit exams in order to graduate, while in others they had to (and still must) pass tests to be promoted from one grade to the next.⁹⁷ Schools, by con-

⁹⁵ Variation among state standards is only partial proof of political discretion because it is also possible that citizens of different states have varying preferences regarding education quality and rigor, and that these differing preferences account for the different standards.

⁹⁶ Greg F. Orlofsky & Lynn Olson, *The State of the States*, EDUC. WK., Jan. 11, 2001, at 86.

⁹⁷ Lynn Olson, *Overboard on Testing?*, EDUC. WK., Jan. 11, 2001, at 23.

trast, were subject primarily to the mild sanction of report cards, which evaluated their performance and identified “low-performing” schools based largely on test scores.

Although even this shaming sanction prompted some change,⁹⁸ relatively few schools received this sanction because they were judged by more lenient criteria than those established by the NCLBA. Schools in at least eighteen states, for example, were judged based on improvements over past performance.⁹⁹ As a result, even a school with relatively low test scores could be considered adequate if it made some improvement over prior years. In other states, schools were given more time than they will have under the NCLBA to raise performance. In Massachusetts, for example, even though the vast majority of students in the state scored below the proficiency level on state tests in 1998, only two schools in the entire state were deemed “underperforming” in 1999–2000.¹⁰⁰ What was true in Massachusetts was true elsewhere: In every state, a very small percentage of schools were labeled as low-performing. Not a single state labeled anywhere close to half of its schools as underperforming.¹⁰¹ Accountability for schools was thus quite soft prior to the NCLBA, and even the relatively mild sanction of labeling schools as low-performing or failing was used sparingly.

As for the third factor, the influence of external competition and internal politics, it is important to recognize that states need not ensure that all schools appear equally good. States interested in attracting firms, for example, need not work to make every school or district exemplary. They have to ensure that some schools within a metropolitan area are attractive to employees of sought-after firms.¹⁰² They also need to ensure there are some strong schools in order to give firms confidence that qualified local employees can be found. A similar motive exists with localities: They are interested in assuring that local schools are considered good. They have little or no reason to assure that neighboring schools are held in equal esteem.

⁹⁸ See Eric A. Hanushek & Margaret E. Raymond, *Improving Educational Quality: How to Evaluate Our Schools?*, in EDUCATION IN THE 21ST CENTURY: MEETING THE CHALLENGES OF A CHANGING WORLD 193, 217–18 (Yolanda K. Kodrzycki ed., 2002) (finding that states with report cards or other forms of accountability posted greater gains on NAEP tests than those that did not sanction or reward schools or disseminate test scores).

⁹⁹ Lynn Olson, *Finding the Right Mix*, EDUC. WK., Jan. 11, 2001, at 12, 17–18.

¹⁰⁰ Ulrich Boser, *Pressure Without Support*, EDUC. WK., Jan. 11, 2001, at 68.

¹⁰¹ See *id.*

¹⁰² Cf. Herzog & Schlottmann, *supra* note 84, at 176–77 (reporting that high-technology firms rank proximity to good schools as important to their locational choices within regions).

Relatively high-performing schools are most likely to be located in suburban areas.¹⁰³ Suburbanites, in turn, generally believe that their schools are good, and this widespread belief buoys property values. Suburbanites in most states also enjoy enough political clout to protect the perceived quality of their local schools.¹⁰⁴ As long as state test scores supported the belief that suburban schools were relatively good, state and local suburban officials could take some comfort in them. State officials could be confident of being competitive for firms and local suburban officials could satisfy parents and homeowners. Put differently, if the relatively few schools deemed low-performing were urban schools with bad reputations, external concerns about competition and internal political dynamics still could lead to general support of a fairly rigorous testing system.

This is not to suggest that testing prior to the NCLBA was uniformly supported by suburbanites or uniformly opposed by urban parents and educators. Some suburban parents oppose testing regimes, especially ones that require frequent testing or impose high stakes on students. They contend that preparing for the tests dumbs down the curriculum in good suburban schools.¹⁰⁵ Conversely, some urban parents and school officials embrace testing because they believe it causes urban schools to raise expectations for students.¹⁰⁶ My point is simply that an accountability system that casts relatively few schools in a bad light is going to generate less opposition than one that denigrates a wider range of schools.

Now consider the No Child Left Behind Act. It essentially changes all three of the factors described above. Instead of ambiguous test scores, which might be spun to a state's advantage, the NCLBA will result in a single judgment: success or failure to make adequate yearly progress. For schools that do not succeed, the ambiguity of test scores will thus be replaced with the certainty of a label: "failing." This label will likely become a chief proxy for school

¹⁰³ See Ryan, *supra* note 91, at 272–75 (discussing school demographics and achievement statistics).

¹⁰⁴ See James E. Ryan & Michael Heise, *The Political Economy of School Choice*, 111 *YALE L.J.* 2043, 2088–91 (2002) (describing successful efforts made by suburbanites to protect physical and financial autonomy of their local schools).

¹⁰⁵ See, e.g., Marc S. Tucker, *The Roots of Backlash*, *EDUC. WK.*, Jan. 9, 2002, at 76 (describing growing opposition to testing among suburban parents); Kathleen Kennedy Manzo, *Protests Over State Testing Widespread*, *EDUC. WK.*, May 16, 2001, at 1 (same); see also James Traub, *The Test Mess*, *N.Y. TIMES*, Apr. 7, 2002, § 6 (Magazine), at 46 (describing views of some parents in New York suburbs).

¹⁰⁶ See, e.g., Casserly, *supra* note 7 (noting urban school districts that have successfully embraced testing).

quality, one that will be difficult to rebut.¹⁰⁷ As acknowledged by Michael E. Ward, North Carolina's Superintendent of Education, "[i]t's going to take a lot of explaining," if and when large numbers of schools fail to make AYP.¹⁰⁸

The experience of Florida confirms that the "failing" label will dominate perceptions of school quality. Florida is one of only two states that explicitly assigns a letter grade, from A to F, to each school based on test scores.¹⁰⁹ A recent study of Florida's system found that the simple distinction between an "A" and "B" grade was valued in the housing market at over \$10,000, a difference of roughly eight percent. Even larger effects were found when analyzing the difference between a grade of "B" and a grade of "C."¹¹⁰ One can easily imagine similar or greater effects on housing prices, and public perceptions more generally, once AYP results are reported. If houses near schools that receive a grade of "B" are worth less than those near schools that receive a grade of "A," it stands to reason that houses near schools deemed "failing" under the NCLBA will be worth less than those near schools that have escaped that label.

In addition to making a more salient judgment about schools, the NCLBA also imposes swifter and harsher sanctions than most state systems. Instead of the relatively mild or rarely used sanctions now in place in many states, Title I schools will face increasingly harsh measures for each year that they fail to make adequate yearly progress. Perhaps most importantly, parents of students in "failing" Title I schools will be told that their children have the option of attending another school.¹¹¹ If those schools continue to fail, in a few years they

¹⁰⁷ See CTR. ON EDUC. POL'Y, *supra* note 1, at 31 (reporting that state officials surveyed said "that the language in NCLB concerning failing schools . . . would place an 'embarrassing label' on schools that they would find difficult to ever 'shake'"); Olson, *supra* note 60 (reporting that chairman of Virginia State Board of Education agreed to comply with federal law "'only under strong protest,'" expressing concern that schools given highest rating under state accountability system might nonetheless "be viewed as 'failing' in some respect under the federal law").

¹⁰⁸ Olson, *supra* note 60 (quoting Ward). Ward also admitted that one of the state's "most intensive efforts has been around a communication plan," presumably to provide an explanation that will placate parents and others concerned by schools' failure to make AYP. *Id.*

¹⁰⁹ Orlofsky & Olson, *supra* note 96. The other state is Colorado. *Id.*

¹¹⁰ DAVID N. FIGLIO & MAURICE E. LUCAS, WHAT'S IN A GRADE? SCHOOL REPORT CARDS AND HOUSE PRICES 2-5 (Nat'l Ctr. for the Study of Privatization in Educ., Occasional Paper No. 29, 2001), available at http://www.ncspe.org/publications_files/730_OP29.pdf. The study purports to control for other measures of school quality, as well as for neighborhood and property attributes. The data, though inconclusive, also suggests that the impact may be temporary. *Id.* at 3-5, 26.

¹¹¹ NCLBA § 1116(b)(1)(E).

either will be shut down and reopened as charter schools, taken over by the state, or taken over by a private management company.¹¹²

We have already seen, in the context of high-stakes testing, how real consequences can change the political dynamics of testing and accountability. In some states with graduation exams, as the date of implementation drew close and predictions suggested widespread student failure, state legislatures voted to postpone the day of reckoning.¹¹³ The California Board of Education, for example, recently voted to delay the implementation of its high school exit exam for two years in light of predictions that close to 100,000 students would not graduate if they had to pass the exams next year.¹¹⁴ Officials in New York and Florida recently made similar efforts to soften the impact of their states' respective high-stakes testing regimes.¹¹⁵ In other states, political protests surrounding testing became more intense and widespread as the date drew near on which high-stakes consequences would attach to test results.¹¹⁶

The sanctions imposed by the NCLBA, as well as the publicity that attends labeling a school a failure, similarly will alter the political dynamics of testing and will generate both external and internal pressure to lower standards. Pursuant to the NCLBA, a number of schools with relatively good reputations will be considered failing because of the Act, including some "good" suburban schools.¹¹⁷ This will alarm state officials interested in attracting firms because it will make it harder to demonstrate that at least suburban schools are of high quality. And as some states begin to manipulate their tests or scoring systems in an effort make their schools look good, this will create pressure on other states to follow suit.

¹¹² § 1116(b)(8).

¹¹³ See Lynn Olson, *States Debate Exam Policies for Diplomas*, EDUC. WK., May 14, 2003, at 1 (describing decisions made by officials in Florida and Massachusetts to postpone requirements that students pass tests in order to graduate); Traub, *supra* note 105 (noting that "most states have postponed exit tests rather than face the music" of denying diplomas to large numbers of students).

¹¹⁴ Greg Winter, *California Will Wait Until 2006 to Require High School Graduates to Pass Exit Exam*, N.Y. TIMES, July 10, 2003, at A17.

¹¹⁵ See Sam Dillon, *Citing Flaw, State Voids Math Scores*, N.Y. TIMES, June 25, 2003, at A1 (describing efforts in New York); Editorial, *Why Testing Can't Fail*, *supra* note 11 (same); Olson, *supra* note 113 (describing efforts in Florida).

¹¹⁶ See, e.g., Matthew I. Pinzur, *Dump the Test, Thousands Demand*, MIAMI HERALD, May 23, 2003, at 1B (describing protests in Florida).

¹¹⁷ See, e.g., Winerip, *supra* note 66 (describing popular local school that likely will be deemed "underperforming" by NCLBA). To cite another example of the gap between popular perceptions and the judgment of the NCLBA, President Bush described a Michigan elementary school that he visited as "excelling," just three months before it was identified as failing to make adequate yearly progress. See Dillon, *supra* note 12.

