

THE EVIDENCE SUGGESTS OTHERWISE | November 2007

THE PANGLOSS INDEX:

How States Game the No Child Left Behind Act

By Kevin Carey

ACKNOWLEDGEMENTS

This publication was made possible by a grant from Carnegie Corporation of New York. The statements made and views expressed are solely the responsibility of the author.

ABOUT THE AUTHOR

KEVIN CAREY is the Research and Policy Manager at Education Sector.

ABOUT EDUCATION SECTOR

Education Sector is an independent education think tank. We are nonprofit and nonpartisan, both a dependable source of sound thinking on policy and an honest broker of evidence in key education debates. We produce original research and policy analysis and promote outstanding work by the nation's most respected education analysts.

© Copyright 2007 Education Sector. All rights reserved.

1201 Connecticut Ave., N.W., Suite 850, Washington, D.C. 20036
202.552.2840 • www.educationsector.org

When policymakers in the White House and Congress wrote the No Child Left Behind Act in 2001, they undoubtedly had places like Birmingham, Alabama, in mind. Nearly half a century after the bombings and protests that helped launch the civil rights movement, Birmingham City Schools was a textbook case of urban education in decline. The district was hemorrhaging students and funding, forcing painful layoffs and the closure of often-crumbling schools. Test scores were among the lowest in the state, particularly for poor black children.

Sadly, not much has changed in the last six years. Less than 40 percent of Birmingham students graduate from high school on time, according to *Education Week*.¹ Test scores still lag the rest of the state; there are still large achievement gaps between black and white children; and the student body and budget continue to shrink every year. For the students who remain, most of whom are black and poor, “the promised land of racial justice” described by Dr. Martin Luther King Jr. from the Birmingham jail must seem very far away.

But you wouldn’t know it by asking the Alabama Department of Education. It says everything is fine, that Birmingham City Schools made “adequate yearly progress” last year under the federal No Child Left Behind Act (NCLB). And only five of the district’s 65 schools are “in need of improvement.” The serious consequences and strong interventions that NCLB’s authors envisioned for chronically underperforming districts like Birmingham are nowhere to be found.

The reason is simple: While NCLB was designed to *raise* achievement standards every year until 2014, when 100 percent of students are required to be “proficient,” the Alabama Department of Education has *lowered* standards annually, to the point where even abjectly failing districts like Birmingham make the grade. And it’s not alone—every one of the accountability-avoidance gambits used in Alabama has been adopted in many other states. Indeed, the most noteworthy thing about Alabama’s elaborate plan to avoid NCLB accountability, and the impact of those actions on Birmingham, is how mundane they really are. Similar stories could be written about states and districts across the nation.

Collectively, these states and districts provide a case study in how determined states can undermine even tightly constructed laws like NCLB. And, as importantly, they provide a cautionary tale for members of Congress working to write the next version of the nation’s most important education law.

The Pangloss Index

A 2006 Education Sector report, *Hot Air: How States Inflate Their Educational Progress Under NCLB*, included a new way of measuring state compliance with NCLB based on data from reports submitted by the states to the federal government.² The report included a composite index of state rankings, based on measures such as student test scores in elementary, middle, and high schools, the percent of schools and districts making “adequate yearly progress” or “AYP,” high school graduation and dropout rates, school violence ratings, and teacher qualifications.³ The top-ranked states reported the best results—the fewest high school dropouts, the most highly qualified teachers, the fewest schools to miss AYP, etc.

Ideally, this index would have shown which states were doing the best job educating their students. And some of the numbers made sense—the lowest ranked state, the District of Columbia, has long been recognized as one of the worst-managed school systems in the country.⁴ Affluent Connecticut, which routinely ranks high on the federal National Assessment of Educational Progress (NAEP), was near the top of the list.

But, on the whole, the index was less indicative of which states were actually doing well than which states had simply chosen to *define* themselves as doing well. Mediocre states made themselves out to be great, and low-performing states snuck into the Top 10, while a number of objectively higher-performing states ranked near the bottom. States with big cities wracked by gang violence reported no dangerous schools; states with serious dropout problems alleged that nearly all students were graduating from high school; states with the nation’s lowest test scores reported the highest. Thus, the list was dubbed “The Pangloss Index,” after the character in Voltaire’s *Candide* who insisted—in the face of all evidence to the contrary—that we live in the best of all possible worlds.

This report includes an updated Pangloss Index, based on a new round of state reports submitted in 2007. As Table 1 shows, many states look about the same—Wisconsin and Iowa are tied for first, distinguishing themselves by insisting that their states house a pair of educational utopias along the upper Mississippi River. In contrast, Massachusetts—which is the highest-performing state in the country according to the NAEP—continues to hold itself to far tougher standards than most, showing up at 46th, near the bottom of the list.

A few states, however, changed their Pangloss Index ranking substantially from 2006 to 2007, and none increased its position more than Alabama. It jumped from

Table 1. The Pangloss Index

State	2007	2006	Change
Iowa	1	2	1
Wisconsin	1	1	0
Nebraska	3	4	1
Kansas	4	6	2
Alabama	5	22	17
Connecticut	6	3	-3
Virginia	7	12	5
Tennessee	8	11	3
North Dakota	9	10	1
South Dakota	10	5	-5
Indiana	11	8	-3
West Virginia	12	7	-5
New Jersey	13	17	4
Montana	14	16	2
Idaho	15	9	-6
Vermont	16	13	-3
Michigan	17	29	12
Texas	18	19	1
Oklahoma	19	13	-6
Illinois	20	26	6
Georgia	21	29	8
Minnesota	22	18	-4
Colorado	23	19	-4
Mississippi	24	13	-11
Ohio	25	24	-1
Pennsylvania	26	28	2

State	2007	2006	Change
New Hampshire	27	23	-4
Maine	27	21	-6
Utah	29	33	4
Washington	30	33	3
Delaware	31	24	-7
North Carolina	32	31	-1
Wyoming	33	42	9
Louisiana	34	38	4
Arkansas	35	32	-3
Arizona	36	35	-1
Rhode Island	37	27	-10
Nevada	38	47	9
California	39	44	5
Maryland	40	49	9
Alaska	41	43	2
South Carolina	42	46	4
Oregon	43	41	-2
New York	44	37	-7
Missouri	45	40	-5
Massachusetts	46	39	-7
Kentucky	47	36	-11
Florida	48	45	-3
New Mexico	49	48	-1
Hawaii	50	51	1
District of Columbia	51	50	-1

22nd to 5th, and a close analysis of the data shows why: a huge increase in the number of schools and districts making AYP. In one year, the percent of schools making the grade increased from half to almost 90 percent, while the number of districts went from two-thirds to 100 percent.

This didn't happen because Alabama students learned much more in 2006 than they did in 2005.⁵ It happened because the state of Alabama, with the full approval of the U.S. Department of Education, manipulated the statistical underpinnings of NCLB in such a way that no other result was possible. Alabama was not exceptional in this regard; many other states did exactly the same thing. Alabama's actions—and the impact those actions have had on the lives of Birmingham schoolchildren—are just one example of a broad national trend.

To understand how a law that has been widely denounced as unduly harsh was transformed into a law that is often comically lenient, it's necessary to start with the early days of NCLB.

School Year 2001–02

Taking Care of Old Business

President Bush signed NCLB into law on January 8, 2002. The basic principles are straightforward: Each state creates common academic standards for all students in reading and math and then tests students once a year to see if the standards are being achieved. Schools and districts that fall short are publicly identified and required to implement a series of reforms that get tougher the longer underperformance persists. The achievement bar gets higher every year, until 2014, when 100 percent of students are required to be “proficient.”

