

No Child Left Behind? **Assessing President Bush's** **Assessment Law**

By Gregory J. Fritzberg

Since the United States Constitution dictates that public education is a state responsibility, one could describe America's recent standards-based reform movement in, well, 50 ways. Excepting recalcitrant Iowa, perhaps, the stories are not all that different. Although Gallup polling has repeatedly demonstrated that most Americans like their local schools, we have also tended to accept the crisis claims about our system as a whole emanating from conservative pundits and a sensation-hungry press since the early 1980s. Politicians have been beholden, many quite happily, to "solving" our educational problems by legislating tougher content and performance standards. Regardless of the still-debated status of American public education today, most people support systemic efforts to articulate clearer and more ambitious learning targets for all students, but the question of who sets these standards and determines our success in fulfilling them has been more problematic.

Arguments about centrally determined "one size fits all" educational standards and assessments are

Gregory J. Fritzberg is an associate professor and chair of the Department of Curriculum and Instruction at Seattle Pacific University, Seattle, Washington.

No Child Left Behind?

certainly familiar in the state of Massachusetts. The mapping techniques—essentially, the tracing of assessment questions to relevant content standards, item by item—and analyses of difficulty employed by Robert Schwartz and his Achieve colleagues in 2001 led them to declare Massachusetts’ assessment system the best they had yet seen (2003). But Achieve is the creation of governors and corporate leaders invested in test-driven accountability, and others are not as sanguine. In a recent collection of essays entitled *Will Standards Save Public Education?* (2000), three prominent reformers (Gary Nash, Linda Nathan, and Richard Murnane) joined small-school advocate Deborah Meier in characterizing the MCAS as “an inch deep and a mile wide” (p. 7). Meier, however, goes further than many of her peers. For Meier, external standards and assessments rob local schools and families of their professional and parental prerogatives, no matter their quality. If children begin to perceive that the adults educating them are simply taking cues from “superiors,” they can lose faith in their schools, their democracy, and ultimately themselves.

As the title of the anthology referred to above suggests, Meier is not without detractors. Indeed, she is outnumbered by proponents of standards-based accountability in Massachusetts and across the nation, as existing state laws unfailingly demonstrate. Moreover, the standards and testing movement got a giant shove forward with the latest reauthorization of the Elementary and Secondary Education Act (ESEA) in January of 2002, entitled *No Child Left Behind* (NCLB). Through NCLB, President George W. Bush utilized education reform to bolster his stature as a moderate and post-Clinton New Democrats supported his federalist vision while simultaneously chiding him for under-funding it. NCLB puts teeth into the previously existing requirement—initiated through the 1994 ESEA reauthorization called the *Improving America’s Schools Act*—that states demonstrate “adequate yearly progress” toward helping all disadvantaged students achieve demanding academic standards. While it grants states receiving ESEA funds surprising latitude in creating or selecting assessment tools to measure academic achievement, NCLB is very strict about what data these tests must produce and the consequences for disappointing results. The purpose of this essay is to critique NCLB’s Title One as it relates to standardized testing and accountability. I will begin by providing some historical background.

A Short History of ESEA Title One

While the federal government had helped states, cities, and towns address educational needs from the time of its Northwest Ordinance land grants, the centrality of education reform in Lyndon Johnson’s “war on poverty” represented a completely different scale. Enhancing educational opportunities for the disadvantaged through the 1965 ESEA was Johnson’s main line of attack, and Title One was ESEA’s “crown jewel” (Jennings, 2001, p. 4), encompassing half of the bill’s funding. A former Texas schoolteacher (for one year), Johnson spoke like a true

believer: “I will never do anything in my entire life, now or in the future, that excites me more, or benefits the nation I serve more... than what we have done with this education bill” (Public Papers of the Presidents, 1966).

The “purity” of Title One funding was an issue from the beginning. All lawmakers saw the bill as a financial boost for schools serving disadvantaged persons, but not everyone was troubled that providing school-level discretion about how to use the funds would lead to some non-participants benefiting alongside qualified students (those below the poverty line, or with neither parent having a high school diploma). The purists won out, and Title One became a funding source for categorical—fiscally compartmentalized—programs which employed separate teachers who typically pulled kids out of their regular classrooms for remedial, or “compensatory,” tutoring. For the vast majority of schools that kept their books straight, the program enjoyed stability and growth throughout the late 1960s and 1970s. A substantial evaluation of Title One called the Sustaining Effects Study (SES) was undertaken between 1976 and 1979 (see Carter, 1984), which showed the overall program to be unsuccessful for severely disadvantaged children, although slightly better in reading than mathematics and during the primary rather than intermediate grades. Still, neither states, districts, or schools were directly accountable for students’ achievement. To borrow Andrew Rotherham’s (2002) pithy phrase, this was “a system of accounting, not of accountability.”

Enter Ronald Reagan, the landslide winner of what political scientist Laurence Iannaccone called the “critical realignment election” of 1980 (p. 1987, p. 62). Proponents of an expensive compensatory program showing only modest and short-term benefits for children were no match for a popular president who favored minimalist government, even if some truth was on their side. Just because a government program had not been especially effective does not mean that it could not be so, and “fading treatment effects” were nothing to be ashamed of with a population of students characterized by chronic poverty, poorly educated parents, and unequal classroom resources apart from their thirty or so minutes per day with a Title One tutor. It was the naïve expectations of the program’s originators that experience had called short, not the potential to equalize educational opportunity in a nation as wealthy as our own. Alas, Title One did survive Reagan, although it took a decade to recoup his funding cuts. But more importantly, it survived Reagan because its proponents responded to legitimate concerns about accountability, its actual ability to provide some “bang for the buck.”

