

# Accountability Systems

This is the first in a series of issue briefs that we hope policymakers, advocates, parents, educators, and the general public will find useful as Congress and the president embark on reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965, currently known as No Child Left Behind (NCLB).

ESEA is the law that governs the vast majority of federal K–12 education programs. Total funding for ESEA programs is \$25 billion for FY 2010, plus an additional \$19 billion in supplemental funding under the American Recovery and Reinvestment Act (ARRA).

While ESEA encompasses dozens of federal programs in a number of specific policy areas, the lion's share of both funding and policy comes by way of the formula-driven Title I<sup>1</sup> program. (Title I was traditionally called "Education for the Disadvantaged," but on February 1, 2010, the Department of Education proposed the name "College and Career Ready Students.") Title I funding for FY 2010 was \$14.5 billion (58 percent of ESEA funding) with a supplement of \$10 billion under ARRA. All of the accountability provisions reviewed in this brief are in Title I of the ESEA statute. (More detailed and up-to-date information on the federal education budget can be found at [www2.ed.gov/about/overview/budget/budget11](http://www2.ed.gov/about/overview/budget/budget11).)

The current ESEA, authorized in 2002, technically expired in 2008, but one-year extensions have been, and will continue to be, triggered automatically until the law is reauthorized.

*"I think it is very difficult for a person who lives in a community to know whether, in fact, his educational system is what it should be, whether if you compare his community to a neighboring community they are doing everything they should be, whether the people that are operating the educational system in a state or local community are as good as they should be.*

*... I wonder if we couldn't have some kind of system of reporting ... through some testing system that would be established [by] which the people at the local community would know periodically ... what progress had been made."*

**Senator Robert Kennedy,**  
U.S. Senate hearing, 1965

## HISTORY OF ACCOUNTABILITY IN ESEA

Testing and accountability provisions have existed in ESEA since 1965. Prior to 1994, however, these guidelines were vague and unenforced. This changed with the passage of President Clinton's Goals 2000 Act and 1994 ESEA reauthorization, which required each state to: (1) develop academic standards, (2) create and administer annual assessments aligned to those standards (once each in grades 3–5, 6–8, and 10–11), and (3) develop a system of adequate yearly progress (AYP) by which to judge student attainment of state academic standards. Few states, however, set clear goals or reported separately on the progress of at-risk groups.

The 2002 reauthorization of ESEA – NCLB – made three key changes to ESEA accountability requirements:

- 1) States would have to disaggregate student achievement data within each state, local education agency, and school; by gender, racial and ethnic group, and English proficiency status; and for immigrants, students with disabilities, and economically disadvantaged students.
- 2) States would have to set AYP goals for closing achievement gaps between students in the above at-risk subgroups and their peers by 2014. Local education agencies and schools would be held accountable for meeting their annual objectives (e.g., increasing the percentage of students proficient in reading from 50 percent to 55 percent).<sup>2</sup>
- 3) Testing would be required annually in reading and math in grades 3–8.

NCLB also required all states to participate in the National Assessment of Educational Progress (NAEP), the nation's report card, as a check on state test results. (Several states had previously not participated in NAEP.)

## THE COMPELLING RATIONALE FOR STANDARDIZED TESTS

History shows that poor and minority students are generally held to lower standards than their peers. They are often told they are doing "fine," and they receive good grades and ultimately high school diplomas – only to graduate from high school and discover that they don't have the knowledge and skills needed to succeed in college and/or the workplace. Intentional or not, biases in the system frequently mask failure until it is too late to remediate.

To understand whether or not students – regardless of race, income, national origin, disability, or ZIP code – have access to a high-quality education, states realized they need a fair way of assessing school performance. Policymakers need a valid way to determine whether **all students are being held to the same high standards and to compare students' achievement of those standards across schools, districts, and states.** Standardized tests are the only truly objective and scientifically valid way to make these critical comparisons.