But mindful of the strong historical role states have played in education, Congress gave states a great deal of latitude in implementing the law. States decide what students need to know, how to test that knowledge, and what score on the test counts as “proficient.” These and myriad other decisions created a major administrative challenge for state departments of education, particularly in states like Alabama, which still hadn't gotten around to fully implementing the previous version of NCLB.

That law, the Improving America's Schools Act (IASA), was enacted under President Clinton in 1994. It was, in many ways, a more radical departure from previous federal education policy than NCLB, requiring states to develop standards, adopt tests, and hold schools accountable for the results. But IASA lacked strong penalties for non-compliance and was weakly enforced by the Clinton administration. While some states created effective accountability systems, others—like Alabama—treated the law not like a mandate but a suggestion, one they declined to take.

When the much-tougher NCLB was enacted, Alabama had to scramble to shore up its standards and move from “norm-referenced” tests (which measure student performance relative to other students) to “criterion-referenced” tests (which measure performance relative to fixed standards). Recognizing the logistical challenges this involved, the U.S. Department of Education gave Alabama permission to use its old, IASA-era accountability system for the upcoming 2002–03 school year.

By that measure, Birmingham City Schools was a mess. For 2001–02, the state gave the district a grade of “D-minus” in elementary school writing, a “D” on the ACT college entrance test, and a “D” on the high school graduation exam. Poor and minority students lagged behind their peers on state tests, while district scores on the norm-referenced Stanford Achievement Test were consistently in the lower percentiles. Two hundred twenty-one students were suspended or expelled for weapons-related incidents in a single year.⁶ Even as Birmingham racked up \$18.5 million worth of undocumented debts, the district was paying its executives a total of \$2.8 million per year, more than Mobile, Montgomery, and Huntsville (which together educate three times as many students) combined.⁷ Despite spending more money per pupil than any of Alabama's other large school systems, Birmingham recorded the lowest test scores.⁸

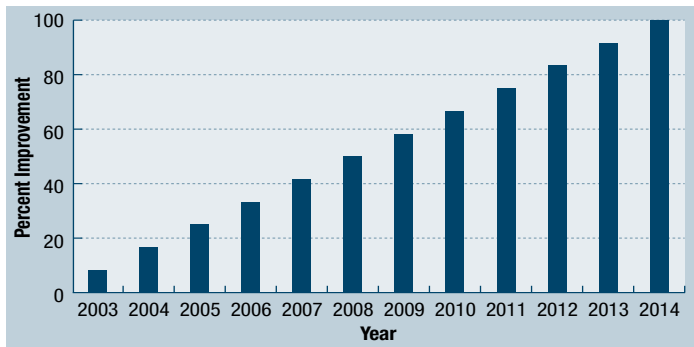
At the time, it would have been natural to assume that the new, tougher NCLB would ratchet up pressure on Birmingham to turn around these sorry results. Instead, a series of actions by the state that were endorsed by the U.S. Department of Education had exactly the opposite effect, masking the city's deep-seated academic problems and reducing the impetus for local education leaders to improve education on behalf of Birmingham students.

School Year 2002–03

Laying the Foundation

States had a number of important decisions to make in 2002 as they established their NCLB accountability plans. In addition to refining their standards and tests, they had to decide what trajectory of improvement they would require their schools to follow. NCLB seemed clear on this point, requiring states to set escalating performance targets, or “annual measurable objectives,” in “equal increments” from 2002 to 2014.⁹ The most obvious way to do this was to start with the current level of school performance, subtract that from the 100 percent target and then divide that figure into equal increments for each year. (See Figure 1.)

Figure 1. Standard Improvement Trajectory



This method spreads out the needed improvement over time, and it was adopted by a number of states. But the U.S. Department of Education decided that “equal” didn’t necessarily mean “the same,” allowing states to require different amounts of improvement in different years. Alabama, along with a number of other states, quickly settled on a trajectory that required schools to make only one-third of their improvement in the first eight years of NCLB. Then—much like the adjustable rate mortgages currently contributing to a wave of foreclosures and bankruptcies nationwide—schools would suddenly have to accomplish two-thirds of their improvement in the final four years. (See Figure 2.)

This “adjustable rate” trajectory seems reckless at first glance, a plan that would lull schools into a false sense of progress only to hit them with hugely difficult performance targets in later years. But states like Alabama knew that history was on their side. If tradition held, Congress would re-write NCLB six or seven years after its original passage.

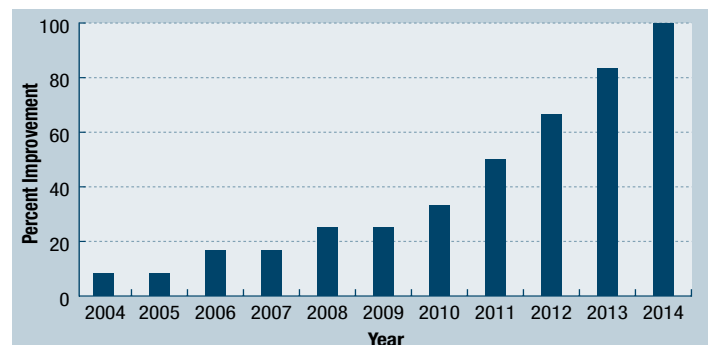
Unlike luckless homeowners, states would have a chance to renegotiate for better terms before the eight-year adjustable rate bill came due. And that’s precisely what occurred. Congress began work on NCLB reauthorization in 2007, with many education officials publicly calling for an extension of the 2014 deadline.

The second major implementation decision dealt with the way states held schools accountable for the performance of disadvantaged students. NCLB was deliberately written to focus resources and attention on closing the historically large “achievement gap” separating low-income and minority students from their more-affluent and white peers. Accordingly, the law held that schools would only be deemed as having made AYP if both the student body as a whole and *every group of disadvantaged students*, including minority and low-income students, students with disabilities, and students with limited English proficiency, met the same performance target. The idea was to help the “subgroups” of students who traditionally have been left behind.

Many schools only have a few students in some subgroups. States argued—with some justification—that test results from very small subgroups should not be used to judge an entire school. The U.S. Department of Education agreed, noting that NCLB required the use of “statistically valid and reliable” measures. The department allowed states to establish a “minimum subgroup size” below which scores would not be counted.

Alabama jumped at the opportunity, establishing a minimum size of 40. (Most states chose a smaller number.¹⁰) In a relatively homogenous state like Alabama (over 95 percent of students are black or white) where many schools and districts are small, this effectively eliminated smaller groups from being counted as a

Figure 2. Adjustable Rate Improvement Trajectory



subgroup under NCLB. Over 80 percent of Alabama schools educate Hispanic students, for example, but only 8.7 percent enroll 40 or more.¹¹

Alabama also included in its initial NCLB accountability plan a third element: “uniform averaging.” NCLB gives states the option to “average data from the [current school year] with data from one or two school years immediately preceding that school year,” in order to smooth out random year-to-year variations and better gauge long-term trends.¹² Alabama adopted this approach—with a twist. Schools would be rated by *either* their current year score *or* the three-year average, whichever was greater. This created a double standard: If a school experienced a one-year drop in performance, it would be buoyed by previous higher results. If, on the other hand, the same school had a one-year increase, that result would stand alone.