While re-authorizing ESEA in 1988, Congress made some very important improvements that reflected the growing consensus (in political circles, at least) around the importance of state-wide standards and assessments. In addition to ordering a comprehensive, longitudinal study of Title One’s effectiveness with disadvantaged students now that SES data was a decade old, the federal government placed comparable evaluation responsibilities on individual states and local educational agencies (LEAs). States were required to identify specific academic

No Child Left Behind?

achievement benchmarks for schools serving Title One (then called Chapter One) students, and to identify schools that failed to make progress toward meeting these goals. States and districts were also obliged to assist unsuccessful schools until they got on track. The 1988 Hawkins-Stafford re-authorization was also the first move away from compartmentalizing Title One services so that only qualified students were in a position to benefit. For schools that served student populations three-fourths of which were economically disadvantaged, hereafter called “Title One Schools,” it now became permissible to invest in school-wide programs that educators felt would best advance their academic objectives, regardless of whether or not non-participants happened to profit from them at the same time.

Again, the enthusiasm around standards-based reform was bipartisan, and a Southern Democrat replaced George Bush Sr. as President after making his name by leading a group of governors that helped draft six national goals for American education. Bill Clinton and congressional supporters crafted *Goals 2000* around these objectives (now eight), and they provided monetary incentives for states to create content and performance standards based on examples developed at the federal level. The Clinton administration’s 1994 re-authorization of ESEA, the *Improving America’s Schools Act* (IASA), modified the previous law in three significant ways. First, the idea of separate standards for Title One students was forsaken. It was now expected that Title One students tackle the same academic content as their more advantaged peers. Second, the move toward school-wide rather than targeted initiatives was extended by the decision to drop the percentage of poor students required for Title One School-status from 75% to 50%, a significant shift that brought many new schools into the fold. Finally, although the idea of districts and schools making “adequate yearly progress” (AYP) was present in the Hawkins-Stafford bill, this was the first time the language had been used. The notion of AYP would create quite a stir, but not for several years.

The AYP expectations created through IASA raised little havoc because they had no teeth. The accountability language was intentionally vague: States were required to define AYP in a way that held districts and schools accountable for “continuous and substantial yearly improvement” toward all students reaching proficiency. Although states were now expected to define “proficiency” for Title One students the same way they did for all others, there was already great variability in the rigor of their statewide standards and assessments, which posed obvious problems for a nationally administered program. A dozen states told federal officials that they expected more than 90% of their Title One students to demonstrate proficiency on statewide assessments, while almost just as many simply hoped that half of their students would be successful (Center on Educational Policy, 2003). This surprising variability of expectations was compounded by the fact that states could define AYP in terms of either a statewide figure, an improvement figure relative to individual districts’ and schools’ own previous performance, their performance specific to reducing the achievement gap between mainstream and

non-mainstream groups, or some combination of these methods. AYP figures were also impacted by vastly different timelines that states were allowed to set for districts and schools to achieve “complete” success, which ranged from 6 to 20 years. The resulting range in percentages of districts and schools within various states that were “identified for improvement” was as wide as the approaches states took to the project in the first place. Arkansas and Wyoming did not identify any failing schools, and Texas identified only 1%, while on the high end Michigan identified 76% and the District of Columbia identified 80% of their Title One schools as sub-standard (Center on Educational Policy, 2003). And I am only referring here to states that complied with IASA’s reporting demands, a mere 17 states. The majority of states negotiated various sorts of waivers, and they still received their allotted funds.

Highlights of the *No Child Left Behind Act*

With strong bipartisan support (87-10 in the Senate), President George W. Bush signed *No Child Left Behind* into law on January 8, 2002, thus re-authorizing for six more years Johnson’s historic ESEA initiative. The most efficient way to introduce Title One of NCLB is to say that it aimed to raise yet again the accountability bar, and also to close the loopholes that had sabotaged its two previous versions. While IASA in 1994 was presented in the wake of *Goals 2000* and the relative novelty of standards-based reform, American educators have since gained familiarity with the idea of standards. Despite the protests of modern-day progressives like Deborah Meier, most educational leaders are focused today on how accountability might help them finally dent the chronic achievement gap between disproportionately poor minorities and their white, middle-class counterparts. The aggressive nature of the new Title One requirements both reflect and enhance the moral and political passions that surround this issue. I will limit my discussion of NCLB’s Title One to its assessment implications, particularly the mechanics of AYP, which is easily the most high-profile piece of the legislation. Other relevant provisions of the new law will likely arise when I make the turn from explanation to critique.

Before proceeding to AYP issues, two broader and very significant changes to previous ESEA laws deserve mention. First, NCLB rejects past notions of “purity” regarding the law’s reach by softening distinctions between service for Title One students and everyone else. As the conduits of federal Title One funds, states are now directed to hold all districts and schools accountable for the performance of their overall student bodies. The states already do this, and overall school performance is more familiar to districts and schools than any of the other subcategories prescribed by the law, but they have done so without the powerful leverage of Title One money. Second, while states are already doing quite a bit of standardized testing, less than half of them formally assess student-learning in each of grades 3 through 8, which NCLB requires to be in place by the 2005-06 school year for both

No Child Left Behind?

reading/language arts and mathematics, along with a high school science assessment two years after that. In my own state of Washington, several years into a new assessment scheme built around 4th, 7th, and 10th grade tests, education officials are already scrambling to fill in the holes.