The NCLBA also will generate anxiety among suburban parents and homeowners. This group will surely demand to know why their local schools, which they thought were good and know are crucial to local property values, have been tagged as failures.¹¹⁸ In those suburbs whose schools receive Title I funding—and it bears repeating that there are many such suburbs—the reaction will be even more intense. Many suburban parents of children in Title I schools undoubtedly will be distressed to learn that students can transfer out of their neighborhood schools. All of this will change the internal politics of testing insofar as it will generate more opposition among politically powerful suburban districts.

The NCLBA thus alters the internal and external pressure on states and localities, and in so doing takes away some options that states had prior to the Act. To be sure, states still have the same basic incentive to make some or most schools look good. But prior to the Act, states could (and did) pursue various routes to achieve this goal. Some combined rigorous standards and tests with soft accountability measures and generous means of assessing schools, which led to relatively few schools being singled out as objectively poor schools. These states could claim to be implementing “world-class” standards, while giving schools plenty of time to meet those standards. As long as test scores did not translate into clearly negative judgments about many schools, and as long as suburban schools performed relatively well, states had the option of administering difficult tests.

After the Act, however, states may no longer have the luxury of using hard tests or stringent scoring systems, because doing so may cause many schools to fail to make AYP. Those schools that fail to make AYP will likely be labeled as “failing.” This label is salient and unambiguous, and it is one that will not necessarily be confined to just a few inner city schools. States thus will face tremendous pressure, from both external competition and internal politics, to avoid having that label attached to many of their schools.

To be fair, it is impossible to know whether those states that adopted rigorous tests would have pressured schools to boost achievement significantly.¹¹⁹ In addition, it is not entirely clear that replacing hard tests that few schools even try to master with easier tests that most do is a step backward. The same is true with regard to easier scoring systems: If states already were giving schools plenty of time to

¹¹⁸ Cf. FIGLIO & LUCAS, *supra* note 110, at 18–19 (documenting negative effect of bad grade for school on local housing prices).

¹¹⁹ It is worth recognizing that the increasingly ubiquitous requirement that students pass certain tests to graduate, if nothing else, would place some indirect pressure on schools to improve.

boost achievement, it is not obvious that lowering scores to make AYP will affect overall levels of performance. The unfortunate truth, however, is that we never will know what *would* have happened, simply because state systems were supplanted by the NCLBA before they had much of a chance to operate. It is quite possible that states with harder tests would have succeeded in boosting achievement on those tests and perhaps also in boosting other academic indicators, such as SAT scores, NAEP scores, or college acceptance rates. This, in turn, might have generated a real race to the top, as other states faced pressure to follow suit by implementing and enforcing rigorous standards and challenging tests.

Although we cannot say what would have happened, we can be reasonably certain what *will* happen. Because of concerns about interjurisdictional competition and internal politics, states will try hard to avoid having large numbers of schools labeled as “failing” by virtue of the NCLBA. They will do whatever they can to make their schools look good according to the new and singularly powerful proxy of AYP. This will be hard to do, at least in the long run, without making tests easier or lowering the scores needed to be considered proficient.

4. *Saved by the NAEP?*

Supporters of the Act nonetheless suggest that states will refrain from lowering their standards because of the NAEP. Recall that a sample of fourth and eighth grade students in every state will have to take the NAEP reading and math tests every other year.¹²⁰ If a state’s students perform well on state tests but poorly on the NAEP, the argument goes, the state will be embarrassed into raising its standards.¹²¹

This argument is not very persuasive. First, state tests and the NAEP need not be identical or even similar. A state whose students perform poorly on the NAEP could easily claim that it prepares students for the state tests but not the NAEP.¹²² Second, NAEP results are not reported for individual students or individual schools, but instead are reported statewide, and even then only for a couple of grade levels.¹²³ State test results, by contrast, will be reported annually for grades three through eight, and at least once in grades ten through twelve.¹²⁴ Those scores, moreover, will be reported for every

¹²⁰ NCLBA § 1111(c)(2).

¹²¹ See *supra* note 56 and accompanying text.

¹²² Cf. Olson, *supra* note 64 (noting that NAEP does not necessarily “reflect what individual states have decided they want their students to know and be able to do”).

¹²³ Reckase, *supra* note 55, at 11–12.

¹²⁴ § 1111(b)(3).

school, and for every subgroup within a school.¹²⁵ Performance on state tests, and the labels and sanctions that attend failure on those tests, surely will be more salient than performance on the NAEP. It thus seems unlikely that state and local officials—or their constituents—will be bothered by a gap between state test results and NAEP results. Consider, for example, whether a state or local official could placate parents whose children are in “failing” schools by informing them that, on average, children in their state perform relatively well on the NAEP.¹²⁶

If states do lower their academic goals, federal intervention here will have been almost perfectly perverse. In a typical race to the bottom among the states, the federal government is thought to be the solution. Federal regulation is offered as the answer to destructive competition.¹²⁷ Federal environmental law, for example, is often, though not universally, defended on such a ground.¹²⁸ Here, however, it is not at all clear that states were engaged in a race to the bottom. At the same time, it *is* fairly clear that the NCLBA could prompt one. Thus, rather than *preventing* a race to the bottom, the federal government may be *creating* one. This is doubly ironic, given that one of the chief goals of the Act is to raise the academic bar in states by pushing them to adopt challenging academic standards. The rhetoric of supporters, moreover, suggests that the Act will overcome the “soft bigotry of low expectations” and leave no child behind in the march to higher standards.¹²⁹ In reality, however, it may be more accurate to

¹²⁵ § 1111(b)(2)(C)(v).

¹²⁶ It is also useful to consider that, according to a recent study, students in most states are more likely to score at the proficient level on state tests than on the NAEP. Olson, *supra* note 64 (noting that only five states “had a smaller share of students scoring at the proficient level on their own tests than on NAEP at the 4th or 8th grade”). This suggests that state and local officials, as well as teachers, are paying more attention to state tests than the NAEP. There is little reason to expect that this trend will change with the NCLBA.

¹²⁷ See Albert Breton, *The Existence and Stability of Interjurisdictional Competition, in COMPETITION AMONG STATES AND LOCAL GOVERNMENTS*, *supra* note 80, at 37, 48–52 (arguing that intergovernmental competition is inherently unstable, but can be stabilized by federal imposition of regulation, subsidies, and regional development policies).

¹²⁸ The most prominent critic of the race-to-the-bottom justification for federal environmental law is Richard L. Revesz, Dean of New York University School of Law. For his critique of this justification and a response to the defenders of it, see Richard L. Revesz, *The Race to the Bottom and Federal Environmental Regulation: A Response to Critics*, 82 MINN. L. REV. 535 (1997), and Revesz, *supra* note 80.

¹²⁹ Press Release, U.S. Dep’t of Educ., Paige Blasts “Soft Bigotry of Low Expectations” (Mar. 12, 2003), available at <http://www.ed.gov/print/news/pressreleases/2003/03/03122003.html>; Press Release, U.S. Dep’t of Educ., Letter Released from U.S. Education Secretary Paige to State School Chiefs on Implementing No Child Left Behind Act (Oct. 23, 2002), available at <http://www.ed.gov/print/news/pressreleases/2002/10/10232002a.html>.

suggest that the Act may leave no child behind in a race to the academic bottom.

B. Promoting Segregation and Exclusion

One of the most heralded aspects of the NCLBA is the requirement that schools meet performance goals for various groups of students, including those who are poor, those for whom English is a second language, and those who are racial and ethnic minorities.¹³⁰ By disaggregating the scores of these students and holding schools responsible for their achievement, the NCLBA promises to shine a needed spotlight on the performance of traditionally disadvantaged and underperforming students. No longer will schools or school systems be able to obscure the academic performance of these groups within aggregate scores. Indeed, this aspect of the NCLBA, more than any other, implements the rhetorical commitment to “leave no child behind.”

Despite the thoroughly admirable purpose of this provision, it unintentionally promotes racial, ethnic, and socioeconomic segregation in at least two ways. The interesting question is whether these incentives to segregate will be dampened by the public school choice provisions in the NCLBA, which in theory could lead to more integration. I begin with the incentives to segregate and then turn to the question of choice. I end with a discussion of the incentives to push students out of school altogether.

1. Incentives to Segregate

The first way in which the NCLBA promotes segregation is by providing administrators of white, middle class schools a reason to exclude African American, Hispanic, and poor students. The mechanics are simple: these students traditionally do not perform as well as their white and more affluent peers on standardized tests.¹³¹ In a recent study, for example, Professors Kane and Staiger concluded that schools that contain an African American or economically disad-

¹³⁰ NCLBA § 1111(b)(2)(C)(v)(II).

¹³¹ See, e.g., Boger, *supra* note 6, at 1437–41 (discussing test score gaps in North Carolina); Larry V. Hedges & Amy Nowell, *Black-White Test Score Convergence Since 1965*, in *THE BLACK-WHITE TEST SCORE GAP* 149, 149–81 (Christopher Jencks & Meredith Phillips eds., 1998) (same); Gary Natriello & Aaron M. Pallas, *The Development and Impact of High-Stakes Testing*, in *RAISING STANDARDS OR RAISING BARRIERS? INEQUALITY AND HIGH-STAKES TESTING IN PUBLIC EDUCATION* 19, 19–38 (Gary Orfield & Mindy L. Kornhaber eds., 2001) (describing achievement gap among students from different backgrounds).

vantaged subgroup are much more likely to fail to make adequate yearly progress than those that do not.¹³²

To improve the chances that a particular school or schools within a district make AYP, administrators have an incentive to minimize the number of African American or poor students in a school or district.¹³³ Importantly, administrators need not exclude all such students. The NCLBA only requires the disaggregation of scores for a subgroup if it is sufficiently large to yield “statistically reliable information.”¹³⁴ Because there is no single formula for determining this figure, the NCLBA allows states to determine the minimum size of subgroups.¹³⁵ That number will vary, but it uniformly will be more than one.¹³⁶ Whatever the precise number is, in many places it will become the dividing line between schools that make AYP and those that do not. In Texas, for example, schools must “count” the performance of racial or ethnic subgroups if at least 10% of the students fall within the subgroup. As Kane and Staiger report, among schools that had exactly 9% Latino students—and thus did not have to disaggregate their scores—42% were rated “exemplary,” while less than 20% of schools with exactly 10% Latino students achieved that status.¹³⁷

Non-Title I schools may have an extra incentive to avoid transfer students, at least those from poor families. Recall that the real sanc-

¹³² Thomas J. Kane & Douglas O. Staiger, *Unintended Consequences of Racial Subgroup Rules*, in *NO CHILD LEFT BEHIND? THE POLITICS AND PRACTICE OF SCHOOL ACCOUNTABILITY* 152, 158 (Paul E. Peterson & Martin R. West eds., 2003).