Alabama’s use of the “uniform averaging” provision illustrates several critical points about state implementation of NCLB. From the beginning, states were aggressively looking for novel interpretations of the statute’s more obscure provisions, stretching and, at times, breaking the clear meaning of the law. Those interpretations *always* made it easier, not harder, for schools to make AYP. And the U.S. Department of Education was a willing partner in this process. In a July 1, 2003 letter approving Alabama’s initial accountability plan (which included the adjustable rate trajectory, the large minimum subgroup size, and the lenient uniform averaging procedure), Assistant Secretary of Education Gene Hickok said, “I join Secretary [of Education Rod] Paige in congratulating you on Alabama’s commitment to holding schools accountable for the achievement of *all* students.”

Had that commitment been real, it would have been welcome in Birmingham. The district’s 2002–03 report card—the last under the old IASA accountability system—was once again full of bad news. Enrollment was down 1,100 students from the previous year. Accountability letter grades of “C” and “D” abounded. Scores on the Stanford Achievement Test averaged less than the 40th percentile, causing the state to designate Birmingham as a “watch with priority” district.¹³

One local observer, a former teacher and school administrator, said at the time: “The sobering reality

is that an entire generation of children has been abandoned. They have been left behind in academic failure, in violence, and in a cycle of moral and civil irresponsibility.”¹⁴ In 2003, the continual loss of enrollment and funding created a new \$14 million deficit, forcing the district to lay off 555 employees and shutter nine schools. In the years afterward, the abandoned buildings and rusting playgrounds would serve as a constant reminder to the community of how far the school district had fallen.¹⁵

School Year 2003–04

Double Standards and Statistics

Fall 2003 began the first full year of NCLB implementation in Alabama. The accountability plan was established and the new criterion-referenced tests were in place. Given the previous dismal results, Birmingham had few reasons for optimism.

Yet, when the results of the new test were released the following summer, they didn’t look so bad: 63 percent of fourth-graders were proficient in reading, 64 percent in math. High school scores were even better: 79 percent proficient in reading, 61 percent in math. And that was low compared to other districts: Statewide, 76 percent of fourth-graders and 86 percent of 11th-graders passed the reading test.

Had Birmingham defied the odds and turned its struggling schools around? Unfortunately, it had not. Alabama had simply adopted an unusually easy test. A comparison of Alabama’s state test results to results from the federal NAEP test makes this clear. Only 23 percent of Alabama fourth-graders were “proficient” on the 2003 NAEP test of reading. While the NAEP proficiency standard is generally recognized as being quite rigorous, only 52 percent of Alabama fourth-graders met the lower, “basic” standard in the same year. (See Table 2.) Alabama was counting as “proficient” a large number of students who couldn’t even meet the lowest federal achievement standard.

In Birmingham, other non-state-created tests suggested that student performance was in decline. Even as 95 percent of Birmingham 12th-graders passed the state’s high school graduation exam in reading, language, math, science, and social studies, the district’s average

score on the ACT dropped from 18.1 in 2003 to 17.4 in 2004 (roughly equivalent to a combined score of 850 on the “old” SAT, which is significantly less than the national average of about 1020. Eighty percent of all students score about 850). In a vestige of the old IASA accountability system, this earned a grade of “D-minus.”

In a state where one in four adults are functionally illiterate, where poverty is high, per-capita income low, and average educational attainment near the bottom nationwide, Alabama education officials were essentially saying “All is well! Out of the gate, we’re three-quarters of the way toward 100 percent proficiency.”¹⁶

Table 2. 2003–2004 Alabama State Test Results vs. the National Assessment of Educational Progress

	Percent Proficient (NAEP)	Percent Basic (NAEP)	Percent Proficient (State)
4th Grade Reading	23%	52%	76%

These lax standards would not, however, have been enough to keep Birmingham off of the NCLB list of low-performing districts. The law’s authors realized that test difficulty would vary dramatically from state to state. (Nearby South Carolina, for example, reported that only 33 percent of fourth-graders were proficient in reading in 2003, even though it had *higher* NAEP scores than Alabama.¹⁷) So they decreed that the first year’s performance target in each state would be based on the actual score of the school that ranked in the 20th percentile in the previous year (in other words, the school that was outperformed by exactly 80 percent of all other schools statewide).¹⁸ That meant that states like Alabama, with relatively high pass rates, would have correspondingly higher performance targets to meet, putting states on a level playing field in terms of the relative difficulty of their initial performance targets.

The 20th percentile rule meant that roughly 80 percent of schools would start above the bar in terms of their *overall* scores.¹⁹ But these schools also had to meet the same standards for their disadvantaged—and thus frequently underperforming—students. If the system had worked as planned, the subgroup accountability provision would mean that many schools and districts would be identified as failing to meet NCLB goals.

Understanding this, the Alabama Department of Education was already hard at work on creating new ways to let more schools and districts off the hook. Near the beginning of NCLB, there’s a short provision which says that “If significant changes are made to a State’s [accountability] plan, such as adoption of new State academic content standards and State student achievement standards, new academic assessments, or a *new definition of adequate yearly progress* (emphasis added), such information shall be submitted to the Secretary [of Education].”²⁰ The U.S. Department of Education rightly decided that such changes would also have to be approved by the secretary.

But, crucially, Congress had imposed no limit on how often states could request changes or how many changes they could make. This turned out to be a serious mistake. Instead of waiting to see how their accountability systems actually worked and adjusting accordingly, many states started to pepper the U.S. Department of Education with proposed amendments to the definition of AYP almost the minute their original plans were approved.²¹ Alabama was no exception.

Alabama’s first request was one of the most significant: It asked to use “confidence intervals” in measuring school performance. Confidence intervals are a “plus or minus” band around a statistical result, like when a public opinion poll is said to be accurate to within plus or minus a few percentage points. It was unorthodox to apply confidence intervals to NCLB results. Confidence intervals are used in polling because polls are based on a sample, or smaller representative group, of the larger population. There’s always a chance that when pollsters call several hundred adults at random, an unusually large number will favor one candidate or another. The confidence interval represents this uncertainty inherent to sampling. It’s standard practice in polling to use a “95-percent” confidence interval, which means there’s a 95 percent chance that the actual opinion of the entire population is within the plus-or-minus band around the result.

NCLB doesn’t use samples. It tests the entire student population in every school. (In fact, schools automatically miss AYP if fewer than 95 percent of students take the test.) There are no issues of sampling error to account for. Nonetheless, states like Alabama made the strange argument that a given school’s test results were indeed a sample, a subset of all the students who possibly could

have attended the school that year, or (alternatively, depending on who you ask) all the students who have attended or will ever attend the school, from the day it opened until the end of time. Alabama also used a 99-percent confidence interval, not 95-percent. To be 99 percent sure that a school's test results represent the mythical larger population, the plus-or-minus band has to be unusually large.

Predictably, the state applied the confidence interval in a way that was biased toward reducing the number of schools missing AYP: If a school's score fell below the performance target but within the confidence interval, it passed. A state that wanted to be really sure—say, 99 percent sure—that schools were truly making progress could just as easily apply the confidence interval in the opposite direction, disqualifying not only all schools below the performance target but also those just *above* the target, but within the confidence interval, on the grounds that their scores were too close to the cutoff for comfort. But no state would ever do this; state amendments to NCLB accountability plans are *always* designed to make it easier for schools to demonstrate success, not harder.