Returning to AYP issues, the architects of NCLB learned some lessons from the previous law. They have removed the ambiguity about how districts and states are to measure AYP. Using 2002-03 data, state officials were directed to establish a baseline figure capturing the proportion of students meeting standards (“proficient”) at the *higher* of either the 20th percentile school or the lowest performing subgroup defined by race, ethnicity, poverty, English-language learner (ELL) status, or disability (P.L. 107-110). Due to the severity of gaps between any of these subgroups and more mainstream students, educational experts are confident that the proportion of students meeting reading/language arts and Mathematics standards (analyzed separately) at the 20th percentile schools will function as the de facto baseline figure against which AYP will be determined (see Kane and Staiger, 2002). NCLB also removes ambiguity about the timeline for progress, ultimate goal, and “day of reckoning” for districts and schools. By 2013-2014, *all* American students should be proficient in reading/language arts, mathematics, science, and likely other subjects, and schools are required to improve in at least a linear fashion toward that end (1/12 of the distance between the state’s baseline percentage of proficient students and full proficiency every year, although initial goals can be set two years out, three years apart after that, and performance can be calculated by means of three-year rolling averages). Given the increased clarity of AYP measures, states have no excuse for neglecting to make the federal government and the general public aware of all schools “identified for improvement” and to provide the necessary assistance with federal money allotted for that purpose. And, last but not least, the Department of Education has gone to great lengths to inform state leaders that the era of easy waivers is over.

Easily the most challenging aspect of NCLB is its attempt to leverage reduction of the achievement gap by requiring disaggregated AYP data for each of the subgroups of students mentioned above: racial and ethnic minorities, English-language learners, and poor and disabled students. But the issue is not disaggregated data per se; seventeen states were disaggregating data into similar categories prior to NCLB. States generally use one of two strategies to evaluate the performance of particular sub-groups. Some states, like California, set a statewide standard for *growth* that applies to each of the sub-groups and the overall population, as opposed to setting proficiency requirements in absolute terms for populations that, on average, rarely start from the same place. Other states, like Bush’s own Texas, have opted for absolute cut-off levels (Kane & Staiger, 2002). NCLB legislates the latter approach, and states employing the relative growth strategy will have to change, as will the majority of states that have not separated data out by sub-groups at all.

What are the stakes attached to the data schools are required to collect for each

of the sub-groups present in sufficient numbers and for the students as a whole? For a school to be recognized as making adequate yearly progress, not only its mean achievement scores for grades 3-8 reading/language arts and mathematics must meet the state's AYP target (along with graduation rates for high schools and a non-test related indicator of choice for elementary and middle schools), but *every* sub-group must meet the same AYP figure for *both* types of tests. Schools that fail to demonstrate AYP for any two consecutive years become "identified for improvement" and must inform students that they are free to attend another public school in the area, with their transportation costs covered. Schools that miss AYP for any three consecutive years must maintain the choice option, as well as pay for supplemental tutoring services provided by local providers. For schools that fail four years running, the district must increase the intensity of its assistance by choosing at least one action from a list of wholesale changes that include installing a completely new curriculum model, replacing the staff, or decreasing the authority of building-level leadership. For schools in their fifth consecutive year without AYP success, continued restructuring, conversion to a charter school, or takeover by the state become necessary (P.L. 107-110). Again, NCLB aims to correct for the laxity of the Hawkins-Stafford era, and districts and schools are feeling the heat.

Theoretical and Technical Problems With the Law

NCLB is "breathtakingly ambitious," to quote Lawrence Hardy (2002, p. 21), in some ways as ambitious as the inaugural version of ESEA. Despite classic Democratic versus Republican ideological differences, there are significant parallels concerning the way the two Presidents presented the law. I described Johnson's enthusiasm in the introduction, and the fanfare surrounding NCLB was no less dramatic, or melodramatic to be more precise: "[NCLB] is the cornerstone of my administration," gushed Bush. "These reforms express my deep belief in our public schools and their mission to build the mind and character of every child, from every background, in every part of America" (quoted in the NCLB Executive Summary). Johnson was a Democrat, and not surprisingly committed more resources to ESEA (adjusting for inflation) than did the Bush administration. But before we heroize Johnson, it must be remembered that education was a cheaper response to poverty than other alternatives. When officials in the Johnson administration formulated poverty policy, they considered more aggressive welfare reforms like income redistribution and national health insurance politically unfeasible and turned instead to education as their major tool of social reform. Johnson did sincerely believe that education was a weapon against poverty, but his Horace Mann-esque rhetoric about the sufficiency of schooling for social mobility undoubtedly had another purpose: it aimed to conceal his compromise with business interests regarding the scope of the welfare state (Kantor & Lowe, 1995).

Still, Bush's rhetoric about eliminating the achievement gap is even more

No Child Left Behind?

unrealistic than Johnson's. While contemporary policy-makers are considerably less naïve about short-term change than Johnson's peers, the current administration fails to commit the necessary resources. Education reform is a cheaper response to poverty than more directly redistributive measures, but it is not as cheap as Bush is pretending. I want to deal with the money problem last, however, because while I believe it is true that the federal resources committed thus far grossly underestimate the magnitude of the edict, readers might dismiss it as mere whining in the absence of other substantive critique. Rotherham's remarks are representative in this regard:

The critics' alternative to the accountability plan is to keep the federal dollars flowing regardless of the results. They have little to offer beyond tired bromides about needing more money for capacity building, innovative partnerships, and a host of other buzzwords that make no difference in the lives of children who attend failing schools. (p. 2)

