

# Conjuring Cut Scores

## *How It Distorts Our Picture of Student Achievement*

By Chester E. Finn, Jr., and Michael J. Petrilli

**N**o Child Left Behind made many promises, one of the most important of them being a pledge to Mr. and Mrs. Smith that they would get an annual snapshot of how their little Susie is doing in school. Mr. and Mrs. Taxpayer would get an honest appraisal of how their local schools and school system are faring. Ms. Brown, Susie's teacher, would get helpful feedback from her pupils' annual testing data. And the children themselves would benefit, too.

So far so good; these are the ideas that underpin 20 years of sensible education reform. But let's return to little Susie Smith and whether the information coming to her parents and teachers is truly reliable and trustworthy. This fourth-grader lives in suburban Detroit, and her parents get word that she has passed Michigan's state test. She's "proficient" in reading and math. Mr. and Mrs. Smith understandably take this as good news; their daughter must be "on grade level" and on track to do well in later grades of school, maybe even go to college.

Would that it were so. Unfortunately, there's a lot that Mr. and Mrs. Smith, and Susie's teachers, don't know. They don't know that Michigan set its "proficiency passing score"—the score a student must attain in order to pass the test—among the lowest in the land. So Susie may be "proficient" in math in the eyes of Michigan edu-

cation bureaucrats, but she still could have scored worse than five-sixths of the other fourth-graders in the country. Susie's parents and teachers also don't know that Michigan has set the bar particularly low for younger students, such that Susie is likely to fail the state test by the time she gets to sixth grade—and certainly when she reaches eighth grade—even if she makes regular progress every year. And they also don't know that "proficiency" on Michigan's state tests has little meaning outside the Wolverine State's borders; if Susie lived in California or Massachusetts or South Carolina, she would have missed the "proficiency" cut-off by a mile.

Mr. and Mrs. Smith know that little Susie is "proficient." What they don't know is that "proficient" doesn't mean much. This is the proficiency illusion.

Standards-based education reform is in deeper trouble than we knew, both the Washington-driven, No Child Left Behind version and the older versions that most states undertook for themselves in the years since *A Nation at Risk* (1983) and the Charlottesville education summit (1989). It's in trouble for multiple reasons. Foremost among these: On the whole, states do a bad job of setting (and maintaining) the standards that matter most—those that define student proficiency for purposes of NCLB and states' own results-based accountability systems.

We've known for years that there's a problem with many states' academic standards—the aspirational statements, widely available on state Web sites, of what students at various grade levels should know and be able to do in particular subjects. Fordham has been appraising state standards since 1997, as has the American Federation of Teachers. A few states do a super job, yet our most recent comprehensive review (2006) found that "two-thirds of schoolchildren in America attend class in states with mediocre (or worse)

---

*Chester E. Finn, Jr., is president of the Thomas B. Fordham Institute, where Michael J. Petrilli is vice president for national programs and policy. This article is adapted from the foreword to The Proficiency Illusion, published by the Thomas B. Fordham Institute and the Northwest Evaluation Association and online at [http://edexcellence.net/doc/The\\_Proficiency\\_Illusion.pdf](http://edexcellence.net/doc/The_Proficiency_Illusion.pdf).*

expectations for what their students should learn.” Instead of setting forth a coherent sequence of skills and content that comprise the essential learnings of a given subject—and doing so in concrete, cumulative terms that send clear signals to educators, parents, and policymakers—many states settle for nebulous, content-lite standards of scant value to those who are supposed to benefit from them.

That’s a serious problem, striking at the very heart of results-based educational accountability. If the desired outcomes of schooling aren’t well stated, what is the likelihood that they will be produced? If teachers, textbook writers, and curriculum planners don’t get decent guidance from state education leaders, and parents have no clarity regarding what their daughters and sons are expected to learn, the odds are slim that school results will be strong.

Yet that problem turns out to be just the opening chapter of an alarming tale. For we also understood that, when it comes to the real traction of standards-based education reform, a state’s posted academic standards aren’t all that matters. They describe the desired outcomes of school, but what is at least as apt to drive actual classroom behavior—and the public’s understanding of how its children and its public education system are doing—is the passing level (a.k.a. the “cut score”) on the state’s actual tests. At day’s end, voters and those they elect to office are likely to define educational success by how many kids pass the state test and how many fail. No matter what the aspirational statements the state sets forth as its educational goals, the rubber meets the road when the testing program determines that Susie (or Michelle or Caleb or Tyrone or Rosa) is or is not “proficient” as determined by her scores on state assessments.

