



A Lens That Distorts

NCLB's faulty way of measuring school quality

No Child Left Behind (NCLB) put schools under the microscope by requiring that they report, annually, the test-score performance of students in grades 3 through 8, and, again, for grade 10. As President Bush said shortly before he signed the bill into law, “We need to know whether a curriculum is working. We need to know whether the teachers, the methodology that teachers use is working. We need to know whether or not people are learning. And if they are, there will be hallelujahs all over the place. But if not, we intend to do something about it.”

Five years later, it has become clear that the microscope NCLB uses to get the information the president said he wanted contains a lens that distorts. Many good schools—both charter schools and inner-city public schools serving the disadvantaged—are not recognized as such, while many poorly performing schools are given a pass. If NCLB is to fulfill its mission, Congress needs to make some major repairs or risk seeing those opposed to all forms of school accountability assume control of the political battleground.

I do not join those opponents of accountability. We do not raise any principled objections to holding schools accountable or testing students annually. On the contrary, the evidence

BY PAUL E. PETERSON

suggests that accountability has had some positive benefits for American education. The reading and math scores of 4th and 8th graders on the National Assessment of Educational Progress (NAEP) have risen steadily since NCLB was put into place. If it cannot be proved that those gains are due to improved school accountability, it is heartening to know that Margaret Raymond and Eric Hanushek found, in more precise estimates of accountability impacts, somewhat larger gains on the NAEP in those states that were the first to put accountability systems into place (see “High-Stakes Research,” *features*, Summer 2003).

Still, the current federally mandated accountability system falls well short of what is needed. The gains made by 4th and 8th graders do not translate into higher levels of performance once students reach the age of 17. Instead, high school achievement has remained as stagnant as ever, and high school graduation rates continue to hover around the industrial world average.

Two things are needed to get the most out of accountability. First, the lens used to look at schools must be reground so that distortions are minimized. Once that repair has been completed, accountability’s bright light needs to shine on the performance of individuals, that is, on students, teachers, and administrators, not just on schools.

A Distorted Prism

Most people would agree that a good school is a place where students are learning, and a poor school is one where that is not happening. But NCLB’s way of measuring school performance does not look directly at how much individual students are learning from one year to the next. Instead, a school is evaluated according to whether or not its students are making Adequate Yearly Progress (AYP) toward full proficiency by 2014. In that year, every tested student must be achieving at the state-determined proficiency level. By next year, the midpoint between 2002 and 2014, the percentage of students proficient at a school is expected to have increased by roughly half the distance from where it was when the law was enacted. Various subgroups of students, defined by ethnicity, gender, economic disadvantage, and need for special education, must be making comparable progress. While some exceptions to those requirements are allowed, schools are said to be A-OK only if the percentage of students scoring at the proficient level is moving forward “adequately.” Schools where that is not happening are identified as “not making AYP” or, after two years, “in need of improvement.” In common parlance, they have “failed.”

Evaluating schools by the AYP measuring stick is typically justified on the grounds that it ensures that “no child shall be left behind.” While this sounds both noble and egalitarian, it in fact expects those schools that had a lower percentage of

students scoring at the proficient level in 2002 to make more rapid progress over the ensuing years than those with higher-performing students. Both are expected to arrive at the same point by 2014 as well as to make more rapid progress each year. It is not unlike a race between the turtle and the rabbit, in which the turtle is asked to complete a two-mile run while the rabbit need only traverse 200 meters in the same stretch of time.

The consequences of this peculiar accountability system are very different, but equally damaging, for two contrasting



The lens used to look at schools must be reground so that

distortions are minimized.

Once that repair has been completed, accountability’s bright light needs to shine on the performance of individuals.

types of schools. For those schools blessed with high-performing students (as a result of learning either at home or in earlier grades), the proficiency standard to which they are held accountable is often much too low. If the country is to have a better-educated citizenry, the schools serving higher-performing students need to lift their performance well above levels of mere “proficiency.” As for those schools whose responsibility is the education of large numbers of low-performing students (as a result of either minimal education at home or bad instruction earlier in life), it is not reasonable to expect that every child will reach the state proficiency standard by the end of 3rd grade and every grade thereafter. Applying that criterion puts schools serving the disadvantaged at risk of being said to “fail,” even if they are doing a fine job of enhancing the skills of their students.