¹³³ See Boger, *supra* note 6, at 1448–49 (recognizing this incentive). Another strategy would be to exclude these students only from testing rather than barring them from a particular school entirely. Earlier studies of state accountability systems documented that schools pursued this strategy by classifying low-income and low-performing students as disabled, which usually exempted them from testing. See, e.g., Julie Berry Cullen & Randall Reback, *Tinkering Toward Accolades: School Gaming Under a Performance Accountability System* (March 2002) (unpublished manuscript, on file with *New York University Law Review*); DAVID N. FIGLIO & LAWRENCE S. GETZLER, *ACCOUNTABILITY, ABILITY AND DISABILITY: GAMING THE SYSTEM* (Nat’l Bureau of Econ. Research, Working Paper No. 9307, 2002), available at <http://www.nber.org/papers/w9307>. Because disabled students must be tested under the NCLBA, however, and because schools must report scores for at least ninety-five percent of their students, classifying students as disabled or otherwise attempting to exclude them from testing will not be a very effective strategy. See § 1111(b)(2)(I)(ii).

¹³⁴ § 1111(b)(2)(C)(v)(II).

¹³⁵ Improving the Academic Achievement of the Disadvantaged, 34 C.F.R. § 200.7(a)(2) (2003); see also Kane & Staiger, *supra* note 132, at 158 (noting that there is no single “magical sample size” that will yield statistically reliable information, and thus states have some freedom to define the minimum sizes differently).

¹³⁶ See Olson, *supra* note 70 (describing different approaches taken by several states in setting minimum subgroup size).

¹³⁷ Kane & Staiger, *supra* note 132, at 162.

tions of the NCLBA are reserved for schools that receive Title I funds.¹³⁸ Although schools that do not receive Title I funds must in theory meet AYP and will have their test results reported, they do not face the public-choice, restructuring, or other accountability provisions that the NCLBA imposes on Title I schools. If a non-Title I school accepts Title I transfer students, however, this might convert it into a Title I school. It is unclear from the NCLBA whether this would happen automatically with even one transfer, or whether it only would happen if enough poor students transfer to bring the poverty level of the chosen school to the requisite level.¹³⁹ Either way, however, accepting transfer students creates the risk that a school once free from the stern accountability provisions of Title I would become subject to them.

While the incentive to shun certain students seems obvious, it is less clear how administrators can achieve this goal should they decide to pursue it. In some instances, the path is straightforward. A number of existing integration plans are voluntary; they are a form of structured school choice, either within or across districts.¹⁴⁰ Schools that accept transfer students who are poor or African American simply might stop doing so.¹⁴¹

Consider, for example, the programs in various metropolitan areas through which urban students, predominantly minority and poor, attend suburban schools.¹⁴² Now imagine that one of those suburban schools finds itself failing to achieve AYP in part or entirely because the transfer students do not meet their benchmark. To the extent suburban school participation was voluntary, there undoubtedly will be pressure within the district to bow out of the program. Long-standing, albeit relatively small-scale, interdistrict choice programs in places such as the Boston, Milwaukee, Rochester, and St. Louis metropolitan areas thus could be at risk.

¹³⁸ See *supra* notes 44–47 and accompanying text.

¹³⁹ Even if schools are not automatically converted to Title I schools by accepting a transfer student, in some districts it would not take many transfers to render a school eligible for Title I funding. For example, in a district in which schools with a ten percent poverty rate are eligible for Title I funding, a school with 135 students, none of whom live in poverty, would only need fifteen poor transfer students to become eligible for Title I funding (and the accompanying accountability provisions).

¹⁴⁰ See Ryan & Heise, *supra* note 104, at 2064–73 (describing various school choice plans).

¹⁴¹ A school's ability to refuse to accept transfer students will depend in part on the choice provisions of the NCLBA, discussed below. See *infra* Part II.B.2.

¹⁴² For a description of these programs, see RICHARD D. KAHLBERG, *ALL TOGETHER NOW: CREATING MIDDLE-CLASS SCHOOLS THROUGH PUBLIC SCHOOL CHOICE* 130–33, 251–54 (2000), and Ryan & Heise, *supra* note 104, at 2070–73.

Structured school choice plans *within* districts may face similar pressures. School districts in Cambridge, Massachusetts; Montclair, New Jersey; and Wake County, North Carolina, among others, sponsor choice plans that seek to create racial or socioeconomic balance within schools. Parents list their preferences among public schools and administrators assign students with an eye toward satisfying those preferences while creating racially or socioeconomically balanced schools.¹⁴³ Surely there will be some pressure in these districts to disband the programs if it turns out that most of the schools are failing to make AYP. Under that scenario, district officials may well conclude that it is better to have at least some schools meeting AYP. One strategy to accomplish this would be to return to a neighborhood school policy, which would result in greater socioeconomic segregation and give those schools higher on the socioeconomic scale a greater chance of making AYP.

For the same reason that existing integration plans may be scaled back or eliminated, it will be that much harder to begin new programs. This may ultimately prove more detrimental to integration than the elimination of the few existing programs. Rates of school segregation are already quite high, due to residential segregation and the ubiquitous policy of neighborhood school assignments. Given that court-ordered desegregation is fading from existence, the only real hope for integration in the near future is through the expansion of voluntary programs, which would generate political controversy even under the best of circumstances.¹⁴⁴ If increased diversity within a school raises the chances that the school will fail to make AYP, the already considerable political obstacles to racial and socioeconomic integration may become insurmountable. In this sense, then, the NCLBA may be even more significant in creating disincentives for schools and districts to integrate than in creating incentives to segregate.

Parents will face similar incentives, which is the second means by which the NCLBA will encourage segregation. Parents with options will be reluctant to choose schools that are failing to make AYP. In some places, this will lead those parents to shy away from more integrated schools, given that racially and socioeconomically integrated schools are more likely to fail to make AYP than predominantly or exclusively white and middle class schools.

¹⁴³ See KAHLENBERG, *supra* note 142, at 116–30; Ryan & Heise, *supra* note 104, at 2065.

¹⁴⁴ Existing voluntary programs include the Metropolitan Council for Educational Opportunity in Boston, controlled choice in Cambridge, and the socioeconomic integration program in Wake County. See KAHLENBERG, *supra* note 142, at 130–33, 251–54 and accompanying text; Ryan & Heise, *supra* note 104, at 2070–73.

Some parents will be able to act on these incentives either by choosing a particular neighborhood or choosing a particular school. In states that offer little or no public school choice, parents will have to move to the “right” neighborhood in order to place their children in middle class schools, which effectively means that exercising this form of choice will be restricted to those who can afford to live in the neighborhoods that host such schools. In states and districts that sponsor school choice, the option to select middle class schools may be more widely available, at least in theory. But if past experience is any guide, the parents who exercise this option will be disproportionately better educated and wealthier than those who do not.¹⁴⁵ If the parents who do choose schools are motivated, as suggested, to avoid schools that fail to make AYP, unfettered public school choice may help promote racial and socioeconomic segregation.

This is not to suggest that a lot of white and middle class parents currently seek out diverse schools or districts. In the past, however, some obviously did. Perhaps in the future, even more would have chosen to do so. Presumably, parents who chose integrated schools looked to a number of factors to inform their judgment as to which schools were best for their children. If more parents equate school quality with test scores, however, they may be less willing to look beyond those scores to judge the quality of a school.

Indeed, there is already some evidence to this effect. Professors Wells and Holme have studied the effect of testing on the demographics of six integrated high schools in different regions of the country.¹⁴⁶ They illustrate how parents of children in these schools have become more skeptical of the value of integration in light of the schools’ relatively poor performance on standardized tests. As a result, relatively good integrated schools, like those in Shaker Heights, Ohio; Englewood, New Jersey; and West Charlotte, North Carolina, have lost or are in the process of losing white students, as well as many middle class African American and Hispanic students.¹⁴⁷ Through interviews with former students, teachers, community members, and local officials, Wells and Holme discovered that test scores have

¹⁴⁵ JOSEPH P. VITERITTI, *CHOOSING EQUALITY: SCHOOL CHOICE, THE CONSTITUTION, AND CIVIL SOCIETY* 9 (1999); *see also* James S. Liebman, *Voice, Not Choice*, 101 *YALE L.J.* 259, 261 (1991) (theorizing that educationally oriented parents, i.e., “educational connoisseurs,” will demand higher quality schools than consumers with less exacting tastes).

¹⁴⁶ Amy Stuart Wells & Jennifer Jellison Holme, *No Accountability for Diversity: Standardized Tests and the Demise of Racially Mixed Schools*, Paper Presented at a Conference on the Resegregation of Southern Schools, Chapel Hill, N.C. (Aug. 30, 2002), available at http://www.civilrightsproject.harvard.edu/research/reseg02/wells_holme.pdf.

¹⁴⁷ *Id.* at 13.

played an important role in prompting white and middle class flight.¹⁴⁸ Whereas these integrated schools once were valued based on a number of criteria, they are now increasingly judged by their test scores alone. As one white graduate of Shaker Heights High School bluntly explained, “[i]f proficiency scores didn’t come down the Blacker the schools get, then white people wouldn’t run away from [them].”¹⁴⁹

Test scores alone do not explain why some integrated schools have lost or are losing white and middle class students, as Wells and Holme recognize. It may be that many parents who would have fled anyway point to test scores as a rationalization. At the same time, however, the current emphasis on standardized test scores has undeniably worked to narrow perceptions of what constitutes a “good” school. This narrowing will continue with the increasing importance of making AYP. To the extent integrated schools have relatively lower test scores and are less likely to make AYP, they are less likely to be judged favorably by parents. Parents with options are thus less likely to keep their children in them or to choose them in the first place. As Wells and Holme point out, reputations of integrated schools already tend to be somewhat fragile, and the emphasis on test scores will likely weaken them even further.¹⁵⁰

2. *NCLBA’s Choice Provisions*

Optimists may point to the choice provisions within the NCLBA and claim that I am spinning a Chicken Little tale. Recall that the NCLBA allows students in Title I schools that fail to make AYP for two consecutive years to attend another public school within the same district.¹⁵¹ Only schools that have made AYP are eligible to receive transfer students. If there are no such schools within the district, the NCLBA and its regulations encourage but do not require districts to arrange for students to attend school in another district.¹⁵² The NCLBA regulations also suggest that lack of space in a “good” school within the same district is not a sufficient reason to deny students their right to choose another school.¹⁵³

¹⁴⁸ *Id.* at 10 (“Overall, we see a trend in our data from school reputations based on a broad array of factors—i.e., athletics, theater, band, curriculum, diversity, national merit scholars, college acceptance of graduates, and student diversity—to much more narrowly defined reputations, based much more heavily on standardized test scores.”).

¹⁴⁹ *Id.* at 16.

¹⁵⁰ *Id.* at 18.

¹⁵¹ NCLBA § 1116(b)(1)(E).

¹⁵² § 1116(b)(11); Improving the Academic Achievement of the Disadvantaged, 34 C.F.R. § 200.44(h)(1) (2003).

¹⁵³ 34 C.F.R. § 200.44(d).