Many testing experts were appalled by the confidence interval idea. Stanford professor David Rogosa called it a “scam” and an example of the “if it could be, it is” approach to measuring school performance.²² W. James Popham, emeritus professor in the UCLA Graduate School of Education and Information Studies, called them a “statistical sleight of hand,” saying:

“Despite outraged assertions from a number of qualified statisticians that applying confidence intervals in AYP calculations is flat-out wrong, federal officials have nonetheless allowed states to use confidence intervals—most likely to limit the number of schools that would otherwise take an AYP nosedive.”²³

Still, the U.S. Department of Education approved the amendment, with some enthusiasm. In fact, it prodded Alabama to go even further. In a May 20, 2004, letter to Alabama, Assistant Secretary of Education Ray Simon said:

“States have great flexibility in the design of their systems and implementation of particular NCLB provisions. If, as you implement your accountability plan, you find additional

elements of your plan that you believe should be refined or amended for next school year to best serve the needs of your students and schools, I encourage you to explore all the areas of flexibility available to your State.”

Alabama didn't need the encouragement. There was much more “flexibility”—read, “leniency”—to come.

School Year 2004–05

Learning From Other States

Birmingham City Schools didn't make AYP in the 2003–04 school year. In large part, this was because it missed badly on the one measure the state hadn't (yet) watered down: participation rates. To ensure accurate results—and to prevent schools from keeping low-achieving students home on testing day—NCLB requires 95 percent of students, both schoolwide and in each subgroup, to take each test. While Birmingham's white students met this standard in 2003, its black and Hispanic students did not. Because it was the first year of the new test, moreover, the uniform averaging provision had yet to kick in. And because the district is relatively large, enrolling over 30,000 students, the minimum subgroup size and confidence intervals weren't enough at the time to offset some truly dismal results.²⁴ Students with disabilities fared particularly badly—only 6 percent of eighth-graders in special education met the already-lenient “proficiency” standard. Other measures pointed to stagnation and decline. Enrollment was down another 2,000 students, and the district once again got a “D-minus” for a 17.4 average ACT score.

But missing AYP had little effect on Birmingham as it began the 2004–05 school year. The old accountability system had been thrown out, and the consequences of missing AYP under the new one wouldn't be felt unless the district fell short for multiple consecutive years. The Alabama Department of Education was taking steps to make sure that would never happen, by devising a new set of revisions to its accountability plan. Statewide, only 23 percent of schools and *no* districts had made AYP. Test participation rates were most often the culprit, an easy enough problem to fix. But the department took no chances, looking to other states for innovative ways to lower the NCLB bar. It found one in neighboring Tennessee.

In the previous year, states had proposed scores of accountability plan amendments to the U.S. Department of Education, every one designed to make it look like more schools and districts were high-performing. The department rejected some of the most egregious ideas.²⁵ But others, like the “Tennessee model,” made it through. This proposal applied specifically to districts. Most states—Alabama included—averaged school- and district-level test results across multiple grades. This made sense; otherwise, districts could be on the hook for over 100 separate scores.²⁶ But that still meant that districts would miss AYP if scores fell short in either reading and math for any of the student subgroups—just what NCLB’s authors intended. Tennessee proposed a much easier standard: Districts would only be identified as “in need of improvement” if they missed the state performance target in *all three grade spans*—elementary, middle, and high school—in the same subject, for two consecutive years. A district could fail two-thirds of its students every year and never be held accountable, as long as it wasn’t exactly the *same* two-thirds. Remarkably, the U.S. Department of Education approved the proposal.

That endorsement opened the floodgates. The following year, 18 states, including Alabama, asked to adopt the Tennessee model, and 10 more filed similar requests the year after that. They all cited the venerable legal principle of precedent: “If Tennessee, why not us?” The U.S. Department of Education had no response, and so the Tennessee model became the norm. Supreme Court Justice Louis Brandeis’ description of states as the “laboratories of democracy” had once again been proven true. In this case, the laboratories were concocting ever-more-innovative ways to thwart the intent of NCLB.

The U.S. Department of Education, meanwhile, was sending more signals of encouragement. President Bush’s re-election in 2004 brought a new Secretary of Education, Margaret Spellings. Her predecessor, Rod Paige, had ended his tenure in a sour relationship with the education establishment. Spellings tried to strike a more collaborative tone, announcing in April 2005 that the department would adopt “a new approach” under which “states will have additional alternatives and flexibility” if they showed evidence of academic improvement and closing the achievement gap.

Taking Spellings at her word, Alabama took steps to address its participation rate problem. Glitches aren’t

uncommon in the first year of a new testing regime, and many Alabama schools and districts increased their participation rates significantly from school year 2003–04 to 2004–05. But the state also lowered the bar by applying the uniform-averaging-with-a-twist procedure that previously had been approved for test scores, giving schools and districts credit for the greater of the current year’s participation rate *or* the three-year average.

Higher participation rates helped a number of schools, and the newer, laxer accountability standards did the rest. Statewide performance on reading and math tests inched up in elementary and middle schools from 2003–04, while staying flat in high school. The AYP results, by contrast, were dramatic. The percent of schools making the grade more than doubled, from 23 percent to 53 percent. The payoff of applying the Tennessee model to districts was even greater—68 percent met all the standards, compared to zero percent the year before. State Superintendent of Education Joe Morton announced the results, saying:

“The gains made by schools this year are very encouraging. Educators statewide worked extremely hard to improve the status of their schools and to achieve greater accountability than has ever been reported in the history of this state’s public education system.”²⁷

In newspaper stories around the state, Morton’s optimistic words were echoed by local educators eager to take credit for moving their schools off the NCLB list. Words like “confidence interval,” “uniform averaging,” and “Tennessee model” were nowhere to be found.

School Year 2005–06

Rise of the ‘Performance Index’

Birmingham City Schools missed AYP again in 2004–05, but with increased participation rates it came much closer than it did the year before. Black, Hispanic, low-income, and limited English proficient students all fell short of the state performance target in fourth-grade reading, despite the fact that—due to the adjustable rate improvement trajectory—the target hadn’t gone up from the year before. But the 99-percent confidence interval did the trick, lowering the bar far enough for each group to get over.

Only special education scores were so bleak (26 percent passing, compared to 83 percent for all students statewide) that even the confidence interval couldn't save them.

The Alabama Department of Education was hard at work on a “solution” to this problem, however, and in this it was not alone. The year 2006 marked four years since NCLB's enactment, more than enough time for states to create and refine their initial accountability plans. Yet the number of states seeking to change those plans by requesting amendments was growing, more than doubling from the year before.²⁸ It was clear at this point that the vast majority of states would *never* be content to stick with a single plan. Instead, they would continue the annual cycle of amendments—always with the aim of lowering standards—as long as they were allowed. Every year a few states would get a new standards-lowering gambit past the U.S. Department of Education, and every following year the remaining states would apply *en masse* for the same consideration.

The political pressure on the U.S. Department of Education to give in to these demands was substantial and growing. Interest groups like the National Education Association, the nation's largest teachers union, routinely denounced the law, as did conservatives who viewed it as an unwarranted expansion of federal power. The National Conference of State Legislatures issued a report questioning the constitutionality of NCLB and calling for Secretary Spellings to “waive requirements of the law.”²⁹ And when the department turned down Connecticut's request to suspend annual testing in grades 3-8—a clear violation of the statute—the state responded by filing a lawsuit in federal court.