One can get a pretty good fix on how much students are learning by tracking individual student test scores from one year to the next. When students at a particular school

are outpacing the typical student in the rest of the state, most people would agree the rate of learning is, at least, better than average. When student gains lag significantly behind average statewide gains, most people might agree that the situation deserves attention by school boards and administrators, if only to make sure that below-average performance in any given year is nothing more than a statistical aberration.



Schools serving the poor, the disabled, and the educationally disadvantaged face a greater challenge, as they must make rapid progress from one year to the next to escape the “failed school” designation.

Any well-designed measuring stick should provide that kind of basic information, especially if it purports to identify schools that are or are not making Adequate Yearly Progress. Martin West and I (“Is Your Child’s School Effective?” *check the facts*, Fall 2006) discovered just how inadequate the AYP measuring tool is when we tracked student progress in Florida. We compared pairs of schools, checking to see whether students were learning more at the one said to be making AYP than at the one said to be failing. Thirty percent of the time the opposite proved to be the case. Any measuring stick that gets something wrong 30 percent of the time is itself a failure.

The errors are systematic. If a school is blessed with initially high-performing students, it is likely to be given a pass by NCLB, even if students are not learning much from one year to the next. Schools serving the poor, the disabled, and the educationally disadvantaged face a greater challenge, as they must make rapid progress from one year to the next to escape the “failed school” designation. As a result, they are

often found to be “failing,” even when the gains made by their students exceed those in “passing” schools.

Simplistic Dichotomy

The imperfections of the NCLB measuring stick are magnified by the fact that it divides all schools into just two categories, pass or fail (“making AYP” or not). The practice borrows from a more common propensity that has unfortunately crept into American education in the name of helping the challenged. Even elite universities, such as my own, allow professors to give students “pass-fail” grades. I have learned from bitter experience that such a grading system both gives students license to do nothing and, ultimately, provides less information to those who rely on grades as a way of ascertaining whether students have learned something. (Generally speaking, the “behind-the-scenes” rule is to treat a pass as a fail, causing further distortions.)

In days gone by, and even now in traditional schools, teachers graded students over a five-point scale that ranges from A to F. NCLB needs to rediscover that ancient practice. States have already shown the benefits of using a multiplicity of cut points. Florida employs an A to F scale, providing a much more intuitive way of telling families and citizens about the quality of their schools. New York grades schools on a four-point scale, which, if not as good as the traditional grading system, would be satisfactory were it not for the fact that, in New York, 4 is good, 1 is bad. (Can you hear the chants? “We’re number four!”)

Regrinding the Lens

A five-point A to F scale that focused strictly on student growth at a school would greatly enhance the transparency of the accountability system. Admittedly, such scrutiny was not possible when NCLB was originally enacted into law, simply because at the time the legislation was passed there was no way in most states of tracking student progress over time. Since 2002, however, several states (North Carolina, Texas, and Florida, for example) have put into place at least the beginnings of systems that allow for tracking of student performance from one year to another. At the time NCLB is reauthorized, Congress needs to mandate such tracking systems in all states, and then ask states to use the systems as a way of identifying which schools are effective, and which are not.

To be sure, not every state could implement such a system immediately, so Congress would need to allow for a period of transition from the current policies to the new ones. Introducing a growth approach via the “safe harbor” provisions of the law may be the politically feasible way to begin. States with high standards and quality information on individuals’ performances over time could be given a second way

of showing that schools are making AYP. If given this option, they will have every incentive to migrate to the new system as quickly as possible, as the distortions of the existing approach intensify.