Some commentators and advocates suggest that the choice provisions could lead to greater racial and socioeconomic integration.¹⁵⁴ The argument is simple. If minority and poor students disproportionately do worse on standardized tests, Title I schools with such students will be more likely to fail to make AYP. As a result, many minority and poor students will have the option to transfer. The schools to which they transfer are more likely to be white and middle class if, again, past performance on standardized tests is any indication.¹⁵⁵ As a result, the operation of the public school choice provision in the NCLBA may promote greater racial and socioeconomic integration.

This is indeed a possibility, and for those who favor greater integration, it is a welcome one. There are reasons to be skeptical, however, that the choice provisions will play out in the way just described. To begin, it is important to recognize that interdistrict choice is not required by the NCLBA. In many metropolitan areas, segregation occurs between rather than within districts,¹⁵⁶ and in these areas the NCLBA choice provision offers little hope of promoting integration. Second, where there is diversity within a given district, space constraints will surely limit the amount of movement.¹⁵⁷ It is inconceivable that states and districts will abide by the regulation that suggests a lack of space is no excuse for failing to guarantee school choice. Saying that space is not a constraint does not make it so. To the extent districts are willing to ignore this regulation, they also may be willing to manipulate space constraints if doing so works to their advantage. That is to say, if the incentives to maintain segregation work in the way I have described, administrators of successful schools may claim that they lack much, if any, space for transfer students.¹⁵⁸

In addition, there does not yet seem to be a great demand for choice. Over 8500 schools were required to offer school choice in 2002–3,¹⁵⁹ but only a “trickle” of parents exercised this option.¹⁶⁰ Part

¹⁵⁴ See, e.g., Taylor, *supra* note 68, at 1755.

¹⁵⁵ *Id.* at 1758.

¹⁵⁶ Ryan & Heise, *supra* note 104, at 2093–96.

¹⁵⁷ See, e.g., Sam Dillon, *New Federal Law May Leave Many Rural Teachers Behind*, N.Y. TIMES, June 23, 2003, at A1 (reporting that parents in New York say promise of school choice is “hollow because there are insufficient seats in better schools to accommodate those requesting transfers”).

¹⁵⁸ Cf. Taylor, *supra* note 68, at 1758–59 (describing various ways in which local superintendents have tried to limit transfers).

¹⁵⁹ The choice requirement began immediately because the NCLBA counted failure to make AYP under the previous Title I reauthorization, the IASA, even though AYP was defined differently under the IASA. Thus, any school that had failed for two consecutive years to make AYP at the time the NCLBA was enacted had to offer school choice. NCLBA § 1116(b)(1)(E), (f)(1).

¹⁶⁰ Boger, *supra* note 6, at 1443–44.

of the explanation may be a lack of information. The choice option is a new one, and it takes time for information to filter out to the public. But this simply points to the likelihood that school officials who are not excited about choice will have opportunities to limit the flow of information and informally discourage the exercise of choice.¹⁶¹

Finally, there is a little-noticed provision in the NCLBA that makes the school choice provision contingent on state permission. The NCLBA requires schools to offer choice unless they are prohibited from doing so by state law.¹⁶² Although this might be an extreme move, it is possible that, if nothing else works, states will enact laws prohibiting school choice. Taken together, all of these obstacles make it unlikely that the NCLBA requirement of offering choice will be sufficient to overcome the strong incentives to maintain or increase racial and socioeconomic segregation.

3. *The Costs of Segregation*

A policy that promotes racial and socioeconomic segregation will be opposed by some, regardless of its effect on academic achievement. Others might argue that we should only be concerned about segregation if it has a detrimental effect on academic achievement. Even if we agree, for argument's sake, with the latter position, there is a strong argument that integration, at least along socioeconomic lines, ought to be pursued. That is, even if our only goals are to increase academic performance and to close the achievement gap, socioeconomic integration can be an effective means to achieve those ends.

Numerous studies demonstrate that predominantly poor schools typically perform much worse than do middle class schools.¹⁶³ Moving poor students to middle class schools, moreover, has been shown to boost their achievement, and to do so without threatening the achievement of their more advantaged peers.¹⁶⁴ To be sure, studies on the asymmetrical benefits of socioeconomic integration are neither uniform nor free from methodological criticism, and the evidence does not suggest that poor students who transfer to middle class schools will immediately start scoring at the proficient level on state

¹⁶¹ See Taylor, *supra* note 68, at 1759 (noting that officials in many districts "delayed in notifying parents of the transfer opportunities and little or no use was made of the media or community meetings to reach out to those least likely to be informed of transfer opportunities").

¹⁶² § 1116(b)(1)(E)(i).

¹⁶³ See, e.g., Boger, *supra* note 6, at 1417–23 (describing studies); Ryan & Heise, *supra* note 104, at 2102–05 (same).

¹⁶⁴ See Ryan & Heise, *supra* note 104, at 2106–07 (discussing evidence).

tests.¹⁶⁵ But the evidence is sufficiently strong to indicate that promoting socioeconomic stratification is not a promising strategy for boosting aggregate achievement levels or closing the achievement gap. It is also abundantly clear that, regardless of the precise benefits of socioeconomic integration, students do not perform well in schools of concentrated poverty. To the extent we wish to boost academic achievement or close the achievement gap, therefore, we should hesitate before promoting socioeconomic segregation. Yet this is precisely what the NCLBA does.

4. *Student Exclusion*

An even more serious threat to disadvantaged students is the problem of student exclusion, which the NCLBA threatens to exacerbate. All types of schools, whether elementary, middle, or high school, must make AYP. Students who perform poorly on state tests obviously hurt schools looking to make AYP. This is why schools, to the extent they can, will work to avoid enrolling those students who are at risk of failing the exams. The same pressure could lead schools to push low-performing students out, either to another school (if one can be found that will accept them) or out of the school system entirely. This temptation presumably will be strongest at the high school level, both because students most typically drop out at this stage and because low-performing high school students are most likely to be farthest behind. Given the connection between performance on tests, socioeconomic status, and race, the students most likely to be targeted for exclusion will be poor and/or racial minorities. Just as these students will suffer from any incentive to segregate created by the NCLBA, they will also suffer, even more dramatically, from any incentive to exclude them from school altogether.

The notion that high-stakes testing increases school dropouts has been debated in the academic literature for some time, and empirical studies have reached different conclusions on the question.¹⁶⁶ Nonetheless, the temptation to exclude low-performing students, enhanced by the NCLBA, can hardly be denied: One less student performing below the proficiency level increases the overall percentage of students who have hit that benchmark. A recent report on New York

¹⁶⁵ See Helen F. Ladd, *School Vouchers: A Critical View*, 16 J. ECON. PERSP. 3, 13–14 (2002) (reviewing evidence); Eric A. Hanushek et al., *Does Peer Ability Affect Student Achievement?* 2–4 (Aug. 2001) (unpublished manuscript) (discussing methodological problems with earlier studies on influence of peers), available at <http://edpro.stanford.edu/eah/papers/peers.aug01.PDF>.

¹⁶⁶ See, e.g., Margaret E. Raymond & Eric A. Hanushek, *High-Stakes Research*, EDUC. NEXT, Summer 2003, at 48 (discussing conflicting studies).

City schools, moreover, lends credence to the view that test-based accountability for schools does indeed provide an incentive to push out low-performing students.¹⁶⁷

The No Child Left Behind Act provides weak protection against this temptation. It requires that graduation rates be included as part of a school's determination of AYP, but it does not say what the rate must be, nor does it demand that the rate increase over a certain period of time.¹⁶⁸ Moreover, graduation rates can only be counted *against* a school when determining AYP. A school with poor test scores, in other words, cannot point to a relatively high graduation rate and thereby make AYP. On the other hand, a school with good test scores but low graduation rates could be at risk of failing to make AYP if the state sets a high target for graduation rates.¹⁶⁹ States thus have little incentive to establish a demanding graduation rate. The lower that rate is set, of course, the easier it is for schools to push out students.¹⁷⁰

To be fair, the NCLBA does require that information about graduation rates be made public.¹⁷¹ Disseminating this information is far from useless, but it remains to be seen whether simply publishing graduation rates will provide sufficient protection for students at risk of being pushed out. If it does not, and if dropout rates increase, the NCLBA could end up further harming those students who obviously need the most help—leaving them, quite literally, behind.

¹⁶⁷ See OFFICE OF PUB. ADVOCATE & ADVOCATES FOR CHILDREN, PUSHING OUT AT-RISK STUDENTS: AN ANALYSIS OF HIGH-SCHOOL DISCHARGE FIGURES (Nov. 21, 2002) (describing various ways in which underperforming students are pushed out of traditional public schools in New York), available at <http://www.advocatesforchildren.org/pubs/pushout-11-20-02.doc>; Tamar Lewin & Jennifer Medina, *To Cut Failure Rate, Schools Shed Students*, N.Y. TIMES, July 31, 2003, at A1 (same).

¹⁶⁸ NCLBA § 1111(b)(2)(C)(vi); see also Improving the Academic Achievement of the Disadvantaged, 34 C.F.R. § 200.19 (2003) (indicating, inter alia, that states may, but are not required to, increase goals for graduation rates).

¹⁶⁹ See NCLBA § 1111(b)(2)(D)(ii). Schools seeking to take advantage of the safe harbor provision must show, in addition to the requisite gain on test scores, some improvement in graduation rates or another “academic indicator,” such as attendance rates. § 1111(b)(2)(I)(i).

¹⁷⁰ Even if graduation rates are set at a demanding level, school districts possess administrative tricks to manipulate the rate. In New York City, for example, students who leave a school are given one of a bewildering array of classifications, which are supposed to indicate whether the student dropped out, transferred to another school, moved out of the city, et cetera. These classifications can be used to mask dropouts; students who are pushed into GED courses, for example, are often counted as transfers rather than dropouts. Lewin & Medina, *supra* note 167. Although the NCLBA regulations indicate that a student who receives a GED does not count as a high school graduate, and they warn that states “must avoid counting a dropout as a transfer,” 34 C.F.R. § 200.19(a), it remains to be seen how vigorously these regulations will be enforced.

¹⁷¹ NCLBA § 1111(b)(2)(C)(vi).

C. *Deterring Good Teachers*

Although social scientists continue to debate the relationship between various educational inputs and outputs, these researchers find common ground on one obvious point: Teachers matter.¹⁷² The better the teacher, the better the student performance, regardless of the student's background. Researchers disagree over which teacher characteristics matter the most—experience, education background, subject matter knowledge, or unquantifiable traits.¹⁷³ But they generally agree that, whatever characteristic is chosen, better teachers tend to be found in middle class schools rather than in high-poverty schools.¹⁷⁴

The larger standards and testing movement, of which the NCLBA is a part, creates two separate problems regarding teachers. First, it will make teaching a less attractive profession to some talented individuals. Second, it will bolster the tendency of good teachers to choose relatively wealthy, white, and high-achieving schools.