The accountability dodge *du jour* in 2006 was the “performance index.” Like minimum subgroup sizes, it was based on a germ of a reasonable idea. NCLB critics often noted that focusing on a single performance level—“proficiency”—could cause schools to ignore students who were either far above or far below the proficiency line. Performance indices blended the percent of students at multiple achievement levels, like the lower “basic” and the higher “advanced,” into a single composite score. Alabama defined the levels in terms of simple roman numerals: Levels I, II, III, and IV, with II being equivalent to “basic,” III meaning “proficient,” and IV meaning “advanced.” The U.S. Department of Education didn't

allow states to offset underperformance at the lower levels by giving schools extra credit for the percent of students reaching the “advanced” level. So Alabama proposed giving schools half credit for every student who missed III but achieved II. The U.S. Department of Education agreed.

There was just one problem: Nearly every student in Alabama was *already* at Level II. Recall that the Level III proficiency standard itself was unusually easy; approximately 75 percent to 85 percent of Alabama students had already reached that standard by 2005, depending on the grade and test. In fourth-grade reading, 99.7 percent of students were at Level II or above. Thus, the practical effect of the performance index was simply to take half of all the non-proficient students in Alabama and count them as proficient, for no other reason than helping their schools make AYP.

Alabama also took steps to address the persistently low-performing special education population. Many states had argued that it was unreasonable to expect students with the severest disabilities to meet common standards. This problem is often exaggerated—disability categories like autism, for example, are growing but still make up less than 4 percent of all special education students, who in turn constitute less than 15 percent of all students.³⁰ From the beginning of NCLB, states were allowed to assess 1 percent of all students—not just students with disabilities, but the whole population—using non-standard assessments. The U.S. Department of Education allowed Alabama to assume that an additional 2 percent of all students had disabilities so severe that they should be counted as proficient, even if they were not.

Alabama also asked to increase its minimum subgroup size from 40 to the greater of 40 or 10 percent of the school population. Statistically speaking, this made no sense, since the accuracy of test results for a given group of students doesn't change depending on how many other students in the school there happen to be. But Alabama had reason for optimism—the U.S. Department had approved a similar request from several other states the previous year.³¹ Precedent had been a winning argument for the Tennessee model, and there was no reason to think the same principle wouldn't apply.

This time, however, Alabama was the victim of bad timing. Just as the U.S. Department of Education was considering Alabama's latest application to water down

AYP standards, the Associated Press published a series of high-profile articles about minimum subgroup sizes, noting that nearly two million, mostly minority, students nationwide belonged to excluded subgroups. Mindful of the fact that NCLB was specifically written to help those students, a group of Congressmen led by Rep. George Miller (D-Calif.), then the ranking member and now the chairman of the House Education and Labor Committee, wrote a letter to Secretary Spellings saying:

“It appears from these press reports that many states have been gaming the system, appealing to the U.S. Department of Education for loopholes that allow the exclusion of millions of minority children’s scores from their accountability subgroups. States must not be allowed to continue to do this any longer.”³²

The U.S. Department of Education subsequently denied Alabama’s request to increase minimum subgroup sizes. But the performance index gambit, combined with all the loopholes and relaxations that had come before, turned out to be more than enough to accomplish the goal of boosting the number of schools and districts making AYP. On August 7, 2006, the Alabama Department of Education released the latest school results along with a press release with the following title: “More Than 87 Percent of Alabama Schools Make AYP: More Schools Meet 100 Percent of the AYP Goals Despite This Year’s Increased Annual Measurable Objectives.”

Commenting on the new school ratings, Alabama Governor Bob Riley said: “Alabama schools are making tremendous strides. We’re seeing unprecedented improvements in the quality of education in our state.” Added State Superintendent of Education Joseph B. Morton: “What’s remarkable is that depending on the different groups of student populations across the state, schools have anywhere from five to 39 goals. Under No Child Left Behind law, if a school does not meet just one of those goals it doesn’t make AYP. That’s an incredibly high standard by anyone’s measures... .” Explaining the source of this fantastic improvement—in just two years, the percent of schools making AYP went from 23 percent to 88 percent, and the percent of districts went from a startling zero percent to 100 percent—Morton said: “Educators are teaching to higher standards, students are studying, and administrators are working to improve the status of their schools and increase the quality of the education being received.”

This was, to say the least, a deeply misleading account of why so many schools and districts in Alabama were suddenly leaping over the NCLB bar. Actual improvements in student test scores were not “unprecedented”—the percent of students scoring at proficient or above increased modestly in most grades, at roughly the same rate as the year before. And the “incredibly high” NCLB standard of meeting up to 39 goals was a fantasy; the Alabama Department of Education had spent the past four years assiduously working to shield schools from being held accountable in exactly that way.

Birmingham’s results make this clear. The district continued to struggle in 2006. Enrollment declined by another 1,400 students, and the average ACT score dropped once again. While local and state officials argued over whether school security guards should be armed with stun guns, the “black flight” phenomenon of minority families with children leaving for better suburban schools continued.³³ Said the mother of one family, which continued to pay the mortgage on its Birmingham house even as it lived in a rented suburban apartment: “I always knew that even if we had to work two or three jobs, our child was not going to Birmingham high schools.” The father of another family said: “We loved where we lived—the area, the community.” But when it came to his sons’ education, “It was either sacrifice them or get out.”³⁴

The parents’ worries were reflected in district test scores. **Table 3** shows the percent of Birmingham students in each grade and subgroup proficient in reading and math in 2006, compared to the state performance target. Dark-blue shaded cells represent scores that exceeded the state performance target, while light-blue shaded cells represent scores that fell short.

Birmingham did more than miss “just one goal” in 2006—it missed most of them. If the state hadn’t chosen the adjustable rate improvement trajectory, which kept state performance targets artificially low in the initial years of NCLB, it would have missed more still. While Birmingham had made significant progress in some grades and subjects over the previous two years, the district was still failing badly with many students, particularly those in special education, who fell short in every subject and grade by a wide margin.

But that wasn’t the message parents and the local community received the day after the scores and school ratings were released. Despite the low test scores, the

Table 3. Birmingham City Schools Test Scores 2006

Grade 3	Reading Score	Reading Target	Math Score	Math Target
All Students	74.45	73	68.32	63
Special Education	31.86	73	30.65	63
Black	74.54	73	68.31	63
Hispanic	62.5	73	64	63
White	84.38	73	71.88	63
Low-Income	72.05	73	65.44	63
Limited English Proficient	65.39	73	68.42	63

Grade 4	Reading Score	Reading Target	Math Score	Math Target
All Students	71.88	73	68.73	67
Special Education	25.26	73	24.19	67
Black	72.05	73	68.45	67
Hispanic	52.17	73	64.58	67
White	84.21	73	83.78	67
Low-Income	69.42	73	66.52	67
Limited English Proficient	50	73	65.3	67

Grade 5	Reading Score	Reading Target	Math Score	Math Target
All Students	70.2	73	69.26	59
Special Education	25.64	73	26.69	59
Black	70.21	73	69.35	59
Hispanic	60.87	73	63.46	59
White	75	73	63.88	59
Low-Income	67.75	73	67.05	59
Limited English Proficient	61.22	73	65.46	59

Grade 6	Reading Score	Reading Target	Math Score	Math Target
All Students	67.92	78	57.85	48
Special Education	23.4	78	19.16	48
Black	67.98	78	57.44	48
Hispanic	62.51	78	75.75	48
White	73.68	78	72.22	48
Low-Income	65.75	78	56.22	48
Limited English Proficient	58.62	78	73.33	48