The advent of high-stakes testing in general, and No Child Left Behind in particular, have underscored this. When NCLB asks whether a school or district is making “adequate yearly progress” in a given year, what it’s really asking is whether an acceptable number of children scored at (or above) the “proficient” level as specified on the state’s tests—and how many failed to do so.

We set out, therefore, to learn more about the meaning of “proficiency” as represented by cut scores on state tests, in particular the scores used for NCLB accountability purposes. Yet the context for this examination is our keen awareness of the interrelatedness of a state’s cut scores and its academic standards. If the latter are nebulous, inscrutable, light on content—or kitchen-sink like in their naïve expectation that kids will learn everything about everything—it’s folly to expect that fooling around with test cut scores will solve the problem. A state needs to be as smart about the one as about the other—a double play that far too few jurisdictions succeed at.

## What We Asked

In the present study, we set out to determine whether states’ “cut scores” on their tests are high, low, or in between. Whether they’ve been rising or falling (i.e., whether it’s been getting harder or easier to pass the state test). And

whether they’re internally consistent as between, say, reading and math, or fourth and eighth grade.

One cannot answer such questions by examining academic standards alone. A state may have awesome standards even as its test is easy to pass. It could have dreadful standards, yet expect plenty from its test-takers—causing serious consternation among teachers, curriculum directors, and others charged with preparing youngsters for such exams. It might have standards that are carefully aligned from one grade to the next, yet be erratic in setting its cut scores.

To examine states’ cut scores carefully, you need a yardstick external to the state itself, something solid and reliable that state-specific results and trends can be compared with. The Northwest Evaluation Association (NWEA) has both a long-lived, rock-steady scale and a computerized assessment called the “Measure of Academic Progress” (MAP) that is used for diagnostic and accountability purposes by schools and school systems in many states. Not all states, to be sure, but it turns out that in a majority of them (26, to be precise), enough kids participate in MAP and the state assessment to allow for useful comparisons to be made and analyses performed.

The NWEA experts (see sidebar, p. 23) accepted the challenge. The three key questions they sought to answer are straightforward and crucial:

- How hard is it to pass each state’s tests?
- Has it been getting easier or harder since enactment of NCLB?
- Are a state’s cut scores consistent from grade to grade? That is, is it as hard (or easy) for a 10-year-old to pass the state’s fourth-grade tests as it is for a 14-year-old to pass the same state’s eighth-grade tests?

## What We Learned

The findings of this inquiry are sobering, indeed alarming. We see, with more precision than previous studies, that “proficiency” varies wildly from state to state, with “passing scores” ranging from the 6th percentile to the 77th. Over the past few years, twice as many states have seen their tests become easier in at least two grades as have seen their tests become more difficult. (Though we note, with some relief, that most state tests have maintained their level of difficulty—such as it is—over this period.) And, only a handful of states peg proficiency expectations consistently across the grades, with the vast majority setting thousands of little Susies up to fail by middle school by aiming precipitously low in elementary school.

What does this mean for educational policy and practice? What does it mean for standards-based reform in general and NCLB in particular? It means big trouble—and those who care about strengthening U.S. K-12 education should be furious. There’s all this testing—too much, surely—yet the testing enterprise is unbelievably slipshod. It’s not just that results vary, but that they vary almost ran-