1. *Reducing the Supply of Teachers*

Standards- and test-based accountability rest on a certain distrust of teachers. Prompted by the belief that schools set low expectations for students, the standards movement seeks to force schools to raise the bar.¹⁷⁵ In setting uniform academic standards, state policymakers necessarily take discretion away from teachers. Local teachers, with some direction from local officials, once determined not only *how* to teach but *what* to teach. Now the state, through standards and tests,

¹⁷² Eric Hanushek, for example, who has argued strongly and consistently that there is no systematic relationship between resources and outcomes, has recently acknowledged that “the effects of teacher quality [are] substantial.” Steven G. Rivkin, Eric A. Hanushek & John F. Kain, *Teachers, Schools and Academic Achievement* 3 (Feb. 2002) (unpublished manuscript), available at http://wc.wustl.edu/workingpapers/Hanushek_Kain_Rivkin_1.pdf. He concludes “that having a high quality teacher throughout elementary school can substantially offset or even eliminate the disadvantage of low socio-economic background.” *Id.* Hanushek and his colleagues nonetheless argue that the differences among teachers are not readily measured by looking to simple, observable characteristics. For Hanushek's views on the general relationship between money and outcomes, see, for example, Eric A. Hanushek, *School Resources and Student Performance*, in *DOES MONEY MATTER?: THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS* 43 (Gary Burtless ed., 1996).

¹⁷³ See Rivkin, Hanushek & Kain, *supra* note 172, at 3 (discussing relevant traits of teachers); Lynn Olson, *The Great Divide*, *EDUC. WK.*, Jan. 9, 2003, at 9 (discussing conflicting studies).

¹⁷⁴ See Olson, *supra* note 173 (describing uniformity of evidence indicating that better teachers tend to be in low-poverty schools).

¹⁷⁵ See *supra* Part I.A.

tells teachers what to cover.¹⁷⁶ Teachers essentially retain the freedom to figure out how to teach material assigned by the state.

Restricting the autonomy of teachers, while perhaps necessary in some instances, can be counterproductive. Establishing standards and requiring periodic testing might protect students against unmotivated teachers who, given the chance, would shirk their responsibilities. At the same time, however, reducing their autonomy can make teaching less attractive to very good teachers. Those teachers who *can* be trusted to motivate and teach their students may find teaching less rewarding the more they are shackled to state standards and tests.¹⁷⁷ Protecting students against bad teachers can thus simultaneously deter good ones from entering or remaining in the profession.

England's experience is instructive. A little more than a decade ago, under Margaret Thatcher's leadership, the British government introduced a national curriculum, which described in precise detail what each child should learn in each grade. The British government also implemented a series of mandatory tests in English, math, and science for students at ages 7, 11, 14, and 16. The tests were designed as "a form of handcuffs," to make sure that teachers followed the national curriculum.¹⁷⁸

Ten years later, England is facing a severe shortage of teachers.¹⁷⁹ Other factors have contributed to the shortage, of course, but the same trend exists elsewhere. As one report observed, in countries where accountability systems have undermined teacher autonomy, like England, Canada, and Australia, there is a "recruitment crisis."¹⁸⁰ In other countries, including many in continental Europe, where teachers still have a good deal of autonomy and are generally esteemed professionals, teacher shortages are much rarer.¹⁸¹

¹⁷⁶ Cf. Donald R. McAdams, *Enemy of the Good*, EDUC. NEXT, Summer 2002, at 23 (suggesting that allowing teachers to set their own standards and develop their own performance measures has been tried and failed because "too many teachers neglected to teach the curriculum or did not teach effectively, and too many children suffered the consequences").

¹⁷⁷ See, e.g., Traub, *supra* note 105 (discussing teachers' reactions to standards and testing regimes in several states).

¹⁷⁸ Mike Baker, *Accountability vs. Autonomy*, EDUC. WK., Oct. 31, 2001, at 48.

¹⁷⁹ *Id.*; Caroline St. John-Brooks, *What They Want is a Little Respect*, TIMES EDUC. SUPPLEMENT (London), Sept. 7, 2001, at 26.

¹⁸⁰ Baker, *supra* note 178; see also Stacey Ash, *Ontario Lags Behind in Education Reform*, CAMBRIDGE REP. (Toronto), Jan. 26, 2002, at A1 (describing teacher shortages in Canada, Britain, and Australia); Chris Bunting, *Is There a Way Out of this Fine Mess?*, TIMES EDUC. SUPPLEMENT (London), Feb. 1, 2002, GET STAFFED (Special Report), at 4 (describing teacher shortages in Britain).

¹⁸¹ Baker, *supra* note 178; St. John-Brooks, *supra* note 179.

The NCLBA, along with state accountability systems, makes life even more unpleasant for teachers because it raises the stakes attached to test results. Schools with poor test scores, or even those that generally have good test scores but have one low-performing subgroup, will not make adequate yearly progress. Teachers in those schools will have to suffer the stigma of being associated with failing schools, which can limit future career opportunities.¹⁸² Teachers who remain in schools that consistently fail to make adequate yearly progress face the possibility of being fired or moved to another school.¹⁸³ In addition to these sanctions, imposed by the NCLBA, some state accountability systems also create the possibility that teachers in low-performing schools will be fired, or that they will face the dispiriting prospect of watching their colleagues receive bonuses for good test results.¹⁸⁴

Attaching consequences to failure may be necessary to provide incentives to take the tests seriously. But it raises the costs associated with failure, which may make teaching even less attractive. It also puts more pressure on teachers to make sure that their students pass the tests, which will require more test-specific preparation. Teachers may resent having to follow state standards, which offer general guidelines as to what material should be covered and which skills taught. But teaching to standards must be relatively liberating as compared to having to teach to a specific test, especially the sort of multiple-choice tests used most often by states. To the extent that spending time preparing for a single standardized test is less interesting or rewarding than is conducting a regular lesson, even one guided by state standards, test-based accountability will force teachers to spend time on tasks they would rather avoid. This will make teaching less attractive to some individuals that we would like to see in classrooms.¹⁸⁵

¹⁸² Jeff Archer, *Increasing the Odds*, EDUC. WK., Jan. 9, 2003, at 52. A recent study of the effect of North Carolina's accountability system on teachers concluded that labeling schools as low-performing exacerbated the departure of teachers from such schools. Charles T. Clotfelter et al., *Do School Accountability Systems Make it More Difficult for Low Performing Schools to Attract and Retain High Quality Teachers?* 12 (Feb. 2003) (unpublished manuscript, on file with *New York University Law Review*).

¹⁸³ NCLBA § 1116(b)(7)–(8).

¹⁸⁴ Charles T. Clotfelter & Helen F. Ladd, *Recognizing and Rewarding Success in Public Schools*, in *HOLDING SCHOOLS ACCOUNTABLE*, *supra* note 18, at 23, 46.

¹⁸⁵ In New York, for example, prior to the NCLBA, the first big tests for students came in the fourth grade. As Traub reports, one teacher in Mamaroneck reported that his students “do so well on the tests because I reinforce them every day.” But then he added, “I don’t know anyone who wants to teach fourth grade”—a sentiment that you hear repeated all the time.” Traub, *supra* note 105.

2. Reinforcing the Unequal Distribution of Good Teachers

At the very least, teaching will be less attractive in those schools where teachers must spend a great deal of time preparing for the tests. This points to the second effect of the NCLBA on teachers, which has to do with their distribution. Attaching consequences to test results creates obvious incentives for teachers to avoid schools that are likely to produce bad results. As already discussed, schools with poor and minority students are more likely to perform poorly on state tests than schools with relatively affluent, white students.¹⁸⁶ Poor and predominantly minority schools already struggle to attract and retain good teachers. Study after study documents that high-poverty and high-minority schools have less qualified and less experienced teachers.¹⁸⁷ Empirical studies also consistently indicate that, when given the chance to choose, teachers systematically move to schools with fewer poor, minority, and low-achieving students.¹⁸⁸ Exactly why more qualified and experienced teachers shy away from high-poverty, high-minority, and low-achieving schools is hard to pin down and most likely is related to a mix of factors, including working conditions, salary, student behavior, parental support, and administrative support.¹⁸⁹ All of these factors, together and singly, point teachers toward relatively high-performing, affluent, and white schools.

The NCLBA will likely reinforce the trend of good teachers exiting challenging schools. Everyone recognizes that a student's test results stem from a combination of in-school and external factors, including ability, socioeconomic status, and parental involvement.¹⁹⁰ These external factors help explain why students from wealthier families tend to do better on tests than students from poorer families: The former come to school more prepared to learn and receive more assistance outside of school than do the latter. High or low test scores in a particular school, therefore, may have less to do with the quality of

¹⁸⁶ See *supra* notes 163–165 and accompanying text; see also *supra* note 131 and accompanying text.

¹⁸⁷ See COMM. ON EDUC. FIN., NAT'L RESEARCH COUNCIL, MAKING MONEY MATTER 169 (Helen F. Ladd & Janet S. Hansen eds., 1999) (describing uniform results of various studies); Olson, *supra* note 173 (same).

¹⁸⁸ See, e.g., Eric A. Hanushek et al., *Why Public Schools Lose Teachers*, J. HUM. RESOURCES (forthcoming 2004) (manuscript at 3, 20–37, on file with *New York University Law Review*) (discussing trends in Texas).

¹⁸⁹ Jennifer Park, *Deciding Factors*, EDUC. WK., Jan. 9, 2003, at 17; Olson, *supra* note 173.

¹⁹⁰ See, e.g., ERIC A. HANUSHEK & MARGARET E. RAYMOND, LESSONS ABOUT THE DESIGN OF STATE ACCOUNTABILITY SYSTEMS 4–9 (Harvard University 2002), available at <http://edpro.stanford.edu/eah/papers/accountability.Harvard.publication%20version.pdf>.

teaching than with factors beyond the control of schools and teachers.¹⁹¹

Adequate yearly progress is pegged to uniform benchmarks of achievement. It follows that a teacher with a classroom of disadvantaged students will generally face a greater challenge than one with a classroom of relatively affluent students. At the same time, the teacher who makes remarkable gains with disadvantaged children but nonetheless fails to bring the students to the AYP benchmark (or within the safe harbor provision) will get little reward under the NCLBA.¹⁹²

Consider two different scenarios. Under the first, imagine a school that is already well above the AYP benchmark in year one. Now imagine that performance holds steady in year two or drops only slightly, so that the students are still hitting the AYP benchmark. Neither that school nor its teachers will be considered failing under the NCLBA. Now change the facts and imagine a school that is well below the AYP benchmark in year one. Suppose that this school makes strong gains in year two, but not enough to meet AYP or to bring the school within the safe harbor provision of the NCLBA. Despite larger annual gains than the first school, this school and its teachers essentially get no credit for the achievement gain.

The safest bet for a teacher, therefore, is to pick schools that are likely to succeed under the NCLBA's framework. These schools are most likely predominantly white, middle class, and located in the suburbs.¹⁹³ Given that the most experienced and talented teachers will usually have the widest array of choices, these are the teachers who will have the opportunity and incentive to choose already successful schools.¹⁹⁴ The best and most experienced teachers will thus have an added incentive to teach in schools that are already performing well.