My first line of critique addresses the very essence of the NCLB plan, namely its assumptions about the impact of high stakes tests on educators' productivity and student performance. No one anywhere near the mainstream decries the articulation of clear learning standards. While Deborah Meier criticizes their formulation by centralized bureaucracies, she takes pains to describe how explicit standards have driven the curriculum in the schools she has led. Yet, noted researcher Frederick Mosteller—whose career has combined incisive critiques of weak educational research with strong affirmations of particular empirical studies that invested sufficient time at the design stage, most recently celebrating Tennessee's Project STAR (Student/Teacher Achievement Ratio; Word, et al., 1994)—was on a team of scholars that sought in vain for any solid empirical proof of a causal connection between standards-based reform and student achievement (Nave, Miech, & Mosteller, 2000). Thus, American educators are about ten years into a movement that is only in its nascent stage of exploration by scholars. This is not to discredit the movement, but we should be modest about it given the fact that it is largely driven by intuition rather than research.

However, NCLB pushes a particular type of standards-based reform that - while it is so common today it appears almost synonymous with it—is much more problematic. The NCLB standards-based strategy emphasizes centrally determined and frequently administered high-stakes tests. The advocates of high-stakes testing believe that the prospect of public praise or shame is the most effective motivator for large bureaucratic institutions such as our state-run schools. There is probably some truth to this, but we also have to ask questions about the kinds of teaching and learning different kinds of high-stakes tests will encourage. Since states have great flexibility under NCLB to select their own tests so long as they are testing frequently enough, those that utilize "basic skills" tests (less expensive to develop and score) will inadvertently narrow the focus of instruction as teachers understandably priori-

tize activities that prepare students for success on these tests. It is easy to see how high-level learning gets slighted in this process, and student motivation and creativity languishes as a result (see Madaus & Clarke, 2001). Indeed, Linda McNeil and Angela Valenzuela (2001) have documented this very phenomenon related to the Texas Assessment of Academic Skills (TAAS) in Bush's home state in recent years.

Instruction tailored to particular tests do help students score better on these tests, as many states' data after implementing standards-based reform has shown. For states that have invested in rigorous tests that measure cognitive and communicative competencies that transcend basic skills, perhaps "teaching to the test" is not all bad. Yet, Audrey Amrein and David Berliner (2002) have raised interesting questions about what they call the "transferability" of the learning gains measured by individual state tests, which of course begs the question of whether it is really learning at all. Amrein and Berliner examined trends on our most prominent standardized tests since the standards-based reform took off in the early 1990s. Most of the test data they examined—the Scholastic Aptitude Test (SAT), the American College Test (ACT), and the Advanced Placement (AP) tests—related mainly to secondary level achievement. Their discussion of the National Assessment of Educational Progress (NAEP) provides the best information about the transferability of learning during the first eight years of school. There is no consistent correlation between having high stakes tests in the elementary and middle grades and students' NAEP mathematics and reading scores as 4th and 8th graders. What did positively effect states' NAEP scores were liberal exclusion policies, allowing students who were likely to perform poorly to avoid the exam, which biased scores upward.

I will discuss NAEP again below, but the point here is that if high-stakes testing and public disclosure alone could help states improve student achievement, consistent with the assumption that all that districts and schools really need is either a "good scare" or the potential for glory, one would expect to see a rise in NAEP scores following the implementation of high stakes, independent of confounding effects related to exclusion rates. A lot of states have had high stakes tests for many years now, and NAEP effects have been consistently absent across the nation (Amrein and Berliner, 2002). Before leaving the topic of the general effectiveness of high-stakes testing reform strategies, I wish to clarify that my worry is that NCLB advocates seem to think that testing *alone* will bring about success, especially if one examines the budget appropriations that have followed the law. I would personally support high-stakes tests (in combination with other factors) in situations where comprehensive organizational and pedagogical changes are enacted simultaneously.

The high-stakes testing approach characterizing NCLB is not only ineffective apart from more substantial investments in public education, it is also highly inequitable. NCLB's approach to holding states accountable for raising the proficiency rates of their student populations relies mainly on states' *own* tests as the measures of their success or failure. Given the considerable variability in the

No Child Left Behind?

general rigor of various states' tests, and even more crucially the variability in where states have set the cut-off score that demarcates "proficiency" from its absence, states with easier tests are going to look better on paper than states that have set the bar higher. The percentages of 8th grade students certified by their states as meeting standards in reading in 2001 ranges from 27% in Maryland to 91% in Texas. The corresponding range in 8th grade mathematics results in 2001 is from 31% in Massachusetts (recall the discussion about the MCAS in the introduction) to 92% in, again, Texas (Linn, Baker, & Betebenner, 2002). The reality is that the Texas Assessment of Academic Skills (TAAS) success that Bush has taken credit for might be more the result of a basic skills focus than any uniquely effective school reform efforts (see McNeil & Valenzuela, 2001).

The problem with such significant variability across states in how stringent they are about identifying students as proficient is that it directly impacts their AYP determinations, to which NCLB attaches numerous consequences that I outlined above. The math is pretty simple. Recall that NCLB instructs states to calculate the annual gains necessary to move from their present proficiency figures to 100% proficiency in 12 years. States like Maryland and Massachusetts have to increase the proportion of their students achieving proficiency by more than 5% each year, while states like Texas do not even have to gain a full percent. Education Secretary Rod Paige recently told state officials that to lower their standards was "not worthy of a great country" (quoted in Center on Educational Policy, 2003), but the illogical nature of the NCLB's AYP calculations pushes states to do just that, especially if there is insufficient federal assistance for the costly tasks of assisting "failing" schools. Whether Paige admits it or not, the AYP provisions create short-term incentives for certain states to water down the rigorous standards they had previously set for themselves.