Distortions across States

Thus far we have focused on how the NCLB accountability lens provides misleading information about school quality within states. Equally disconcerting, the accountability measuring stick provides grossly misleading results when states are compared to one another. The cause of the distortion: allowing each state to establish its own standards and its own definition of proficiency, thereby generating 50 different definitions of the same concept.

The meaning of the word “standard” has its origins in the flag or emblem carried high in battle in order to martial a fighting force toward a fixed objective. If standard-bearers head off in inconsistent directions, they direct portions of the battle force toward divergent objectives, opening the army to flank attacks. Accordingly, “standard” came to mean something that was fixed, such as the specific weight for an official coin or the unchanging value of a precious metal against which the value of paper currency could be compared.

It is thus an oxymoron to hold students accountable to more than one standard, as is the case under NCLB, which allows each state to establish its own standard, no matter how widely it diverges from some national definition. Frederick Hess and I show that a very few states—only Massachusetts, Maine, and South Carolina—have as high a definition of proficiency as the one originally set nationally by those who administer the NAEP. (For more, see “Keeping an Eye on State Standards,” *features*, Summer 2006.) Standards in most states fall far short of that national mark, North Carolina, Oklahoma, and Tennessee being the most extreme laggards. So, by official definitions, Johnny may be deemed a proficient reader in North Carolina but not if he should move to South Carolina.

So varied are state standards that the relative rigor of a state’s proficiency definition is a better predictor of the percentage of schools said to be “failing” (not making AYP) than the overall quality of student performance in the state, as estimated by average NAEP scores. The correlation between the proficiency standard and the percentage of schools failing to make AYP rate is 0.44. The correlation between the actual level of student proficiency on the NAEP in the state and the percentage of schools identified as “failing” is only negative 0.31.

Clearly, AYP is giving information that is at least as much political as it is substantive. In Massachusetts, for example, 43 percent of the students failed to make AYP, despite the fact that the state has the highest-performing students in the country. Why? Because Massachusetts has one of the highest standards in the country, a standard as high as the one

NAEP uses. Conversely, only 7 percent of the schools in Tennessee are failing, though the state ranks near the bottom in terms of school performance. Why? Because Tennessee has one of the lowest operational definitions of proficiency in the country. The pattern nationwide is laid out in Table 1.

We are not necessarily proposing the NAEP or Massachusetts standard of “proficiency” as the correct one. If one is going



The accountability measuring stick provides grossly misleading results when states are compared to one another. The cause of the distortion: allowing each state to establish its own standards and its own definition of proficiency, thereby generating 50 different definitions of the same concept.

to expect every child to reach that level by the year 2014, one can be quite certain it won’t happen. Even world leaders in education do not come close to reaching that goal. In 8th-grade math, for instance, only 73 percent of the students in Singapore are proficient by the NAEP definition of the word, despite the fact that Singapore has the highest-performing math students (see Figure 1).

That fact helps to clarify a basic dilemma that NCLB confronts as long as it continues to use the 2014 goal of full proficiency as its benchmark. Either the word “proficiency” will have to be dumbed down to mean little more than “basic” understanding of the given material, or a new way of measuring school performance must be introduced.

The simplest solution: use a high standard, such as the one employed by the NAEP, when holding students accountable for reaching full proficiency, if they are to receive an academic diploma, but hold schools accountable for achieving a high but realistic rate of student growth from one year to the next.

To Get Fewer Failing Schools, Set Easier Standards (Table 1)

States whose students score poorly on the National Assessment of Educational Progress (NAEP) tend to have more schools that fail according to No Child Left Behind (NCLB) guidelines. But the connection (correlation = -0.31) is not as tight as one might expect, as many states with low-performing students report fewer failing schools simply because they set lower standards (correlation = 0.44).