Teaching in such schools will not only be less risky, given the stakes involved, but it may also be more enjoyable on a day-to-day basis. As already suggested, in schools whose students can easily pass state tests, teachers may not have to spend much time on test preparation. In schools with students at risk of failing the tests, by contrast, test preparation will occupy a large portion of classroom time.¹⁹⁵ If it

¹⁹¹ See, e.g., Elmore, *supra* note 8, at 37 ("Most high-performing schools simply reflect the social capital of their students (they are primarily schools with students of high socioeconomic status), rather than the internal capacity of the schools themselves.").

¹⁹² NCLBA § 1111(b)(2)(E), (I)(i).

¹⁹³ See *supra* notes 163–165 and accompanying text; see also *supra* note 131 and accompanying text.

¹⁹⁴ Boger, *supra* note 6, at 1445–46.

¹⁹⁵ See, e.g., Traub, *supra* note 105 (contrasting test preparation in Scarsdale, New York, a high SES school, with that in Mamaroneck, a lower SES school).

is reasonable to suppose that teachers—especially the more creative and innovative ones—would prefer to spend less rather than more time on test preparation, this is yet another reason for them to choose high-performing schools.

3. *Requiring “Highly Qualified” Teachers*

The NCLBA is not silent on the issue of teacher quality. It requires that all teachers in Title I schools be nominally “highly qualified.” But the criteria that make a teacher “highly qualified” for purposes of the NCLBA are not perfect, or even very good, proxies for actual quality.

For newly hired teachers to be considered “highly qualified” under the NCLBA, they must be licensed or certified by the state. Middle and high school teachers also must demonstrate subject-matter knowledge, either by having majored in the subject in college or by passing a “rigorous state test.”¹⁹⁶ All existing teachers also must be licensed or certified, and existing middle and high school teachers must also demonstrate subject-matter knowledge. They can do this either through the same means as newly hired teachers or through “a high objective state uniform standard of evaluation.”¹⁹⁷

These requirements should lead to some improvement, especially in urban schools. City students are often taught by teachers who are not certified and by those who are teaching subjects for which they have had no academic preparation.¹⁹⁸ The NCLBA should change this, assuming the supply of teachers is sufficient. But it will fall short of ensuring that all teachers are actually good or that the best teachers are in the schools that need the most help.

The basic problem is similar to the problem that plagues the NCLBA’s approach to adequate yearly progress. While requiring states to enforce nominal requirements for teachers, it leaves the substance of those requirements to the states. Moreover, although the NCLBA seems to establish two separate requirements—licensure or certification on the one hand, and demonstrating subject-matter knowledge on the other—these seemingly independent obligations can be collapsed in those states whose licensing or certification exams test for subject-matter knowledge.¹⁹⁹

Whether the NCLBA’s requirement that teachers be “highly qualified” is a meaningful proxy for quality, therefore, depends prima-

¹⁹⁶ § 9101(23); Improving the Academic Achievement of the Disadvantaged, 34 C.F.R. § 200.56 (2003).

¹⁹⁷ NCLBA § 9101(23)(C).

¹⁹⁸ Olson, *supra* note 173.

¹⁹⁹ See 34 C.F.R. § 200.56.

rily on the state's requirements for certification or licensure.²⁰⁰ Studies of these requirements do not inspire much confidence. In many states, teaching candidates must pass a state test in order to be certified or licensed. Yet the cutoff scores for passing the exams are often remarkably low, allowing even those who score in the lowest tenth percentile to qualify for a license or certification.²⁰¹ The tests themselves, moreover, are not uniformly rigorous. Indeed, four states actually use a test for licensing that twelve other states use to determine *eligibility* for teacher preparation programs.²⁰²

It is not at all clear that states have much incentive to stiffen their licensure or certification requirements. In fact, it is possible that the NCLBA may cause states to lower their standards if doing so is necessary to meet the requirement that teachers be nominally "highly qualified," especially in states facing teacher shortages.

Even if they remain unchanged, we should take little comfort in existing standards for certification or licensure. Indeed, because of their lack of rigor, commentators suggest that current licensure or certification systems generally bar only the "weakest of the weak" from entry into the profession and fall far short of guaranteeing teacher quality.²⁰³ Some argue that these systems also deter some talented individuals from entering the profession because they require time-consuming teacher preparation courses as a prerequisite to sitting for a teaching exam.²⁰⁴ Licensure and certification thus might be both insufficient and unnecessary to ensure that teachers are indeed "highly qualified."

²⁰⁰ In those states that do not test for subject-matter knowledge as part of their licensure or certification processes, new and existing teachers presumably will have to take an additional test as a result of the NCLBA. Although this separate testing requirement could in theory result in better-prepared teachers, in practice it will all depend on the rigor of the state tests, over which states themselves have complete control. If states already have low requirements for certification and licensure, it is hard to imagine states making subject-matter tests difficult, especially if they face a shortage of teachers.

²⁰¹ CTR. ON EDUC. POL'Y, *supra* note 1, at 94.

²⁰² *Id.* at 94–95.

²⁰³ SANDRA HUANG ET AL., EDUC. TRUST, INTERPRET WITH CAUTION: THE FIRST STATE TITLE II REPORTS ON THE QUALITY OF TEACHER PREPARATION PROGRAMS 6–8 (2002); Frederick Hess, *Expired Licenses*, in IMPLEMENTING THE NO CHILD LEFT BEHIND ACT 53 (National Center on Education and the Economy et al. eds., 2003), available at http://www.ppionline.org/documents/Ed_NCLB_0403.pdf.

²⁰⁴ See, e.g., Hess, *supra* note 203, at 64–66. Rural districts have also objected to the requirement that middle- and high school teachers be licensed or certified to teach in every subject. They argue that this requirement is especially onerous for rural schools, where low enrollments necessitate assigning teachers to cover a number of different classes. Rural district officials predict that they will both lose existing teachers and have an even more difficult time recruiting new ones. See Dillon, *supra* note 157.

We may grant, for argument's sake, that a teacher who meets the NCLBA criteria is more qualified than one who does not. But this does not make the former "highly qualified" in any real sense of the phrase. Indeed, given the uncertainty as to which teacher characteristics are most likely to boost student achievement, it is folly to suppose that the NCLBA criteria do any more than ensure that teachers are minimally qualified. Teachers who meet the NCLBA criteria will still vary greatly in "actual" quality—including their ability to boost academic achievement. And those who are better than others will still be able to avoid—indeed, the NCLBA will give them even more reason to avoid—challenging schools. In short, the NCLBA's explicit guarantee of "highly qualified" teachers in every school seems no match for the Act's implicit message that teachers who want to succeed should choose already successful schools.

III SHIFTING FOCUS

Relying on absolute achievement levels to assess school performance may be counterproductive. It would seem to follow that, short of abandoning test-based accountability altogether, focusing on rates of growth is the solution.²⁰⁵ It is not a panacea, of course, as I explain below; focusing on rates of growth creates its own set of problems and would face political obstacles. Nonetheless, a value-added system of accountability would provide a more accurate picture of school quality and would not generate the same perverse incentives that have been unleashed by the NCLBA. After describing the promise and limitations of a value-added system, this Part explains what opportunities exist for states to follow such an approach, both now and in the future.

A. *AYP, Value-Added Assessments, and the Accountability Dilemma*

To appreciate the promise of value-added assessments, it is important to revisit why AYP is a relatively useless measure of school quality. As mentioned above, student performance is the product of a number of factors, some of which schools can control, others of which are beyond a school's ability to influence. A student's score on a standardized test is the result of both school and teacher inputs, as well as

²⁰⁵ Again, I recognize that there are a number of strong arguments for abandoning test-based accountability altogether, and that the problems I identify here might contribute to the general case against testing. But, in this Article, I want to put that debate to one side and to assume that test-based accountability is here to stay, at least for the foreseeable future. See *supra* note 11 and accompanying text.

a host of exogenous factors, including innate ability, socioeconomic status, parental involvement, community stability, and peers.²⁰⁶ Because of the influence of these exogenous factors, looking to whether students in a school hit a uniform benchmark of achievement—the current approach to measuring AYP—actually tells us very little about the quality of the school itself.

For example, schools with relatively advantaged students typically post better test scores than those with relatively disadvantaged students. But it does not follow that the former school is better at educating its students than the latter; the scores may simply reflect the fact that the former school has “stronger” students than the latter. Put differently, it is difficult, if not impossible, to know exactly what value a school has added to students’ knowledge and abilities simply by looking at one year’s test scores.²⁰⁷ Looking at whether a school’s scores have improved from one year to the next offers no more accuracy, and perhaps less, both because exogenous factors influence scores each year and because each year a new crop of students is tested.²⁰⁸ If students in the first year of testing are relatively weak and those in the second year are relatively strong, the school’s test scores will rise, but not necessarily because of any improvement in teaching or administration.

If AYP were just an aspiration, we might not care why or how a school’s students reach the proficiency level. We would only hope that they do. But AYP is more than this. It is the basis used for sanctioning schools. Title I schools that do not meet AYP are presumed to be doing something wrong, which is why these schools must submit improvement plans, why their students eventually are allowed to choose another school, and why the schools ultimately might be subject to reconstitution or state takeover. Schools that meet AYP are left alone. To the extent AYP offers an inaccurate snapshot of school quality, the NCLBA may punish relatively effective schools and reward relatively ineffective ones.²⁰⁹

²⁰⁶ See, e.g., Clotfelter & Ladd, *supra* note 184, at 26; Hanushek & Raymond, *supra* note 98, at 201.

²⁰⁷ Social scientists, for once, seem to be in complete agreement on this point. See, e.g., Clotfelter & Ladd, *supra* note 184, at 26; Hanushek & Raymond, *supra* note 98, at 202–03; Robert H. Meyer, *Value-Added Indicators of School Performance*, in *IMPROVING AMERICA’S SCHOOLS: THE ROLE OF INCENTIVES* 197, 213–14 (Eric A. Hanushek & Dale W. Jorgenson eds., 1996); Dale Ballou, *Sizing Up Test Scores*, *EDUC. NEXT*, Summer 2002, at 10.

²⁰⁸ See Hanushek & Raymond, *supra* note 98, at 204; Helen F. Ladd & Randall P. Walsh, *Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right*, 21 *ECON. EDUC. REV.* 1, 2–4 (2002).

²⁰⁹ Cf. Robert H. Meyer, *Comments on Chapters Two, Three, and Four*, in *HOLDING SCHOOLS ACCOUNTABLE*, *supra* note 18, at 137, 139 (noting that failing to isolate school

Notice that the inaccuracies and perverse incentives produced by the current approach to AYP are intertwined. To the extent AYP may reflect the nature of the student body rather than the nature of the school, it is a poor way to measure schools. At the same time, to the extent administrators understand what influences a school's scores, they also know that they may improve those scores by attracting some types of students and excluding others. Moreover, AYP demands different levels of achievement growth from different schools, and, in particular, it requires the lowest-performing schools to post the greatest gains. Yet no attention has been paid to whether the level of growth demanded is at all feasible, nor is much credit given to schools that accomplish more than average levels of growth but still fall short of the uniform AYP benchmark or the safe harbor target.²¹⁰ If the levels of growth required are indeed not feasible, this creates the pressure, described above, to make the tests easier or to lower the scores necessary to be deemed proficient.