A couple of caveats are necessary here. First, Texas officials are actually making their tests more difficult now, instead of languishing where they are at. However, this is probably because they are lamenting their national reputation for being soft on standards, rather than being due to any policy consequences arising from NCLB. Second, NCLB references the valuable data available from NAEP—drawn from biennial samples of 4th and 8th graders in each state who take its reading and mathematics sections—although it does not spell out explicitly how it can be used to compare the rigor of individual state's standards. In fact, NAEP data from 1990 to 2000 validates the concerns I mentioned above. If roughly nine-tenths of Texas schoolchildren meet state standards on the basis of their performance on the TAAS and only one quarter to one half (depending on whether we are talking about reading or mathematics tests) of Maryland students meet their state's standards, one would expect Texas to significantly outperform Maryland on the NAEP. But Maryland has repeatedly outsourced Texas in terms of students demonstrating proficiency on the 8th grade mathematics section, although Texas has always been close. NAEP's purpose for NCLB is to bring inconsistencies like this to light, but again there are

no clear policy consequences of NAEP data, and thus the incentive for states to “aim low” to enhance their chances of making AYP in the short term remains.

Another major problem with NCLB, beyond the variability in rigor that results from states being allowed to judge themselves—assessing achievement by their own tests—arises from overstating what these tests can tell us in the first place. In other words, is AYP totally about a school’s progress in terms of teaching and learning, or are the calculations contaminated by statistical problems? In a paper presented at a 2002 conference in Massachusetts and summarized recently in an essay called “Randomly Accountable,” Thomas Kane and Douglas Staiger (Jeffrey Geppert assisted them with the shorter essay) reveal how much statistical noise is associated with measuring AYP. NCLB employs a “cross cohort” model to determine if states are making AYP in particular grades from one year to the next. One year’s third graders are the next year’s fourth graders; the grade-specific tests assess different children each year. Conclusions about a school’s performance on this basis are clouded by sample error. In a study of 300,000 3rd-5th graders in North Carolina, Kane and Staiger determined that sample error combined with random occurrences during test times (such as a disruptive student or a barking dog) accounted for about *three quarters* of the variance in test scores in successive years, a bit more for small schools and a bit less for large schools. In other words, given that the average American elementary school has 68 kids in each grade (Kane, Staiger, & Geppert, 2002), a set of 4th graders in one year is not necessarily like the group that arrives the following year, and these variations are largely out of educators’ control.

There is some flexibility in NCLB for states to deal with measurement error. The Center on Educational Policy (2003) suggests that states should compute a figure that statisticians call standard error of proportion (SEP), or more commonly, a “confidence interval.” My own state of Washington has taken this advice, allowing districts and schools to add a SEP figure to their actual percentage of students meeting standard to arrive at the final number they compare with the state’s AYP threshold. NCLB also allows schools to employ three year rolling averages, which will reduce statistical noise. Yet, the sheer magnitude of Kane’s, Staiger’s, and Geppert’s estimation of measurement imprecision—confirmed in a recent study by the Center for Research on Evaluation, Standards, and Student Testing (CRESST; Linn & Haug, 2002)—still calls into question a federal reform plan that pushes states to place so much emphasis on single tests. After reviewing their study prior to publication, David Grissmer at the Rand Corporation pulled no punches: “The question is, are we picking out lucky schools or good schools, and unlucky schools or bad schools? The answer is, we’re picking out lucky and unlucky schools” (Olson, 2001).

The use of confidence intervals is even more critical for all of the sub-group calculations that must also demonstrate AYP for a particular school to be deemed effective. Again, NCLB is somewhat flexible—some would just say ambiguous—about the numerical point at which groups of African American, Hispanic, Native American, Pacific Islander, economically disadvantaged, disabled, or ELL students

No Child Left Behind?

become large enough to merit holding a school accountable for their AYP. The states' answers to this question have ranged from a low of 20 students to a high of 70, with at least 95% of students in any sub-group required to participate in testing (all non-participating students beyond the 5% maximum are assumed to be below standard). These determinations involve a sticky trade-off. If a state sets the minimum number of students that can comprise a particular sub-group too low, then measurement errors are likely so large as to make the computations for sub-groups close to the minimum rather meaningless (or at least difficult to explain to the public), yet still high-stakes since to be judged successful schools must meet AYP not only overall but for every single sub-group in both reading and mathematics. If the minimum number of students comprising a sub-group is set too high, on the other hand, then a state effectively excuses from accountability all schools that are serving disadvantaged students in small enough numbers to escape their holding sub-group status.

Unlike Washington state, California's effort to implement sub-group accountability through cash rewards to especially successful schools ignored issues of measurement error until *Orange County Register* reporters called them on it. Kane and Staiger summarized California's sub-group rules as bluntly as Grissmer's remarks concerning AYP overall:

California's subgroup (sic) rules are analogous to a system that makes every school flip a coin once for each subgroup, and then gives cash awards only to schools that get a 'heads' on *every* flip (original emphasis). Schools with more subgroups must flip the coin more times and, therefore, are put at a purely statistical disadvantage relative to schools with fewer subgroups. (p. 15)

As California's experience illustrates, one of the inequities surrounding AYP measurement ambiguities associated with sub-groups is that diverse schools are especially vulnerable to the NCLB mandate that schools must meet AYP for every sub-group. These schools are being asked to reduce the achievement gap – meeting the statewide AYP figure or reducing the proportion of students not meeting standard by 10% annually (see the “safe harbor” provision; P.L. 107-110) – for every sub-group simultaneously. As Kane and Staiger noted, there is definitely a perverse incentive for districts and schools to avoid integration at play here, which runs counter to our historic aspirations for our common schools. The fewer a school's sub-groups, the fewer its chances for failure.