State	Percentage of Schools Not Making Adequate Yearly Progress (2005)	Percentage of Students Proficient on the NAEP (2005)*	Rigor of State Proficiency Standards (2005)
U.S. Median	23	32	C
Hawaii	66	22	B+
Florida	64	29	C
Rhode Island	56	28	B-
District of Columbia	55	10	C
Nevada	53	23	C
New Mexico	53	18	B-
South Carolina	53	29	A
Alabama	47	20	D-
New Hampshire	47	39	
Idaho	43	34	D
Massachusetts	43	45	A
North Carolina	42	32	D-
Alaska	41	29	D+
Indiana	40	32	C-
New Jersey	39	39	C
California	38	23	B
Missouri	35	30	A
Oregon	32	33	C
Illinois	27	30	C
Delaware	26	33	C-
Kentucky	26	27	C+
Colorado	25	35	D
Maryland	25	32	C
Ohio	24	36	C
Maine	23	35	A
Connecticut	20	37	C
New York	20	33	C
Washington	20	37	C
Wyoming	20	35	A
Pennsylvania	19	36	C
Virginia	19	36	D+
Georgia	18	26	D-
West Virginia	17	23	D-
Louisiana	16	20	C
South Dakota	16	36	D+
Arizona	14	25	D+
Vermont	14	39	
Minnesota	13	41	
Texas	13	31	D+
Utah	13	32	D+
Michigan	12	32	C-
Mississippi	11	17	D-
Iowa	9	34	D+
Kansas	9	37	C-
North Dakota	9	37	C
Montana	7	37	C+
Tennessee	7	25	F
Oklahoma	3	25	F
Wisconsin	2	36	C-
Arkansas		28	B-
Nebraska		35	D

Correlation = -0.31

Correlation = 0.44

Note: Percentage of students proficient on the NAEP is equal to the average of the state's 4th- and 8th-grade math and reading proficiency levels.

SOURCES: National Center for Education Statistics, State Education Reforms website (nces.ed.gov/programs/statereform/), Table 1.6; National Center for Education Statistics, National Assessment of Educational Progress; Paul E. Peterson and Frederick M. Hess, "Keeping an Eye on State Standards," *Education Next*, Summer 2006

Accountability

Who should be held accountable? That question brings me to NCLB's final distortion: exactly who or what is being held responsible. In ordinary language, only individuals, not entities such as schools, can be held accountable. We hold drivers, not cars, responsible for accidents. Or, if cars are faulty, we hold responsible those who made them. But under NCLB, only entities (schools, school districts, states), not students, teachers, or administrators, are held responsible for what is happening. To fix the NCLB accountability system, we need to find ways of holding accountable the individuals, that is, the students and teachers, who are involved in the education process.

At one time, student promotion to the next grade was conditional on performance, and graduation from high school depended on learning a specific body of material. Gradually, it has become standard practice to promote virtually all students from one grade to the next, regardless of whether they have learned the material. Such practices are justified on the grounds that holding a child back for poor performance only undermines self-esteem and aggravates learning problems. Minimal high-school graduation requirements are similarly justified on the grounds that having a diploma is better than not receiving one, regardless of what is learned.

Recently, some cities and states have introduced policies that return to more traditional practices. The results have been surprisingly promising. In Florida, the performance of 3rd graders jumped the first year they were expected to pass a test if they were to move on to 4th grade (see "Getting Ahead by Staying Behind," *research*, Spring 2006). Those held back benefit from being required to repeat the 3rd grade. In Massachusetts, the expectation that students pass a 10th-grade test if they are to graduate from high school spiked student performance the first year the law was introduced, with continuing gains in subsequent years. Internationally, Ludger Woessmann has shown that students score higher in countries that require students to perform well on comprehensive examinations than in countries that, like the United States, have no such expectations (see "Crowd Control," *research*, Summer 2003).

Teachers and administrators should be held accountable as well. Once the other elements of a well-designed accountability system have been

put in place, it is reasonable to hold teachers accountable for student learning. As Thomas Kane and his colleagues have shown (see “Photo Finish,” *research*, Winter 2007), the best measure of teacher quality in any given year is how much students learned from that same teacher the preceding year. The research simply confirms what every school child knows: certain teachers are consistently effective, while others are not.