### An esteemed National Academies of Sciences panel concluded:

*"Can scores on one test be made interpretable in terms of scores on other tests? Can we have more uniform data about student performance from our healthy hodgepodge of state and local programs?"*

*"After deliberation that lasted nine months, involving intensive review of the technical literature and consideration of every possible methodological nuance, the committee's answer was a blunt 'no.'"*

**Michael Feuer,**

Executive Director of the Division of Behavioral and Social Sciences and Education, National Research Council, National Academies of Sciences, in "Moderating the Debate," Harvard Education Press, 2006

In recognition of this, in 1989, the National Governors Association (NGA), led by then-Governor Clinton and supported by President George H.W. Bush, launched an initiative for new state standards and assessments, which was then codified into federal law in 1994 and accompanied by \$2 billion in start-up and implementation funding over the next six years. It is critical to note that states use this information to guide their decisionmaking and policy-setting on all fronts – funding; curricula; supports and interventions for students, teachers, and schools; etc. States are not collecting data to collect data; they are using this information to improve their education systems.

Whether or not we continue to obtain information from standardized tests and use it to inform policy decisions from the White House to the classroom will have serious, long-term implications for students. **If we revert to a patchwork of standards and assessments that vary according to political pressure or societal and community biases**, historically disadvantaged students, whether intentionally or unintentionally, will be mislabeled as achieving high standards when in fact they are not. In turn, **the schools in which poor and minority students are enrolled are likely to be overlooked when it comes to badly needed investments in teaching and learning** and in formulating and implementing fundamental reforms in chronically failing schools.

## WE DON'T NEED MORE TESTING OR LESS TESTING; WE NEED BETTER TESTS

As the American Federation of Teachers (AFT) states: "When done appropriately, annual state-administered tests provide useful feedback about student learning and can guide the system to ensure that schools, teachers, and staff get the supports they need to help all students meet expectations."<sup>3</sup>

It is widely recognized that many (if not most) state tests fall short of the ideal in both the range and level of material they cover and the way in which they assess student knowledge and skills (e.g., over-reliance on multiple choice or "fill in the bubble" questions). President Obama, Secretary Duncan, and most governors and state school chiefs concur that an overhaul of state assessment systems is long overdue.

Therefore, state assessment systems are currently targeted for improvement through \$350 million in federal funds that have been set aside under the Race to the Top fund to create a multistate assessment system aligned with college- and career-ready standards. The hope is that in joining together around a set of common, high standards, states can avoid the tendency to "dumb down" student assessments and instead produce results that allow valid interstate comparisons.

## DISAGGREGATION OF DATA IS CRUCIAL TO HIGHLIGHTING ACHIEVEMENT GAPS

Before the passage of NCLB, state and local accountability systems typically analyzed and published student achievement data based only on average student achievement. This type of reporting hid critical information about subgroups of students — particularly gaps between their achievement and that of their more advantaged peers. Federal law requires that student achievement data be disaggregated by race, ethnicity, gender, disability status, migrant status, English proficiency, and status as economically disadvantaged. (Note: School lunch is used most often, but not always, to measure economic disadvantage.)

Prior to 2002:<sup>4</sup>

- Only 11 states disaggregated (i.e., analyzed, compared, and reported) achievement data by gender or ethnicity;
- Only six states disaggregated for economically disadvantaged students;
- Only seven states disaggregated for English proficiency status;
- Only one state disaggregated for migrant students; and
- Only one state (Texas) had a state goal of narrowing or closing achievement gaps between any of these groups.

## TIMELINES AND ANNUAL BENCHMARKS ARE A NECESSARY PART OF ACCOUNTABILITY SYSTEMS

Critics claim that goals and timelines cause states to lower or dumb down standards to make meeting such goals easier, producing overly rosy reports of academic progress and gap-closing.

Some states clearly have lowered their achievement bars, but overall, the arguments pitting goals and timelines against standards are specious. Unless there are specific goals for students to meet, having high standards is an empty exercise. And history shows that without federal requirements, many states and school districts would have *neither* high standards *nor* real timelines.