All of this suggests that a more appropriate basis for accountability would be one that isolates the quality of the school. This is precisely what so-called "value-added" methods of assessment attempt to do.²¹¹ Pioneered about a decade ago by Professor William Sanders, value-added assessments had been adopted by districts in twenty-one states, and Sanders's methodology still forms the basis of Tennessee's Value-Added Assessment System.²¹² Other states looked in whole or in part to growth levels in assessing schools.²¹³ Value-added assessments have also been the focus of a sizeable academic literature, most of it favorable.²¹⁴

quality in accountability systems can lead to cancellation of effective programs and continuation of those that do not work, and can also send wrong signals to parents and students about which schools are good).

²¹⁰ The Department of Education, as discussed below, has approved some state AYP plans that give schools partial credit for bringing students closer to the proficiency benchmark. See *infra* note 231 and accompanying text. In this limited way, the Act, at least as implemented by the states and the Department of Education, does give schools some credit for improving the performance of students who nonetheless fall below the proficiency benchmark.

²¹¹ Meyer, *supra* note 209, at 139.

²¹² Jeff Archer, *Sanders 101*, EDUC. WK., May 5, 1999, at 26; see Lynn Olson, *Education Scholars Finding New 'Value' in Student Test Data*, EDUC. WK., Nov. 20, 2002, at 1.

²¹³ See, e.g., Olson, *supra* note 99 (reporting that eighteen states assessed schools based on improvements over past performance).

²¹⁴ See, e.g., Hanushek & Raymond, *supra* note 98, at 205–213 (explaining why value-added approach is superior to others); Meyer, *supra* note 207, at 213–16 (same); Jay P. Greene, *The Business Model*, EDUC. NEXT, Summer 2002, at 20 (same); Anita A. Summers, *Expert Measures*, EDUC. NEXT, Summer 2002 (endorsing value-added approach). Not all academics have embraced the value-added approach; some have reservations due to various measurement and methodological problems. See, e.g., Ladd &

Although this method is fairly complex in its details, its basic approach is to focus on achievement gains over time for the same individual or groups of students.²¹⁵ The underlying supposition is that if we know how much a student's achievement has improved from one year to the next, we have a much better sense of what value the school has added to the student's academic performance. The reason is fairly simple: Exogenous factors that affect achievement will influence achievement every year that a student is tested. By focusing on gains made or lost by the same students, rather than overall levels of achievement, those exogenous factors are cancelled out.²¹⁶ At least that is the idea.

The complicating factor is that exogenous factors, like socioeconomic status, appear to affect not only overall achievement, but also rates of progress.²¹⁷ The influence may be less on the latter than the former,²¹⁸ but it still exists. Thus, students from middle class families tend both to perform better on each test and to make more progress annually than do their poorer peers. Looking just at gains, therefore, may also give a somewhat misleading picture of school effectiveness, insofar as rates of progress are influenced by factors beyond the school's control.

There are two potential responses to this problem. The first is to control directly for those exogenous factors believed to influence growth, such as the race and socioeconomic status of both individual students and the student body.²¹⁹ The Dallas school system employs this approach. It essentially establishes predicted achievement gains based on past test scores, which are themselves adjusted to take into account various exogenous factors like race and socioeconomic status. The result is that annual growth targets differ for different types of students and the schools that educate them.²²⁰

Walsh, *supra* note 208, at 3–4; Ballou, *supra* note 207. Even the critics of value-added methods, however, do not suggest that focusing on absolute achievement levels is a superior method of accountability. See, e.g., Ladd & Walsh, *supra* note 208, at 16 (emphasizing “that, as a measure of school effectiveness, gains in student performance are far superior to the alternative of relying on the average level of student achievement”).

²¹⁵ William L. Sanders et al., *The Tennessee Value-Added Assessment System: A Quantitative, Outcomes-Based Approach to Educational Assessment*, in *GRADING TEACHERS, GRADING SCHOOLS: IS STUDENT ACHIEVEMENT A VALID EVALUATION MEASURE?* 137 (Jason Millman ed., 1997).

²¹⁶ Hanushek & Raymond, *supra* note 98, at 205; Sanders et al., *supra* note 215, at 138.

²¹⁷ See Clotfelter & Ladd, *supra* note 184, at 56; Ballou, *supra* note 207.

²¹⁸ See Hanushek & Raymond, *supra* note 98, at 205 (speculating that exogenous factors have “relatively small” influence on growth rates).

²¹⁹ Meyer, *supra* note 209, at 141–43.

²²⁰ For a discussion of the Dallas approach, see Clotfelter & Ladd, *supra* note 184, at 33–37.

The second is to let individual students or groups of students essentially act as their own control. Tennessee focuses on individual students and North Carolina on groups.²²¹ In each state, growth targets are based on the past performance of individual students or groups of students. Past performance is used to establish a baseline rate of growth for an individual or group of students. The idea is that once we know the “normal” rate of growth for an individual or group of students, we can then use this baseline to establish predicted rates of growth for those individuals or groups. Schools can then be judged based on whether the actual rates of growth exceed or fall short of the predicted rates.²²²

Value-added approaches generally, and those that try to control for the influence of exogenous factors on growth in particular, are not perfect in isolating the contribution of schools to a student’s academic progress.²²³ But they nonetheless provide a more accurate portrait of a school’s quality than does the current approach to AYP. To the extent that value-added assessments help isolate school quality, they also help level the playing field for all schools by taking away the advantage of having an affluent student body or the disadvantage of having a poor one. This, in turn, would lessen or eliminate the incentive to shape the student body, because regardless of the starting point of the students, the school would be judged on the basis of what it added to the student’s knowledge and abilities. If growth targets are set with an eye toward past performance, moreover, this would reduce the incentive to dumb down tests or make scoring systems more generous.

But value-added assessments create their own problems and can generate political controversy, especially in the climate that produced the NCLBA. In attempting to isolate a school’s performance, value-added approaches generally ignore absolute levels of achievement. Because the focus is on achievement growth, schools can be judged effective even if most or all of their students are performing below average, or below an established proficiency level. Value-added systems that take into account factors like race or socioeconomic status exacerbate this problem, at least politically, because they essentially

²²¹ For a description of Tennessee’s approach, see Sanders et al., *supra* note 215. For a description of North Carolina’s approach, see Ladd & Walsh, *supra* note 208, at 6–7, and N.C. DEP’T OF PUB. INSTRUCTION, STATE BOARD OF EDUCATION, SETTING THE ANNUAL GROWTH STANDARDS: “THE FORMULA,” 1 ACCOUNTABILITY BRIEF (Jan.2002), available at <http://64.233.161.104/search?q=cache:ZX-LTJPOuGIJ:www.ncpublicschools.org/vol2/forlisa/SettingAnnuGrowthStandards.html+&hl=en>.

²²² See Ladd & Walsh, *supra* note 208, at 6–7.

²²³ See Ballou, *supra* note 207 (discussing limitations of value-added approach); Meyer, *supra* note 209, at 141–42 (same).

establish different growth targets for different types of students.²²⁴ What makes these systems potentially more accurate measures of school performance is precisely what makes them politically controversial.²²⁵

Any government interested in implementing a test-based accountability system thus faces a dilemma. Creating a “fair” system of accountability, which isolates a school’s performance, may not be enough to push all schools and students to reach a sufficient standard of achievement. It may also be politically unacceptable insofar as it tolerates different absolute rates of achievement and is therefore in tension with any rhetorical commitment to leave no child behind. On the other hand, requiring all schools and students to reach a common benchmark over the same time period may be both unfair and counterproductive, insofar as it creates incentives to lower the benchmark or take actions that will make it more difficult for some students to succeed.

There may be theoretical solutions to this dilemma, but I fail to see one, and the literature on accountability is close to silent on this point.²²⁶ The basic problem is an informational one. Because we do not have good information about what schools, teachers, and students are capable of achieving over a certain period of time, any accountability system is bound to ask for too little or too much. If it asks for too little, it may be self-limiting. If it asks for too much, it may be self-defeating, for all of the reasons I have described.

It is tempting to suppose that a hybrid system, which combines a focus on absolute targets and growth in achievement, could solve the problem of asking too much and asking too little. But this is a false hope, primarily because new students enter the school system each year. To see the problem, imagine an accountability system that looks

²²⁴ See Richard F. Elmore et al., *The New Accountability in State Education Reform: From Process to Performance*, in HOLDING SCHOOLS ACCOUNTABLE, *supra* note 18, at 65, 82, 93–94 (acknowledging political difficulties of holding students to different standards).

²²⁵ Adding to the political difficulties confronting value-added systems is their complexity. The description of these systems here is very bare bones; in reality, they require fairly sophisticated computations that will likely be lost on anyone not trained in statistics. To the extent that politicians, teachers, and parents cannot understand the basis for accountability, support for that system may diminish. Elmore et al., *supra* note 224, at 92–93; cf. Summers, *supra* note 214, at 19 (acknowledging that support could wane but arguing that value-added systems should not be rejected because they are less than fully transparent to teachers and public).

²²⁶ The one commentator I have found who does more than simply acknowledge some tension between focusing on growth and hitting uniform benchmarks of achievement suggests that it is possible to combine a value-added approach with one that focuses on absolute performance goals. Meyer, *supra* note 207, at 220. All that is required, he points out, is that schools with low-performing students must post greater gains. *Id.* Rather than resolving the dilemma I describe here, however, this “solution” simply restates it.

to past achievement growth within a state to create a more realistic time period for reaching the ultimate goal of having all students pass state tests. The time period could be set explicitly with an eye toward giving the lowest performing schools or students a reasonable period of time to reach the proficiency level. Alternatively, the time period could vary for different groups of students and schools, based on how far those students and schools are from reaching the proficiency level. To the extent that the time period over which students and schools must hit proficiency targets is more realistic, this would undoubtedly be an improvement over the current system.

But students are a moving target. Even if the goal of one hundred percent proficiency were met in a particular year, this does not guarantee that all students entering school after this point will score at the proficient level from beginning to end. If students continue to enter school unprepared and begin the testing period below grade level, one of two things must happen: Either new growth targets will have to be established and the whole cycle will have to begin anew, or schools will again be punished in part for having weak students. However appealing and pragmatic a hybrid system might seem, at some point either growth rates or absolute achievement must be the dominant basis for accountability. In the end, then, accountability systems either must focus on realistic growth targets and run the risk that they are tolerating low achievement by some or all students, or they must focus on an ambitious common benchmark and hope for a miracle.