Once the information is available to track student progress from one year to the next, one can identify the classrooms in which the most, and least, learning is taking place. That information can be used to reward the high performers and to counsel the low performers, who should be dismissed if they remain consistently ineffective classroom teachers. Of course, any teacher can have a bad year, and any accountability system may make an error, so all personnel decisions must be made by administrators who are fully informed of particular circumstances. But until teachers are held responsible for the performance of their students, it is unlikely that accountability systems will prove effective.

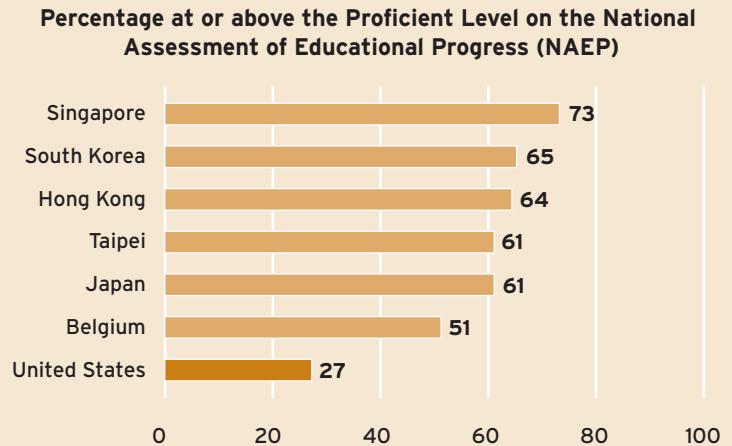
Finally, an effective accountability system requires strong administrative leaders, who should be held responsible for the learning gains realized at their school.

The Political Problem

It is rumored that influential interest groups in Washington and key members of Congress are considering many of the changes I have proposed in this essay. Let us hope so, but one should not be optimistic about the outcome of the legislative process in the absence of strong, sustained public support for reforms along these lines. The defects in NCLB, as originally written, are not accidental. The law took the form that it did because Congress navigated among powerful political interests—those of unions, suburbanites, state and local education officials, and other interested parties. Had teachers been held accountable, union opposition would have blocked the law’s enactment. Had states not been given the option to set any standard they wanted, many state and local officials would have balked at excessive federal control. Had states been required to put in place a data collection system that tracked student performance over time, privacy fanatics would have insisted that every child had a right not to be known, even to those responsible for the child’s education. Had every school been measured for growth in student performance, many a suburban district (as well as its board and superintendent) would have been found wanting.

Unrealistic Expectations or Low Standards (Figure 1)

Using the NAEP proficiency standards for 8th-grade math, the percentage of students scoring at or above proficiency in countries with the highest-performing students is much higher than in the United States, but is still a long way from 100 percent proficiency.



Note: For countries other than the U.S., percentages have been estimated by linking the National Assessment for Educational Progress (NAEP) to student performance on the Trends in International Mathematics and Science Study (TIMSS).

SOURCE: Gary W. Phillips, “Linking NAEP Achievement Levels to TIMSS,” (American Institutes for Research, 2007)

Had students been held accountable, groups of students and parents would have raised strenuous objections.

In 2001, lawmakers displayed sheer political genius when they came up with a law that could be sold to the public as an egalitarian policy that would leave no child behind. By insisting that every child reach a minimum level of performance in 2014 (well beyond the political lifetime of many of the key decisionmakers), the law left most parts of the educational system untouched and limited the rigors of accountability to the schools that served the most challenging populations. If that was politically shrewd, it is educationally problematic. The best and the brightest were given a pass. Meanwhile, excellent schools serving the most challenged, whether charter schools or high-quality inner-city public schools, were placed at the greatest risk of being called failures, even when they were successes.

None of this can be altered without regrinding and polishing the lens through which NCLB’s accountability light shines on America’s schools. If we cannot soon come to believe what we are shown, the whole microscope will be tossed into history’s dustbin.

Paul E. Peterson is professor of government at Harvard University and a senior fellow at the Hoover Institution. He serves as editor-in-chief of Education Next.