The false choice between high standards and clear timelines is made painfully clear when we examine high school graduation. None of the "dumbing-down" criticisms of testing apply to graduation rates. Graduation is a more clear-cut outcome than a test score, and it's a more life-changing one. Despite this reality, too few states have measurable goals for increasing graduation rates or for closing graduation gaps between groups of students.

As always, the politics of education tend to mask failure. There is pressure to avoid setting goals that are hard to meet because, frankly, doing so is stressful for adults. Experience and history both tell us that discomfort in the short term is a necessary prerequisite for change that benefits students in the long run.

We shouldn't have to choose between high standards and ambitious goals for student achievement. Any effort to establish a world-class U.S. public education system that meets the needs of all students requires both.

## AMBITIOUS GOALS AND TIMELINES HAVE IMPROVED ACADEMIC PROGRESS FOR AT-RISK STUDENTS

A number of rigorous research studies have demonstrated that strengthened accountability over the last decade is associated with improved student achievement, especially for low-income and minority students.<sup>5</sup>

A majority of states have seen improvements in overall student achievement *and* a narrowing of achievement gaps since NCLB was enacted.<sup>6</sup> For example, results from the 2005 NAEP show more progress was made by nine-year-olds in reading from 2000 to 2005 than in the previous 28 years combined.<sup>7</sup> NCLB has certainly not "fixed" public education, nor has it closed the achievement gap; but tangible improvements have resulted from this legislation, and they provide critical information we can use to strengthen the next reauthorization.

Moreover, despite assertions to the contrary, these results have not come about due to more students dropping out of school. While national graduation rates are unacceptably low, a 2009 study conducted by the America's Promise Alliance found that graduation rates actually *inched up* in recent years, from 66 percent in 1995 to 71 percent in 2005. The rate still falls far short of acceptable, and it is not a huge increase, but it is a significant one — certainly not the decline some have falsely claimed.<sup>8</sup>

## HIGH GOALS AND CREDIT FOR GROWTH ARE COMPATIBLE

NCLB has a *stated* goal of all students in all subgroups reaching proficiency by 2014. But the NCLB "safe harbor" provision allows a school or district to meet AYP regardless of where it is on the path to 100 percent proficiency as long as it *reduces*, by at least 10 percent, the percentage of students overall and in key demographic subgroups who are not proficient. This provision overrides NCLB's "100 percent by 2014" goals, making it moot in an increasing number of cases.

The table to the right illustrates what constitutes AYP for any school between "Year X" and "Year X +1."

There are arguments to be made about whether nine percentage points is too much expected annual progress for a school at 10 percent proficiency — or whether two percentage points for a school at 80 percent proficiency is too little — and certainly there is room for improvement in moving toward a more nuanced model. But the fact remains: **Under current law, even with a stated goal of 100 percent proficiency by 2014, all schools get credit for growth short of that goal, and no school or district has to be at 100 percent proficiency in 2014 or any other year. Not even close.**

NCLB Safe Harbor Provision for AYP

% Proficient Year X	Safe Harbor AYP Year X +1
10%	19%
20%	28%
30%	37%
40%	46%
50%	55%
60%	64%
70%	73%
80%	82%
90%	91%

### Key Findings on Post-NCLB Student Achievement

1. In most states with three or more years of comparable test data, student achievement in reading and math has gone up since 2002, the year NCLB was enacted.
2. There is more evidence of achievement gaps between groups of students narrowing since 2002 than of gaps widening. Still, the magnitude of the gaps is often substantial.
3. In nine of the 13 states with sufficient data to determine pre- and post-NCLB trends, average yearly gains in test scores were greater after NCLB took effect than before.