Grade 7	Reading Score	Reading Target	Math Score	Math Target
All Students	63.68	63	41.93	40
Special Education	21.43	63	8.89	40
Black	64.05	63	41.71	40
Hispanic	36.67	63	36.36	40
White	62.07	63	62.07	40
Low-Income	61.95	63	36.56	40
Limited English Proficient	30.76	63	31.8	40

Grade 8	Reading Score	Reading Target	Math Score	Math Target
All Students	60	51	56.6	48
Special Education	20.17	51	24.57	48
Black	59.73	51	56.31	48
Hispanic	60.87	51	65.52	48
White	82.14	51	67.86	48
Low-Income	56.18	51	53.52	48
Limited English Proficient	44.44	51	60.87	48

Grade 11	Reading Score	Reading Target	Math Score	Math Target
All Students	79.12	84	77.89	73
Special Education	35.72	84	46.2	73
Black	79.38	84	77.94	73
Hispanic	54.55	84	72.73	73
White	72.72	84	72.73	73
Low-Income	75.6	84	74.85	73
Limited English Proficient	n/a		n/a	

accumulation of accountability loopholes had finally done the job—Birmingham made AYP. Echoing the celebratory tone of the governor and state superintendent, local school board president Odessa Ashley was quoted the next day in the *Birmingham News* saying: “I am overwhelmed with the progress we have made in the Birmingham public schools system. And it is because of the instructional leaders we have.”³⁵

Parents were also missing the benefits that come with schools being identified as needing improvement under NCLB. School that miss AYP for two years in a row are required to give parents the option of transferring their children to another, higher-performing school within the district. If schools fall short three years in a row, they have to offer free after-school tutoring. But because Alabama started NCLB a year late and identified few schools as missing AYP in 2006, not many parents in Birmingham or elsewhere received the choice and tutoring benefits. And no Birmingham schools had reached the more serious, final stages of NCLB accountability, the “restructuring” that can potentially result in schools being closed down. NCLB had promised tough action on behalf of children in persistently failing schools. In Alabama—four years and counting since the law’s enactment—that promise had yet to be delivered.

School Year 2006–07

Good for Schools, Bad for Parents, Children

With the work of ensuring that the vast majority of schools and districts made AYP regardless of underperformance basically finished, the Alabama Department of Education made no major changes to the state accountability plan in 2007. In 2006, the U.S. Department of Education had said that Alabama and other states would be allowed to write off 2 percent of all students as too severely disabled to test (and therefore “proficient”) “for this year only.” In 2007, the department approved an identical request “for this year only” again.

With performance targets starting to rise and no new gimmicks on the books, it wasn’t surprising that the number of schools making AYP dipped to 82 percent. They weren’t all the *same* schools, however, which meant that many schools were no longer “in need of improvement” due to missing AYP for two years in a row.

That was more than enough for Alabama’s political leaders to present the results as another great leap forward. Said Governor Riley in August 2007:

“Alabama’s schools have made tremendous progress during the past few years. Just three years ago, only 23 percent of schools met all their yearly progress goals. This year, we’ve raised the bar and saw a sharp decrease in the number of schools designated as needing ‘school improvement.’”

That meant that even fewer parents and students received school choice benefits, a fact that the Alabama Department of Education deemed important enough to merit a congratulatory press release, which began:

“Montgomery, Ala. —More positive news for Alabama schools following the release of the 2007 Adequate Yearly Progress (AYP) report. The data indicates a 70 percent decrease from last year in the number of Title I schools ... identified for School Improvement. That means fewer schools in Alabama must offer School Choice for the upcoming school year. School Choice provides parents alternatives on where their children can attend school.”

This single paragraph encapsulates the essential problem with state implementation of NCLB. Accountability systems can make life uncomfortable for underperforming public schools. That’s the point: to throw the stark light of public scrutiny on schools that consistently fail to educate students well and to create strong incentives—even requirements—for education leaders to take actions to improve. Those actions can be painful but necessary, because the best interests of adults in the education system aren’t always the same as the best interests of students. Conversely, “more positive news for Alabama schools” can be more negative news for Alabama parents and children. That the Alabama Department of Education not only admitted this but announced it in a press release shows exactly whose side it is on.

Alabama is not alone among the states in this respect, or even the worst offender. Time after time over the last five years, state departments of education acted to protect their schools and districts from the consequences of NCLB accountability, even though those actions badly

undermined the intent of the law, to the detriment of the state's students. Every one of the accountability-avoidance gambits used in Alabama has been adopted in many other states. As a result, states like Wisconsin and Oklahoma identified an even smaller percentage of schools as missing AYP.

Some states went even further. The U.S. Department of Education recently cited the Missouri Department of Education for systematically violating the law and refusing to identify schools as “in need of improvement.” Unlike Alabama, Missouri didn't bother to game the system—it *broke* the system, unapologetically. Claiming both ignorance and recalcitrance, the Missouri Department of Education's director of federal programs attributed the lawbreaking to “a combination of not agreeing with aspects of the law and not being clear about what was expected of us in some cases.”³⁶

To be sure, state departments of education aren't filled with bad people who don't care about education and actively want to ruin the lives of schoolchildren. But they are political organizations, vulnerable to pressure from special interest groups like those that represent local school boards, superintendents, and other members of the education establishment. They often see their principal clients as fellow educators, working hard to meet unreasonable demands. State departments were designed and built as bureaucracies, organizations whose essential function is compliance with rules. If you see NCLB as just another set of onerous federal rules to comply with in order to receive federal funds, then gaming the system makes a kind of sense. It's certainly cheaper and easier than actually turning around low-performing schools.

There are exceptions, of course. Some states have done better than others in complying with the letter *and* spirit of NCLB. It's instructive to note that Massachusetts—

46th on the Pangloss Index—posted significant gains on the NAEP from 2003 to 2007, despite the fact that it was *already* the highest-performing state in the nation when the law was enacted. While Alabama has boosted fourth-grade NAEP scores, its eighth-graders—the only students who were enrolled in NCLB-tested grades over the entire five-year span of the law—made almost no progress at all.³⁷

School Year 2007–08 and Beyond

Birmingham City Schools began the 2007–08 school year diminished in many ways. Over the last five years, nearly one-quarter of the student population has bled away. An emergency plan earlier in 2007 to cut \$15 million from the budget staved off financial takeover by the state, but the ongoing exodus of families and children makes future cuts all but inevitable. The students who remain are badly in need of outside help, of someone willing to step forward with the time, resources, and attention needed to turn the failing district around.

So far, the Alabama Department of Education has not risen to the challenge. Instead of working to solve the problem, it seems to have spent most of the last five years devising ever-more-creative ways to deny that it exists. Instead of holding Birmingham's leaders accountable, it has given them a series of free passes. The result, inevitably, is a district that is not nearly as effective as its children need it to be.

Hopefully, education leaders in Alabama will realize that they can best serve the state's citizens by being more honest about the state's educational failures and challenging schools and districts to improve. In the meantime, Congress should make sure that such actions are far less optional than they are today.

RECOMMENDATIONS

Congress is currently working to revise NCLB. And if there's one thing it can count on, it's that state departments of education will not change their essential nature anytime soon. There are also limits—both political and institutional—to which the U.S. Department of Education can be expected to resist constant state pressure to water down accountability. Congress would thus do well to write the next version of NCLB with the lessons of how states reacted to the current version firmly in mind.