*(Continued on page 28)*

# The Proficiency Illusion

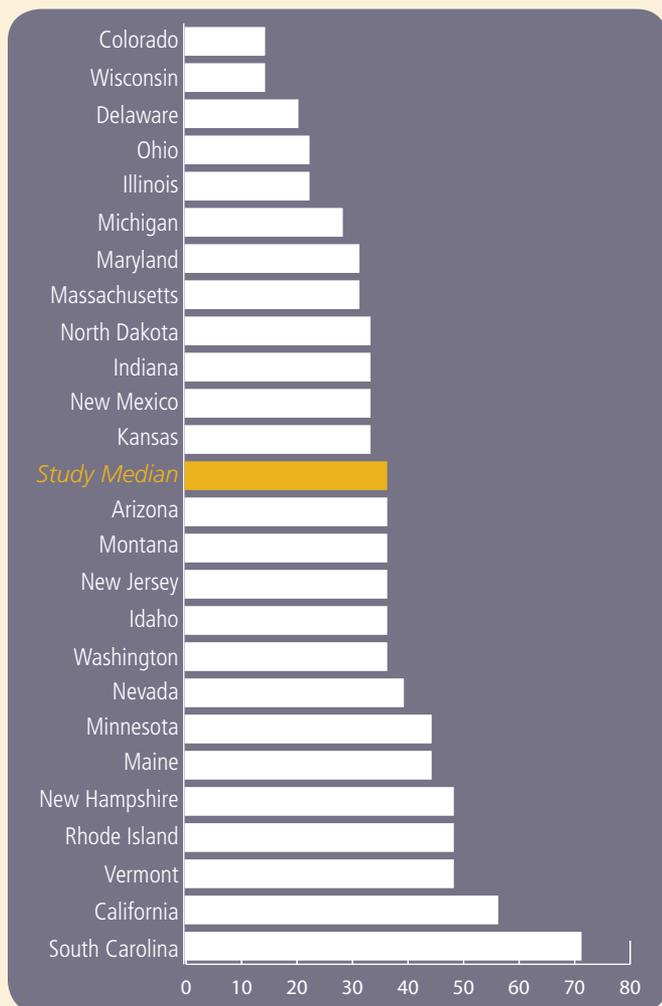
By John Cronin, Michael Dahlin, Deborah Adkins, and G. Gage Kingsbury

At the heart of NCLB is the call for all children to become “proficient” in reading and math by 2014. Yet that law expects each state to define proficiency as it sees fit and to design its own tests. Serious problems have arisen as a result. We summarize four of them here. For the full results of our study, read *The Proficiency Illusion* online at [http://edexcellence.net/doc/The\\_Proficiency\\_Illusion.pdf](http://edexcellence.net/doc/The_Proficiency_Illusion.pdf).

## I. State tests vary greatly in their difficulty.

To compare how difficult it is to score proficient on states’ tests, we needed to convert the states’ proficiency cut scores to a single common scale. Our Measures of Academic Progress (MAP), a computerized adaptive test, provided that scale; having done extensive norming studies with MAP, we were able to estimate the percentile scores on MAP corresponding to each state’s cut scores. As Figure 1 shows, we found that eighth-grade reading cut scores ranged from the 14th percentile (Colorado\*) to the 71st percentile (South Carolina).

**Figure 1** – Grade 8 estimated reading proficiency cut scores for 2006 (ranked by MAP percentile)



## II. Differences in state proficiency cut scores can be seen in the rigor of the assessment items.

The differences in proficiency cut scores are not numerical artifacts. They represent real differences in the assessment items that students are expected to be able to answer. To illustrate this point, we selected several states to represent the range of proficiency cut scores used for grade 4 reading and math. We then extracted questions from the MAP item pool that were equivalent in difficulty to the proficiency cut scores for each of these states. Using the MAP items shown below, we can compare what “proficiency” requires in reading and math in several different states.

To make comparison easier, all the reading items focused on a single skill that is commonly required in all state standards: the ability to distinguish fact from opinion. Almost all reading curricula have introduced this concept prior to fourth grade.

For mathematics, we extracted examples of items from the MAP item bank with difficulty ratings equivalent to five states’ proficiency cut scores in algebraic concepts. None of the items requires computational abilities that would be beyond the scope of a typical grade 4 curriculum.

**Reading Exhibit 1** – Grade 4 MAP item with difficulty equivalent to Colorado’s proficiency cut score (scale score 187, 11th percentile)

**Alec saw Missy running down the street. Alec saw Paul running after Missy. Paul was yelling, “Missy, stop! Wait for me!”**

### What do we know for sure?

- A. Missy is Paul’s big sister, and she is mad at him.
- B. Paul is mad at Missy and is chasing her down the street.
- C. Alec saw Paul running after Missy and calling for her to wait.**
- D. Alec tried to stop Missy because Paul wanted to talk to her.