Amendments to NCLB and Alternatives For the Future

The American public expects and deserves state and federal governments to hold our public schools accountable for the quality of education they provide for our children. While a positive step in that direction, the vague accountability language in the 1994 IASA re-authorization of ESEA described in this essay made that task difficult, and there was a real need for NCLB to clarify these ambiguities.

However, with *No Child Left Behind* the federal government has overstepped its bounds and taken ESEA badly off course. Their decision to apply AYP requirements and corresponding sanctions to all students in all schools expands unnecessarily the federal government's historic focus on socio-economically disadvantaged children. Given its vast scope, the unrealistic and inequitable goals are very problematic. NCLB goals are unrealistic because in the absence of comprehensive and significant federal investment in disadvantaged children, academic success for 100% of our disadvantaged children is fantastical. A twelve year time frame helps cloud this fact, but even for Bush's father's *America 2000* goals, the day of reckoning eventually came for the National Educational Goals about universal school readiness, literacy, and safe schools. NCLB goals are inequitable because the sub-group rules concerning AYP are disproportionately difficult for diverse schools, especially given the statistical measurement challenges that have yet to be worked out.

Title One accountability provisions should return to an exclusive focus on the socio-economically disadvantaged, and the problem of states essentially judging their own performance has to be solved. The suggestion by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) that NAEP serve as the primary evaluation tool makes sense, as does their suggestion that we initially define success in terms of the percentage of students at NAEP's "basic" level. The consensus of those that have studied NAEP's cut-off levels for distinguishing "proficient" from "basic" (not only CRESST, but also the General Accounting Office, the National Academy of Science, and the National Academy of Education) is that NAEP's proficiency bar is set very high relative to other standardized tests. In no state has the proportion of students attaining NAEP proficiency in reading and mathematics reached even *one-half* (Linn, et al., 2002), and significantly increasing the numbers of disadvantaged students who attain the basic level is a more attainable goal. Even better than adjusting our NAEP expectations, though, would be to create a new examination solely for Title One that balances attainability with rigor, demanding higher-level thinking on the part of students while maintaining cut-off levels that will keep states motivated to improve because they know success is at least conceivable. The new examination might take a "value added" approach rather than NCLB's current cross cohort approach; a value added approach looks at how groups of students grow during their years in a school, as opposed to regularly evaluating achievement at particular grade levels, even though the students in these grades change every year. Most states employ the cross cohort approach even in the absence of NCLB regulations, so there might be reasons that federal officials have done likewise. At the very least, however, federal evaluators need to cleanse the current AYP system from contamination associated with sample error.

In a 1999 book called *In the Shadow of "Excellence": Recovering a Vision of Educational Opportunity For All* (see also Fritzberg, 2000), I wrote about the achievement gap between whites and non-whites and the middle-class and the poor, the very problem NCLB addresses. Like most state-level school officials (see Center

No Child Left Behind?

on Educational Policy, 2003), I support NCLB's mandate that we pay explicit attention to how particular student populations are faring in school. However, tests and sanctions—apart from more sufficient support for states to help the coming onslaught of “failing” schools to improve—will not solve anything. In Gerald Bracey's twelfth *Phi Delta Kappan* report “on the condition of public education” in America (2002), he articulates a conspiracy theory that accuses the Bush administration of establishing an accountability system so stringent that they know huge numbers of districts and schools will repeatedly fail, thus ushering in like a “Trojan horse” (Bracey's language) renewed, and now legitimated calls for vouchers. Some unnamed state school officials said the very same thing (Center on Educational Policy, 2003). I choose not to weigh in on the federal motives behind NCLB, but I do see NCLB as a significant step backward in the national government's contribution to education reform. Although the final version of the 1994 *Goals 2000* initiative—passed during the same year as the previous ESEA re-authorization—disappointed me, the conversations that took place in Congress prior to its passing were the right ones. Encouraged by a commissioned report about the standards movement from the National Council on Educational Standards and Testing (NCEST), many congressional Democrats urged the government to require that states develop “opportunity-to-learn” standards that would instill some integrity into their calls for content and performance standards and corresponding assessments. If the mantra that “all children can learn” is to have any credibility, they argued, states must be held accountable for the systems that serve them, more specifically the quality of facilities, curriculum, and teachers that students in both wealthy and poor communities should receive. President Clinton's (Hillary's, to be exact) health care legislation had not come to fruition and he needed to pass a domestic bill, so he ultimately bailed on his demands for opportunity-to-learn standards, but the debates preceding the final outcome were important.