Five Recommendations for Congress

To ensure that states better comply with the spirit of public school accountability, and to give students in cities like Birmingham a fighting chance at a high-quality education, Congress should consider the following recommendations when it writes the next version of No Child Left Behind:

- 1) Close the Loopholes.** While the U.S. Department of Education rightly rejected some state proposals to water down NCLB standards, it let others through. These proposals have now become institutionalized to the point where it will be difficult for the department to root them out of state accountability plans. Therefore, Congress should explicitly prohibit the following:
 - *Minimum subgroup sizes larger than 25:* It's reasonable to not hold schools accountable for very small groups of students, because those results can be statistically unreliable. But that interest must be balanced with the need to focus resources and attention on minority students and others who—by definition—make up only a fraction of the school population. Minimum subgroup sizes of 25 or less strike a fair middle ground.
 - *Confidence intervals of any kind:* Confidence intervals would make sense if states only tested a sample of students, but they're statistically inappropriate when testing entire populations. Minimum subgroup sizes are sufficient to ensure reliable results; adding confidence intervals on top of them is redundant and results in schools "meeting" performance targets even when falling far short of the actual goal.
- *The "Tennessee model":* The practice—now used by most states—of identifying districts as "in need of improvement" only when they miss AYP in the same subject in all three grade spans—elementary, middle, and high school—for two consecutive years has eviscerated district-level accountability under NCLB. It allows districts to fail two-thirds of their students every year and never be identified, as long as they're not exactly the *same* students. This runs directly counter to NCLB's stated goal of holding districts accountable for all students in both reading and math. It is particularly problematic given that successful school improvement strategies nearly always require strong leadership and coordination at the district level.
- *Fake "Performance Indices":* It makes sense to give schools incentives to focus on more than a single performance level, so the highest- and lowest-performing students aren't ignored. Performance indices theoretically accomplish that goal by giving schools credit for the percent of students who reach multiple standards, including the lower "basic" level of achievement. But performance indices only work when "basic" is a legitimate standard, not (as in Alabama) a level that every student with a pulse has already met. States should only be allowed to give schools credit for the percent of students reaching "basic" or its equivalent on the state test if that level (A) is at least equivalent to the NAEP "basic" level in terms of the percent of students meeting the standard, or (B) was not reached by at least 15 percent of students in 2007.
- *Adjustable Rate Improvement Trajectories:* By the time Congress finishes the new version of NCLB, most states that chose back-loaded improvement trajectories will be close to the year when the performance bill begins to come due. In that sense, most of the damage from this strategy has already been done. If, however, Congress allows states to extend the 2014 deadline for 100 percent proficiency in exchange for increasing the rigor of their standards, it should require states to create a straight-line improvement trajectory comprised of equal, annual increments.³⁸ (States that want to *front-load* improvement and reach 100 percent proficiency a few years early should, of course, be allowed to do so.)

2) Be Specific. 2007 brought a host of proposals from Congressional leaders, bipartisan commissions, and others to change NCLB’s core accountability system. While the proposals vary, many—like the idea of incorporating “multiple measures” other than standardized test scores in reading and math into state accountability systems—would give states more flexibility to implement NCLB than they have today.

Congress should be extremely cautious in this regard. NCLB is, in many ways, a narrowly written statute (in fact, this is one of the chief sources of criticism against it). Yet states have managed to find many opportunities to undermine the law within the small spaces provided by subgroup sizes, confidence intervals, etc. Increasing state flexibility will provide a whole new set of opportunities to sabotage accountability, and many states will inevitably seize the chance. To prevent this from happening, Congress should be as specific as possible when altering NCLB’s core accountability system, minimizing states’ ability to deviate from the core tenets of the law.

To take one example, some have proposed that any additional or “multiple” AYP measures beyond reading and math tests be confined to only a limited number of measures that address significant academic subjects and are available to all students in a given state. Others have called for the multiple measures concept to be applied far more loosely and broadly, giving states a great deal of flexibility to measure whatever they want, however they please. Given the track record of states in using their current, more limited flexibility under NCLB, the latter course would almost surely lead to the effective dismantlement of the Elementary and Secondary Education Act as a catalyst for school improvement.

3) Improve Peer Review. When states submitted their first NCLB accountability plans in 2002, the plans were reviewed by “peer review” teams that traveled to each state for a day to discuss the plans with state department staff. Amendments to state plans, however, have simply been reviewed internally by the U.S. Department of Education. This process clearly has been inadequate, and the U.S. Department of Education itself established a superior model when it invited state applications to experiment with “growth model” accountability measures. NCLB has been

frequently—and justifiably—criticized for its lack of attention to the year-to-year progress of individual students. If a student starts the school year four grade levels behind, and the school makes an heroic effort to help the student catch up three years, the result is not different than if it had ignored that student completely: the student is still “not proficient.” This should change. But the danger of using growth measures is that states could give schools credit for *any* growth, even if it’s not enough to help students eventually catch up. During the IASA era, when states had much more discretion to design their accountability systems, a number did exactly that.

To ensure that the growth model experiments wouldn’t undermine NCLB, the U.S. Department of Education convened a diverse mix of outside experts, including researchers, advocates, and education leaders, to review and approve state proposals. Some of the reviewers had been openly skeptical of the growth model concept. The process worked well, with the review panel ultimately approving a number of state proposals, but only those that limited extra accountability credit to students whose growth put them on a trajectory to become proficient within a defined period of time. Congress should expand this process to include *all* state accountability plans, and subsequent revisions to those plans, under the next version of NCLB. This will insulate the process from political pressure and the inevitable bureaucratic instinct to placate and compromise.

4) Stick to the Plan. When Congress gave states wide latitude to amend their accountability plans, it probably didn’t envision the never-ending barrage of standards-lowering proposals that subsequently occurred. It’s become clear that too many states see the amendment process not as a chance to fine-tune plans and correct mistakes, but a way to annually lower the achievement bar even as they tell the public that standards are being raised.

Therefore, Congress should require states to submit a two-year accountability plan, subject to the peer review process described above, after which states will have one opportunity (barring emergency or revision of technical errors) to revise the plan for another two years. This will shut down the thriving marketplace for innovative accountability dodges that is currently eroding the foundation of NCLB.

5) Promote Shared Standards. The profusion of state schemes to evade NCLB accountability is due in large part to the fact that every state is allowed and encouraged to devise its own unique system of standards, tests, and accountability. There are good reasons for some state standards to differ—perspectives on mid-19th century history differ between Alabama and Massachusetts, for good reason. But in most cases, and especially

in the core subjects of reading and math, states and students are better off working together to develop shared standards. Congress should provide stronger incentives, financial and otherwise, for states to voluntarily collaborate on a range of shared standards—not just for academic subject matter and testing, but also for the accountability systems that undergird NCLB.

Data Sources and Methodology

This report is based on data submitted by state departments of education to the U.S. Department of Education through reports called Consolidated State Performance Reports (CSPRs). The latest CSPRs were submitted in 2007 and were provided to Education Sector by the U.S. Department of Education.

The “Pangloss Index” found in Table 1 of this report is calculated by aggregating state rankings on 11 measures derived from the CSPRs. Those measures are:

- The number of persistently dangerous schools as a percentage of all schools, 2005–2006.
- Statewide high school graduation rates, 2004–2005.
- Statewide high school dropout rates, 2004–2005.
- Percent of core academic classes taught by highly qualified teachers, 2005–2006.
- Percent of qualified Title I paraprofessionals, 2005–2006.
- Percent of public schools that made adequate yearly progress, 2005–2006.
- Percent of public school districts that made adequate yearly progress, 2005–2006.
- Average percent of students proficient or advanced in fourth-grade math and reading 2005–2006. (Note: This amount, as well as the average

proficiency rates in eighth grade and high school, is calculated by averaging separately reported reading and math proficiency rates. If a state reported a proficiency rate in one subject but not the other, the proficiency rate for the reported subject was used.)