**Center for Education Policy,**  
"Answering the Question That Matters Most: Has Student Achievement Increased Since No Child Left Behind?" 2007

## THE "CURRICULUM NARROWING" TREND IS MINISCULE OR NONEXISTENT

The U.S. Department of Education has conducted studies of time spent on specific subjects over the last 20 years, consistently finding no curriculum narrowing, i.e., decreasing time spent on certain subjects. In fact, slight *decreases* (not statistically significant) were found in both English and math instruction. A 2007 study by the Institute for Education Sciences found that as a percentage of total hours in the student school week, English instruction in grades 1–4 was only slightly, and not statistically significantly, different in 2003–04 than in 1987–88 (35.5 percent versus 35.0 percent). From 1999–2000 to 2003–04, the amount of time spent on math instruction in grades 1–4 actually declined slightly (from 17.4 percent to 16.5 percent of total student hours in the school week).<sup>9</sup>

A nonscientific, self-reporting survey of educators conducted by the Center for Education Policy (CEP) found that since 2002, 56 percent of all districts had not reduced time spent on social studies, science, arts and music, lunch, or recess to fit in more time for reading and math; 44 percent reported some narrowing.

When students can't read, write, or do math at grade level, a focus on those subjects for at least a period of time may be justified, given that these skills are prerequisites for success in all other subjects. And this is what the CEP study found, though even here curricula narrowing occurred in only a little less than half of districts. Specifically, the study found that in districts where at least one school was identified as not making AYP, 51 percent decreased time spent on social studies; conversely, only 31 percent in districts with no AYP-identified schools did.

## IMPROVING LOCAL ACCOUNTABILITY POLICIES AND PRACTICES

Research shows that schools react differently to the demand for accountability according to their own policies and practices of institutional accountability. In a recent book, Richard Elmore and colleagues (Harvard Graduate School of Education) observed that schools with strong internal accountability and sound instructional practices react more positively to external accountability systems than schools that are not organized this way:

*The editors hypothesize that schools are more successful (i.e., more learning occurs) when formal and informal accountability mechanisms are aligned with individuals' internalized notions of accountability and responsibility. A strongly "aligned" school would be one where, for instance, there is coherence among teachers' and administrators' expectations for student success and a philosophical agreement on how that success is obtained. In contrast, a highly misaligned school is one with a relatively "weak or dysfunctional internal accountability system." For example, it may be a school where "the principal forces teachers to adhere to rules that they know result in poor achievement outcomes."<sup>10</sup>*

One of the most dysfunctional practices associated with strong accountability systems is narrowly "teaching to the test." Two points are essential here: (1) not everyone does it; and (2) it is *not* an effective approach, either to improve instruction *or* to raise test scores.

Assessment expert James Popham (University of California, Los Angeles School of Education and Information Studies) distinguishes between "curriculum teaching" and "item teaching" (i.e., teaching to the test). Curriculum teaching involves a focus on the full body of knowledge and skills students need to master a subject. It is the kind of learning that standards-based reform – i.e., high standards aligned with good assessments and rich curricula – aims to foster. "Item teaching" is the narrow drill and kill, teaching-to-the-test approach that is intended only to raise test scores rather than expand student knowledge and skills.

**Research shows that "item teaching" (i.e., test preparation) does not always follow from strong accountability, and when it does, it does not work either to improve student learning or to raise test scores.**

In a seminal study in Chicago, researchers found that in classrooms in which teachers employed "authentic instruction" – i.e., a focus on the full body of skills needed to master a subject – score gains on the Illinois Test of Basic Skills (ITBS) exceeded the national average by 20 percent. Conversely, in classes in which students were given few authentic assignments – i.e., in which there was a narrow, drill and kill approach – ITBS gains were much *less* than the national average. Moreover, the type of instruction offered was found to be a function of "teacher disposition and choices" rather than the particular characteristics or achievement levels of the students being taught.<sup>11</sup>

In a comprehensive review of the "teaching to the test" issue, Craig Jerald concludes:

*Accountability and standardized tests need not be in conflict with good instruction. ... Teaching to the test by dumbing down instruction offers only a kind of fool's gold, promising a payoff that does not deliver. The choice between good instruction and good test scores is a false one.<sup>12</sup>*