Almost all fourth-graders answer this item correctly. It contains a very simple passage and asks the student to identify the facts in the passage without making an inference. The student does not have to understand terms like “fact” or “opinion” to correctly answer the question.

\* Colorado currently reports the state’s “partially proficient” level of academic performance on its state test as “proficient” for NCLB purposes, while using the higher “proficient” level for internal state evaluation purposes. In effect, Colorado has two standards: an easier standard for NCLB, and a harder standard for internal state use. For purposes of fairly comparing Colorado to other states, we used its NCLB-reported standard. Consequently, all subsequent references to “proficient” or “proficiency” in Colorado should be understood as referring to the NCLB-reported standard.

**Reading Exhibit 2** – Grade 4 MAP item with difficulty equivalent to Wisconsin’s proficiency cut score (scale score 191, 16th percentile)

**Which sentence tells a fact, not an opinion?**

- A. Cats are better than dogs.
- B. Cats climb trees better than dogs.**
- C. Cats are prettier than dogs.
- D. Cats have nicer fur than dogs.

This item is also quite easy for most fourth-graders and does not require reading a passage. It does introduce the terms fact and opinion, however, and some of the distinctions between fact and opinion are subtle. For example, some children may believe that the differences in cat and dog fur are fact.

**Reading Exhibit 3** – Grade 4 MAP item with difficulty equivalent to North Dakota’s proficiency cut score (scale score 199, 29th percentile)

**Summer is great! I’m going to visit my uncle’s ranch in July. I will be a really good rider by August. This will be the best vacation ever!**

**Which sentence is a statement of fact?**

- A. Summer is great!
- B. I’m going to visit my uncle’s ranch in July.**
- C. I will be a really good rider by August.
- D. This will be the best vacation ever!

Most fourth-graders answer this item correctly. The differences between fact and opinion in this item are considerably more subtle than in the prior item. For example, many fourth-graders are likely to believe that “Summer is great!” is not a matter of opinion.

**Reading Exhibit 4** – Grade 4 MAP item with difficulty equivalent to California’s proficiency cut score (scale score 204, 43rd percentile)

**The entertainment event of the year happens this Friday with the premiere of Grande O. Partie’s spectacular film Bonzo in the White House. This movie will make you laugh and cry! The acting and directing are the best you’ll see this year. Don’t miss the opening night of this landmark film—Bonzo in the White House. It will be a classic.**

**What is a fact about this movie?**

- A. It is the best film of the year.
- B. You have to see it Friday.
- C. It opens this Friday.**
- D. It has better actors than any other movie.

Just over half of fourth-graders from the MAP norm group answer this item correctly. The question requires the student to navigate a longer passage with more sophisticated vocabulary.

Indeed, the student has to know or infer the meaning of “premiere” to answer the question correctly.

**Reading Exhibit 5** – Grade 4 MAP item with difficulty equivalent to Massachusetts’s proficiency cut score (scale score 211, 65th percentile)

**Read the excerpt from “How Much Land Does a Man Need?” by Leo Tolstoy.**

So Pahom was well contented, and everything would have been right if the neighboring peasants would only not have trespassed on his wheatfields and meadows. He appealed to them most civilly, but they still went on: now the herdsmen would let the village cows stray into his meadows, then horses from the night pasture would get among his corn. Pahom turned them out again and again, and forgave their owners, and for a long time he forbore to prosecute anyone. But at last he lost patience and complained to the District Court.

**What is a fact from this passage?**

- A. Pahom owns a vast amount of land.
- B. The peasant’s intentions are evil.
- C. Pahom is a wealthy man.
- D. Pahom complained to the District Court.**

This item is clearly the most challenging to read (it is Tolstoy after all), and the majority of fourth-graders in the NWEA norm group got it wrong. The passage is long relative to the others and contains very sophisticated vocabulary. At least three of the options identify potential facts in the passage that have to be evaluated.

**Math Exhibit 1** – Grade 4 MAP item with difficulty equivalent to Colorado’s proficiency cut score (scale score 191, 8th percentile)

**Tina had some marbles. David gave her 5 more marbles. Now Tina has 15 marbles. How many marbles were in Tina’s bag at first?**

**What is this problem asking?**

- A. How many marbles does Tina have now?
- B. How many marbles did David give to Tina?
- C. Where did Tina get the marbles?
- D. How many marbles was Tina holding before David came along?**
- E. How many marbles do Tina and David have together?