- Average percent of students proficient or advanced in eighth-grade math and reading, 2005–2006.
- Average percent of students proficient or advanced in high school math and reading, 2005–2006.
- Percent of teachers receiving high-quality professional development, 2003–2004. (Note: States were not required to report this measure in 2007. To facilitate comparisons with the 2006 Pangloss Index, the same data were used for the 2007 that were used in 2006, which were derived from CSPRs submitted in 2005).

For each measure, states were ranked so that the states reporting the most positive results were ranked highest. For example, while states were ranked highest if they reported the highest high school graduation rates and highest percent of schools making adequate yearly progress, they were also ranked highest if they reported the lowest number of persistently dangerous schools and the lowest high school dropout rates.

Some states did not submit data for some measures. In those cases, states were, for ranking purposes, assigned the median value of those states that did submit data.

Endnotes

- ¹ Available online at <http://www.edweek.org/apps/maps/>.
- ² Kevin Carey, *Hot Air: How States Inflate Their Progress Under NCLB* (Washington, DC: Education Sector, 2006).
- ³ See “Data Sources and Methodology” for a complete list of factors and methodology used to create the index.
- ⁴ See, for example, Dan Keating and V. Dion Haynes, “Can D.C. Schools Be Fixed?” *The Washington Post*, June 10, 2007.
- ⁵ The state reports used to compile the Pangloss Index report data from the previous year; the 2006 report contained 2005 achievement data while the 2007 report contained 2006 achievement data.
- ⁶ *State Board of Education Report Card for Birmingham City Schools, 2001–2002*.
- ⁷ Jen Sansbury, “Birmingham Audit Can’t Verify Brown School Transactions,” *The Atlanta Journal-Constitution*, February 4, 2003; Carla Crowder, “City Schools Pay Big for Brass,” *The Birmingham News*, December 10, 2001.
- ⁸ David Nichols, “Birmingham Has Abandoned Its Children,” *The Birmingham News*, July 13, 2003.
- ⁹ For more detailed information on Annual Measurable Objectives, see Erin Dillon and Andrew J. Rotherham, *States’ Evidence: What It Means to Make ‘Adequate Yearly Progress’ Under NCLB* (Washington, DC: Education Sector, 2007).
- ¹⁰ In 2003, 31 states chose a minimum subgroup size smaller than 40. Twelve states chose 40, and seven chose minimum subgroup sizes larger than 40. See Mary Fulton, *Minimum Subgroup Size for Adequate Yearly Progress (AYP), State Trends and Highlights* (Denver: Education Commission of the States, 2006).
- ¹¹ Author’s calculations based on enrollment from 2005–2006 (the latest available) from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data.
- ¹² No Child Left Behind Act of 2001, Public Law 107-110, 107th Cong., (January 8, 2002), Sect. 1111(b)(2)(J).
- ¹³ *State Board of Education Report Card for Birmingham City Schools, 2002–2003*.
- ¹⁴ David Nichols, “Birmingham Has Abandoned Its Children.”
- ¹⁵ Marie Leech, “Abandoned Schools Leave Chasms in Communities,” *The Birmingham News*, February 13, 2007.
- ¹⁶ Available online at <http://literacy-council.org/initiative.htm>, see also https://www.casas.org/lit/litcode/Detail.CFM?census__AREAID=1.
- ¹⁷ Available online at <http://www2.edtrust.org/edtrust/summaries2004/SouthCarolina.pdf>.
- ¹⁸ Because Alabama was late in establishing criterion-referenced tests and had 2002–2003 scores to use, its performance targets for 2003–2004 were based on test scores from 2003–2004.
- ¹⁹ The actual percentage varies because this calculation was weighted for school size.
- ²⁰ No Child Left Behind Act of 2001, Section 1111(f)(2).
- ²¹ Ellen Forte Fast and William J. Erpenbach, *Revisiting Statewide Educational Accountability Under NCLB: A Summary of State Requests in 2003–2004 for Amendments to State Accountability Plans* (Washington, DC: The Council of Chief State School Officers, 2004).
- ²² David Rogosa, *The “99% Confidence” Scam: Utah-Style Calculations*, November 2003.
- ²³ W. James Popham, “AYP Wriggle Room Running Out,” *Educational Leadership*, September 2005.
- ²⁴ The larger the number of students being measured, the smaller the confidence interval, since more students increases the accuracy of the “sample.”
- ²⁵ It is difficult to document precisely which proposals were made and rejected, since the U.S. Department of Education does not post state requests online, and, according to William J. Erpenbach and Ellen Forte, *Statewide Educational Accountability Under the No Child Left Behind Act—A Report on 2005 Amendments to State Plans* (Washington, DC: The Council of Chief State School Officers, 2005), the U.S. Department of Education often rejects state requests via undocumented “off-the-record conversations.”
- ²⁶ Two subjects (reading and math) multiplied by nine grades (3–8 and once in high school) multiplied by eight subgroups (Hispanic, white, black, Asian, Native American, low-income, special education, and English language learner) equals 144 scores.
- ²⁷ Alabama Department of Education Press Release August 11, 2005.
- ²⁸ Ellen Forte and William J. Erpenbach, *Statewide Educational Accountability Under the No Child Left Behind Act—A Report on 2006 Amendments to State Plans* (Washington, DC: The Council of Chief State School Officers, 2006).
- ²⁹ *Delivering the Promise: State Recommendations for Improving No Child Left Behind* (Washington, DC: National Conference of State Legislatures, 2005).
- ³⁰ Erin Dillon, *Labeled: The Students Behind NCLB’s ‘Disabilities’ Designation* (Washington, DC: Education Sector, 2007).
- ³¹ Ellen Forte and William J. Erpenbach, *Statewide Educational Accountability Under the No Child Left Behind Act—A Report on 2006 Amendments to State Plans*.
- ³² “House Democrats Ask Spellings to Act Quickly to Correct Problem With No Child Left Behind,” Press Release, April 28, 2006.
- ³³ *The Birmingham News*, “It’s a Shock to the System,” March 8, 2006.
- ³⁴ Barnett Wright and Kelli Hewett Taylor, “City’s Black Students Leaving for Suburbs: Enrollment Down 20% Over 5 Years as Parents Abandon Birmingham Schools,” *The Birmingham News*, August 6, 2006.
- ³⁵ Patrick Hickerson, “Area Schools Improve in Achieving Goals in Birmingham, 80% Make Grade,” *Birmingham News*, August 8, 2006.
- ³⁶ Andrew Brownstein, “Missouri Cited for Accountability Breakdown, ED Pushes Greater Oversight of Districts,” (Tampa, FL: Thompson Publications, September 28, 2007).

³⁷ From 2003 to 2007, the average scale score of Alabama eighth-graders on the NAEP increased from 262 to 266 in mathematics, and declined from 253 to 252 in reading. Neither change was statistically significant.

³⁸ This approach has been suggested by The Education Trust, see: <http://www2.edtrust.org/NR/rdonlyres/5A150FED-85FD-4535-8DF6-737A536EB0FB/0/EdTrustNCLBRecommendations41607.pdf>.