NCLB's results-oriented approach to reducing the achievement gap reflects a change in our understanding of equal educational opportunity that occurred at the same time as the inaugural ESEA. After James Coleman's famous report to the U.S. Civil Rights Commission called *Equality of Educational Opportunity* in 1966, he and others that followed him began to speak of equal educational opportunity not in terms of provision and access, but in terms of concrete results (see also Coleman, 1968). The idea was that the gaping achievement gap was actually *prima facie* evidence of our lack of commitment to equal opportunity, and that if radicals like Richard Herrnstein (1971) were wrong about race-based differences in mean intelligence, then we could judge our progress on equal educational opportunity in the future by actual reductions in the achievement gap. Like other laws before it, NCLB will not solve the problem of unequal academic achievement across ethnic groups, but its attention to sub-group performance will keep the issue on the table. But again, resolving the achievement gap demands more knowledge than we have now and more investment

Gregory J. Fritzberg

than we have made thus far. As for the knowledge piece, Richard Elmore's (2003) blunt assessment of our current ability to fix our schools is worth noting:

The premise that educators know what to do and all they need are the correct incentives to do it is essentially wrong. *Some* educators know what to do; *most* don't. *Some* are able to learn what to do on their own; *most* are not. . . . The main lesson of the reform movement thus far is that increasing performance in schools is complex and difficult work—much more difficult than simply changing policy. (original emphasis, p. 28)

Elmore is probably right about school reform at the national level, which lends weight to Maris Vinovskis' (2003) complaint that federal policy-makers have under-invested in research and development since the very beginning of the excellence movement. At the building level, however, Robert Slavin's tireless work on how to best invest school-wide Title One funds—whether it be his own Success For All program or another comprehensive reform plan—offers some hope (see Slavin, 2001). In other words, it is bringing effective Title One reform *to scale* that remains elusive, not an absence of alternatives for action in our local schools.

I have saved the investment—read money—issue for last because it is the easiest for critics of public schools to dismiss as self-interested excuse-making on the part of educators. But it is imperative that the federal government do more to help beleaguered states implement NCLB mandates in the context of a temperamental economy. William Mathis (2003) has examined 10 different state-sponsored studies of how much money it might cost to help all students meet their standards, as NCLB will eventually require. Of the nine studies that addressed annual per-pupil costs, 6 of them estimated required increases of between 30% and 46%, one study arrived at an estimate of 24%, and two studies ended up with 15% figures. Moreover, eight of the ten studies acknowledged that their reliance on traditional cost-projection methodologies likely under-estimated the costs for remedial (Title One) students by about *one-half*. Total spending on American public education in 2000 was about \$423 billion (*Digest of Education Statistics*, 2001). Assuming a conservative rise in nationwide costs associated with NCLB of 20%, we are looking at an \$84.5 billion increase to truly assist the plethora of failing schools; assuming a 35% increase in costs, the additional amount required is \$148 billion. Do these figures seem inflated? Consider that North Carolina estimates that a full 60% of their schools will fail to meet NCLB standards, Vermont estimates 80% over a three year period, and Louisiana is preparing for a stunning 85% failure rate (Fletcher, 2003). NCLB's \$18 billion authorization for Title One was already weak, but the Administration's most recent budget request asks for a third less than is authorized, and Bush has called his \$12.3 billion figure "more than enough money." With the National Governors Association recently estimating that the 50 states together will face a \$58 billion budget deficit due to recent economic woes (Mathis, 2003), the difference between various states' projections of their own needs and the federal contribution call Bush's remarks into serious question.

No Child Left Behind?

“Money is not the answer to everything, but it is a pretty good indication of the nation’s priorities,” Senator Edward Kennedy said in 2001 as the more expensive New Democrats’ proposal for NCLB funding called the “Three R’s” plan was being debated (Robelen, 2001). As for the “old” Democrats, Senator Christopher Dodd of Connecticut and Representative George Miller of California re-introduced in early 2003 a significantly more expensive answer to NCLB called the “Comprehensive Act to Leave No Child Behind” (S. 448/H.R. 936). The Dodd-Miller proposal was backed by the prominent Children’s Defense Fund (CDF) but had no real chance of success in the current Congress. Still, the text of the bill was important because it outlined a much more comprehensive (as it said) and holistic approach to educating children in our country, explicitly recognizing and building upon natural connections between public education and relevant social services that serve Title One children. In the CDF’s thorough explanation of the “alternative” NCLB (2002), they revealed the ironic way in which NCLB borrowed the organization’s motto but failed to live up to its implications. Perhaps the more authentic Dodd-Miller version, or something like it, will slowly gain support in Congress as the political winds change in the coming years.

Conclusion

In 2003, President Bush asked Congress for \$87 billion to help America “rebuild” Iraq, and efforts to improve our national security at home have also been expensive. Americans are in the midst of a feverish national effort to protect ourselves in a newly menacing global context. As a citizen and a parent, I do not wish to belittle such an endeavor, but we have some work to do at home as well. However, while financial concerns got the last word in this essay, I hope that it does not dominate the larger picture of NCLB’s Title One presented here. My purpose was to address Title One’s historical aspirations, its present manifestation in NCLB, its technical shortcomings, and, yes, its resources for success. Almost 4 decades old, Title One remains an almost sacred national commitment to enhancing the lives of under-privileged children, and it is incumbent on the present generation of politicians and educators to do it justice. Or, better said, to do *them* justice.