The NCLBA is hoping for a miracle. Yet it is simultaneously decreasing the odds that this miracle will happen. By using an unrealistic goal as a basis for accountability, the Act creates incentives that make it harder to achieve that goal, especially for poor and minority students facing even lower odds of being taught by good teachers. Defenders of the Act, including the Secretary of Education, deride those who call the Act unrealistic as apologists for failure and suggest not so subtly that they are bigoted for not believing that all students can succeed.²²⁷ If the incentives play out in the way I have described, however, the NCLBA will work to ensure that many students indeed are left behind, especially those who are already disadvantaged. Moreover, given the connection between test performance and socio-economic status, the ultimate irony is that the NCLBA may reinforce the impression that disadvantaged students cannot achieve at high levels and that the vast majority of schools that teach such students are failures.

²²⁷ See *supra* note 129.

Because value-added systems do not create the same unproductive incentives, and do not paint as inaccurate a picture of school quality, on balance they seem preferable to the approach codified by the NCLBA. To be sure, they are an imperfect solution and carry the risk of asking too little of some students and schools. It is nonetheless possible to envision a system that sets a fairly aggressive target for growth, one that takes past growth rates into account but demands more than the replication of past performance. Ideally, such a system would provide rewards for schools that exceed their growth targets, and would study such schools to determine what contributed to their success and whether it could be replicated. Such a system, again, would not be perfect, but it would be much more productive than the one demanded by the language of the NCLBA.

The question remains as to how to move, politically, from the system codified by the NCLBA to a system based on achievement growth. As I explain below, this may not be as difficult as it first appears.

B. Banking on Volatility

States should begin by recognizing that education reform is notoriously beset by fads.²²⁸ Almost immediately after a new reform is introduced, supporters and opponents of the reform point to studies that “prove” its efficacy or futility. The fight among social scientists regarding the effect of high-stakes testing on student performance, for example, has already begun.²²⁹ Most often, because the reforms (predictably) fail to produce significant and uncontested improvements in a short period of time, politicians and the public lose interest, especially if another new reform is dangled in front of them, promising the impossible. Demonstrating again the perpetual triumph of hope over experience, politicians and the public often discard the “failed” reform and rush to embrace the new one.

The odds are quite good that the NCLBA is another fad. By 2014, Congress, the President, state officials, and the world will have changed. Schools and students may still be subject to standards and

²²⁸ Camilla Persson Benbow & Julian C. Stanley, *Inequity in Equity: How ‘Equity’ Can Lead to Inequity for High-Potential Students*, 2 *PSYCHOL. PUB. POL’Y & L.* 249, 268–70 (1996); James Van Patten & James Bolding, *Reform and Reconstruction in Education: Commitment or Fad*, 44 *J. EDUC. PHIL. & HIST.* 40 (1994); Larry Cuban, *Reforming Again, Again, and Again*, *EDUC. RESEARCHER*, Jan.–Feb. 1990, at 3, 3.

²²⁹ See Debra Viadero, *Researchers Debate Impact of Tests*, *EDUC. WK.*, Feb. 5, 2003, at 1 (discussing conflicting studies regarding impact of tests on student performance); Raymond & Hanushek, *supra* note 166, at 48 (criticizing methodology of study finding that high-stakes testing does little to improve achievement and suggesting that such testing does indeed improve achievement).

testing, but it is unlikely that public schools will be operating under the existing framework of the NCLBA. If states take the Act literally and operate on the assumption that it is here to stay, the perverse incentives I have described will influence their responses (and the responses of teachers and parents) to the Act.

But states might consider operating on the assumption that the NCLBA, too, shall pass. I am not proposing that states simply ignore the Act's mandates. Rather, I suggest that they take advantage of the flexibility that the Department of Education has shown with regard to state compliance plans. As mentioned above, the Department has approved roughly twenty state plans that postpone a good deal of achievement growth until the last few years of the twelve-year period, despite language in the Act that seems to prohibit such an approach.²³⁰ States ought to use this freedom to establish growth targets that are at least somewhat realistic over the next five to eight years. Ideally, those targets should be sufficiently ambitious to push schools to do more than replicate the status quo, but not so unrealistic as to create pressure to lower standards or punish schools that are demonstrating real growth—admittedly a very difficult balance to achieve.

With the Department's approval, some states, like Massachusetts, Rhode Island, Minnesota, and Vermont, have already moved in this direction. They have created a "performance index" for determining AYP, which gives schools full credit for students who hit the proficiency benchmark and partial credit for students who have moved closer to the mark.²³¹ Other states should consider following suit and also explore whether there are additional ways to incorporate value-added assessments within a proficiency index used to determine AYP. They might also explore whether the strategic use of rewards for schools that exceed their targeted rates of growth are effective and then study those successful schools in an effort to determine the cause of their success.²³²

²³⁰ See *supra* notes 69–70 and accompanying text.

²³¹ State accountability workbooks, which were submitted to the U.S. Department of Education for approval, describe each state's approach to AYP. The accountability workbooks of Massachusetts, Rhode Island, Minnesota, and Vermont, along with those from all other states, can be accessed through the Council of Chief State School Officers' website. See Council of Chief State Officers, State Accountability Workbooks, at http://www.ccsso.org/federal_programs/NCLB/1935.cfm (last updated Feb. 24, 2004); see also Olson, *supra* note 70 (noting that some state accountability plans give partial credit to schools that post achievement gains but fail to reach established absolute benchmarks); Lynn Olson, *Approval of States' ESEA Plans Suggest Flexibility*, *EDUC. WK.*, Jan. 22, 2003, at 14 (same).

²³² North Carolina employs this approach in its state accountability system. Ladd & Walsh, *supra* note 208, at 6–7.

These reforms will not solve the dilemma I described above, nor will they completely eliminate the perverse incentives created by the Act. It is probably impossible to hit upon just the right balance between pushing schools to attain higher absolute levels of achievement and assuring that schools doing a good job with the students they have are not unfairly punished. Without that balance, the perverse incentives I have described cannot be completely cabined. But surely states can improve upon the system envisioned by the literal language of the NCLBA. To the extent that the Department of Education is willing to approve plans that suppose very large gains will occur in the last few years of the twelve-year period, states ought to use that flexibility wisely. Should the NCLBA prove enduring, this strategy, as suggested earlier, will have to be abandoned. But that bridge can be crossed later. In the meantime, it is important that states do not start down a self-destructive path.

CONCLUSION

The NCLBA offers an opportunity to consider a broad array of issues within the standards-and-testing movement, as I suggested in the Introduction. It also offers an opportunity to consider an issue that has received relatively little scholarly attention, namely, the proper role of the federal government in education law and policy. Historically, that role has been fairly limited and primarily directed toward special programs, usually targeted at particular populations such as the poor or the disabled.²³³ With the NCLBA, the federal government has moved to center stage in education policy. Although this is not the place to conduct a full-blown analysis or construct a complete model, some tentative observations are possible.

In my view, the Act shows why the federal government should get off the federalism fence. In an attempt to drive education policy without intruding too greatly upon state authority, the federal government has combined regulatory stringency regarding AYP with regulatory laxity regarding the quality of standards and assessments. This will likely prove to be an unworkable compromise. The federal government instead should first determine, to the extent possible, whether state competition and internal political dynamics are sufficient to push states to establish and enforce rigorous academic standards over a reasonable period of time. As it stands, there is not yet enough empirical evidence to make a conclusion one way or the other. But in some ways, that is precisely the point: Before rushing in, the federal government ought to have a reasonable basis for concluding

²³³ See Cohen, *supra* note 73, at 75.

that states cannot be trusted to establish and work toward achieving rigorous goals.

Should it be determined that states cannot be trusted, there is no good substitute for federal control of standards and tests. The federal government can create all the monetary rewards and sanctions it likes, but if states are the sole judges of whether their standards are sufficiently rigorous, those rewards and sanctions will either be futile or counterproductive. To see this, imagine if the federal government provided monetary incentives for states to establish "safe" air pollution levels, but left it up to the states to determine which levels are safe. These incentives would either cause states to set whatever pollution levels they desire, or, worse yet, to relax existing standards so that they can be achieved more easily.

To be sure, national standards and tests may be political non-starters, despite apparent public support for them in polls.²³⁴ But they are the only real solution if states, left to their own devices, would set academic standards too low. Both the first President Bush and President Clinton appear to have grasped this, at least partially, which is presumably why each advocated (voluntary) national standards and tests.²³⁵

If it turns out that states can be trusted or that competition among them will lead to higher standards, the federal government should not interfere with state accountability systems. It might encourage the creation of standards and assessments by helping to fund their creation, but it should not dictate how to determine whether one school is failing and another succeeding. Instead, it should allow states to determine this for themselves by, for example, allowing states to define adequate yearly progress. This was the approach taken in the IASA,²³⁶ the precursor to the No Child Left Behind Act, and there simply was not enough evidence to conclude that this approach was unsuccessful. To be sure, states defined their goals differently. Some looked to absolute achievement levels, others to growth, and others to a combination of the two. But different definitions of AYP are not a problem if states are sincerely interested in boosting achievement levels. If this is the case, states will pressure

²³⁴ See Lowell C. Rose & Alec M. Gallup, *The 34th Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools*, 84 PHI DELTA KAPPAN 41, 44-45, 53 (2002) (finding that sixty-eight percent of those polled favor national test and sixty-six percent favor national curriculum).

²³⁵ See RAVITCH, *supra* note 18, at 1-6, 28-29, 138-39 (outlining proposals by Presidents Bush and Clinton and their emphasis on voluntary testing); David J. Hoff, *Strong Words Underscore National Testing Questions*, EDUC. WK., Feb. 18, 1998, at 26 (discussing debate about national tests).

²³⁶ See IASA § 6311(b)(2)(I).

themselves and their neighbors to adopt whatever accountability system and definition of AYP is the most effective in achieving the goal of better academic performance.

In addition to helping fund the creation of standards and assessments, the federal government could also assist states and localities by engaging in the one task that it can perform better than any state: redistribution. In the literature on interjurisdictional competition, discussed earlier, authors disagree as to whether competition among states promotes efficiency. They uniformly agree, however, that such competition will not likely promote equity, because it works against the redistribution of wealth. Equity, they contend, must be the federal government's responsibility, because the federal government is in a better position than the states to redistribute wealth.²³⁷

Redistribution of wealth in this context might enhance not only equity but also competition. Relatively poor jurisdictions, such as property-poor school districts or relatively poor states, are at an obvious disadvantage when competing with relatively wealthy ones. If states and local districts have the right incentives to work toward rigorous achievement goals, the federal government should limit itself to assuring that they have sufficient means to reach those goals. Additional funding obviously is not the only solution to the woes of poor states and school districts, but it is wrong to suppose it is irrelevant. And, again, if the real problem is not money but the fact that states and local districts are not ambitious enough in setting enforceable achievement goals, the federal government should set those goals itself.

I recognize that these options are not especially attractive or even plausible in today's political climate. But who would have guessed, even ten years ago, that a Republican president, with huge bipartisan support, would enact the most intrusive federal education legislation in our nation's history?

²³⁷ See *supra* notes 81–82.