## ENDNOTES

- <sup>1</sup> For an overview of the Title I program, see the U.S. Department of Education Web site: [www2.ed.gov/programs/titleiparta/index.html](http://www2.ed.gov/programs/titleiparta/index.html).
- <sup>2</sup> Section 1116 of ESEA prescribes a tiered menu of interventions that states and districts may, or in some cases must, take if districts or schools do not meet their annual goals. The options that states or districts have to pursue vary according to the number of years that a district or school does not meet AYP, though there is a great degree of latitude in both principle and (even more so) in practice. The interventions that follow from the ESEA accountability system will be the subject of a subsequent issue brief.
- <sup>3</sup> American Federation of Teachers (2006), *Smart Testing: Let's Get It Right*. <http://archive.aft.org/pubs-reports/downloads/teachers/Testingbrief.pdf>.
- <sup>4</sup> *No State Left Behind: The Challenges and Opportunities of ESEA 2001*. Report prepared by the Education Commission of the States (ECS) under the auspices of the State Education Policy Network, a partnership of the Council of Chief State School Officers, ECS, Education Leaders Council, National Association of State Boards of Education, National Conference of State Legislatures, and National Governors Association. [www.ecs.org/clearinghouse/32/37/3237.pdf](http://www.ecs.org/clearinghouse/32/37/3237.pdf).
- <sup>5</sup> Charles Barone (2009), *Are We There Yet? What Policymakers Can Learn from Tennessee's Growth Model*. Education Sector Technical Reports. [www.educationsector.org/usr\\_doc/Are\\_We\\_There\\_Yet.pdf](http://www.educationsector.org/usr_doc/Are_We_There_Yet.pdf). Also see Eric A. Hanushek and Margaret E. Raymond (2003), "Lessons about the Design of State Accountability Systems," in *No Child Left Behind? The Politics and Practice of School Accountability*, Paul E. Peterson and Martin R. West, eds., Brookings Institution, p.127-52.
- <sup>6</sup> Center on Education Policy.
- <sup>7</sup> [www.ed.gov/nclb/overview/importance/nclbworking.html](http://www.ed.gov/nclb/overview/importance/nclbworking.html).
- <sup>8</sup> [www.americaspromise.org/Resources/Research-and-Reports/c/Cities-In-Crisis-2009.aspx](http://www.americaspromise.org/Resources/Research-and-Reports/c/Cities-In-Crisis-2009.aspx).
- <sup>9</sup> Institute for Education Sciences, National Center for Education Statistics (2007), *Changes in Instructional Hours in Four Subjects by Public School Teachers of Grades 1-4*. <http://nces.ed.gov/pubSearch/pubsinfo.asp?pubid=2007305>.
- <sup>10</sup> Martin Carnoy, Richard Elmore, and Leslie Siskin (eds.) (2003), *The New Accountability: High Schools and High-Stakes Testing*. Routledge-Falmer. Reviewed by Sharon L. Nichols (2004) in *Education Review*.
- <sup>11</sup> Fred D. Newmann, Anthony Birk, and Jenny Nagaoka (2001), *Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence?* Consortium on Chicago School Research. <http://ccsr.uchicago.edu/publications/p0a02.pdf>.
- <sup>12</sup> Craig D. Jerald (2006), "Teach to the Test? Just Say No." The Center for Comprehensive School Reform and Improvement, Learning Point Associates. [www.centerforcsri.org/files/CenterIssueBriefJuly06.pdf](http://www.centerforcsri.org/files/CenterIssueBriefJuly06.pdf).

**DEMOCRATS**  
for EDUCATION REFORM

*Democrats for Education Reform (DFER) is a political action committee whose mission is to encourage a more productive dialogue within the Democratic Party on the need to fundamentally reform American public education. DFER operates on all levels of government to educate elected officials and support reform-minded candidates for public office.*  
[www.dfer.org](http://www.dfer.org)

EDUCATION  
EQUALITY  
PROJECT  
[www.edequality.org](http://www.edequality.org)

*The Education Equality Project (EEP) is leading a civil rights movement to eliminate the racial and ethnic achievement gap in public education. EEP is a national, bipartisan advocacy group of elected officials, civil rights leaders, and education reformers working to bring equity to our public education system.*  
[www.edequality.org](http://www.edequality.org)  
[www.twitter.com/EdEquality](https://twitter.com/EdEquality)