References

- Amrein, A.L. & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved from <http://epaa.asu.edu/epaa/v10n18/>.
- Bracey, G. W. (2002). The 12th Bracey report on the condition of public education. *Phi Delta Kappan*, 84(2), 135-150.
- Carter, L.F. (1984). The sustaining effects of compensatory and elementary education. *Educational Researcher*, 13(7), 4-13.
- Center on Educational Policy. (2003). State and federal efforts to implement the No Child Left

Gregory J. Fritzberg

- Behind act. Retrieved from http://www.cep-dc.org/pubs/nclb_full_report_jan2003/nclb_full_report_jan2003.htm
- Children's Defense Fund. (2002). *The state of children in America's union: A 2002 action guide to leave no child behind*. Washington, DC.
- Coleman, James. (1968). The concept of equality of educational opportunity. Equal educational opportunity: A special issue of the *Harvard Educational Review*.
- Elmore, R. F. (2003). Change and improvement in educational reform. In Gordon, D. T. (Ed.), *A nation reformed?: American education 20 years after a nation at risk* (pp. 23-38). Cambridge, MA: Harvard Education Press.
- Fritzberg, G. J. (1999). *In the shadow of "excellence": Recovering a vision of educational opportunity for all*. San Francisco: Caddo Gap Press.
- Fritzberg, G. J. (2000). Equal educational opportunity versus "excellence": Keeping the pressure on Goliath. *Educational Foundations*, 14(2), 51-71.
- Goldsmith, S. S. (2002). MCAS circa 2001: As good as it should be? *Journal of education*, 183(1), 68-94.
- Hardy, L. (2002). A new federal role. *American School Board Journal*, 189(9), 20-24.
- Heubert, J. P. (2001). High-stakes testing and civil rights: Standards of appropriate test use and a strategy for enforcing them. In Orfield, G., & Kornhaber, M. L. (Eds.), *Raising standards or raising barriers?* (pp. 179-194). New York: The Century Foundation.
- Iannaccone, Laurence. (1987). From equity to excellence: Political context and dynamics. In Boyd, W. L., & Kerchner, C. T. (Eds.), *The Politics of Excellence and Choice in Education*. New York: The Falmer Press.
- Jennings, J. (2001). Title I: Its legislative history and its promise. In Borman, G. D., Stringfield, S. C., & Slavin, R. E. (Eds.), *Title I: Compensatory education at the crossroads* (pp. 1-24). Mahwah, NJ: Lawrence Erlbaum Associates.
- Jennings, J. (2002). Knocking on your door. *American School Board Journal*, 189(9), 25-27.
- Kane, T. J., & Staiger, D. O. (2002, June). *Racial sub-group rules in school accountability systems*. Paper presented at the "Taking Account of Accountability: Assessing Politics and Policy" Conference, Cambridge, MA.
- Kane, T. J., Staiger, D. O., & Geppert, J. (2002). Randomly accountable. *Education Next*. Retrieved September 22, 2003 from <http://www.educationnext.org/20021/56.html>
- Kantor, H., & Lowe, R. (1995). Class, race, and the emergence of federal education policy." *Educational Researcher*. 24(3), 4-11, 21.
- Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind act of 2001. *Educational Researcher*, 31(6), 3-16.
- Madaus, G., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In Orfield, G., & Kornhaber, M. L. (Eds.), *Raising standards or raising barriers?* (pp. 85-106). New York: The Century Foundation.
- Mathis, W. J. (2003). No child left behind: Costs and benefits. *Phi Delta Kappan*, 84(9), 679-686.
- McNeil, L., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In Orfield, G., & Kornhaber, M. L. (Eds.), *Raising standards or raising barriers?* (pp. 127-150). New York: The Century Foundation.

No Child Left Behind?

- Meier, D. (2000). Educating a democracy. In Meier, D., *Will standards save public education?* (pp. 3-31). Boston: Beacon Press.
- Murnane, R. J. (2000). The case for standards. In Meier, D., *Will standards save public education?* (pp. 57-63). Boston: Beacon Press.
- Nash, G. B. (2000). Expert opinion. In Meier, D., *Will standards save public education?* (pp. 44-49). Boston: Beacon Press.
- Nathan, L. (2000). Habits of mind. In Meier, D., *Will standards save public education?* (pp. 50-56). Boston: Beacon Press.
- National Center For Educational Statistics. (2001). *Digest of education statistics*. Washington DC: U.S. Department of Education.
- Nave, B., Miech, E., & Mosteller, F. (2000). A lapse in standards: Linking standards-based reform with student achievement. *Phi Delta Kappan*, 82(2), 128-132.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).
- Olson, L. (2001, May). Study questions reliability of single-year test-score gains. *Education Week on the Web*, 20(37). Retrieved from <http://www.edweek.org/ew/ewstory.cfm?slug=37brookings.h20&keywords=single%2Dyear>
- Rotherham, A. (2002). A new partnership. *Education Next*. Retrieved September 22, 2003 from <http://www.educationnext.org/20021/36.html>
- Slavin, R. E. (2001). How Title One can be become the engine of reform in America's schools. In Borman, G. D., Stringfield, S. C., & Slavin, R. E. (Eds.), *Title I: Compensatory education at the crossroads* (pp. 235-260). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schwartz, R. B. (2003). The emerging state leadership role in education reform: Notes of a participant-observer. In Gordon, D. T. (Ed.), *A nation reformed?: American education 20 years after a nation at risk* (pp. 131-151). Cambridge, MA: Harvard Education Press.
- U.S. Department of Education. The No Child Left Behind act of 2001: executive summary. (n.d.). Retrieved May 15, 2003 from <http://www.ed.gov/offices/OESE/esea/exec-sum.html>
- Vinovskis, M. A. (2003). Missed opportunities: Why the federal response to a nation at risk was inadequate. In Gordon, D. T. (Ed.), *A nation reformed?: American education 20 years after a nation at risk* (pp. 115-130). Cambridge, MA: Harvard Education Press.
- Word, E., Johnston, J., Bain, H.P., Fulton, B.D., Zaharias, J.B., Achilles, C.M., Lintz, M.N., Folger, J., & Breda, C. The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report 1985-1990. Nashville, TN: Tennessee State Department of Education, 1994.