This item, which reflects the Colorado NCLB proficiency cut score, is easily answered by most fourth-graders.

ers. It requires that students understand the basic concept of addition and find the right question to answer, although students need not actually solve the problem.

**Math Exhibit 2** – Grade 4 MAP item with difficulty equivalent to Illinois’s proficiency cut score (scale score 197, 15th percentile)

**Marissa has 3 pieces of candy. Mark gives her some more candy. Now she has 8 pieces of candy. Marissa wants to know how many pieces of candy Mark gave her.**

**Which number sentence would she use?**

- A.  $3 + 8 = ?$
- B.  **$3 + ? = 8$**
- C.  $? \times 3 = 8$
- D.  $8 + ? = 3$
- E.  $? - 3 = 8$

This item, reflecting the Illinois cut score, is slightly more demanding but is also easily answered by most fourth-graders. It requires the student to go beyond understanding the question to setting up the solution to a one-step addition problem.

**Math Exhibit 3** – Grade 4 MAP item with difficulty equivalent to Texas’s proficiency cut score (scale score 205, 34th percentile)

**Chia has a collection of seashells. She wants to put her 117 shells into storage boxes. If each storage box holds 9 shells, how many boxes will she use?**

**Which equation best represents how to solve this problem?**

- A.  $9 - 117 = ?$
- B.  $9 \div 117 = ?$
- C.  $117 \times 9 = ?$
- D.  $117 + 9 = ?$
- E.  **$117 \div 9 = ?$**

This item, at a difficulty level equivalent to the Texas cut score, is answered correctly by most fourth-graders but is harder than the previous two. The student not only must be able to set up the solution to a simple problem, but must also know how to frame a division problem in order to answer the question correctly.

**Math Exhibit 4** – Grade 4 MAP item with difficulty equivalent to California’s proficiency cut score (scale score 212, 55th percentile)

**$8 + 9 = 10 + ?$**

- A. 6
- B. 9
- C. 17
- D. **7**
- E. 6

When the proficiency expectations in grade 4 mathematics range from setting up simple addition problems to solving complex, multi-step multiplication problems, then meeting these expectations achieves no real equity.

Most fourth-grade students in the MAP norm group do not answer this question correctly. The more advanced concept of balance or equivalency within an equation is introduced in this item. This concept is fundamental to algebra and makes this much more than a simple arithmetic problem. The student must know how to solve a problem by balancing the equation.

**Math Exhibit 5** – Grade 4 MAP item with difficulty equivalent to Massachusetts’s proficiency cut score (scale score 220, 77th percentile)

**The rocket car was already going 190 miles per hour when the timer started his watch. How fast, in miles per hour, was the rocket car going seven minutes later if it increased its speed by 15 miles per hour every minute?**

- A. 205
- B. **295**
- C. 900
- D. 1330
- E. 2850

This is obviously the most demanding item of the set and is not answered correctly by most fourth-graders within the MAP norm group. The student must understand how to set up a multiplication problem using either a two-step equation,  $190 + (7 \times 15) = ?$ , or a multi-step equation,  $190 + (15+15+15+15+15+15+15) = ?$

These examples from reading and mathematics make it apparent that the states we studied lack a shared concept of proficiency. Indeed, their expectations are so diverse that they risk undermining a core objective of NCLB—to advance educational equality by ensuring that all students achieve their states’ proficiency expectations. When the proficiency expectations in grade 4 mathematics range from setting up simple addition problems to solving complex, multi-step multiplication problems, then meeting these expectations achieves no real equity. The reading examples, too, show that “proficiency” by no means indicates educational equality. A student who can navigate the California or Massachusetts reading requirements has clearly achieved a much different level of competence than has one who just meets the Colorado or Wisconsin proficiency standard.

The proficiency expectations have a profound effect on the delivery of instruction in many states. Because of the consequences associated with failure to make adequate yearly progress (AYP), there is evidence that instruction in many classrooms and schools is geared toward ensuring that students who perform near the proficiency bar pass the state test (Neal and Whitmore-Schanzenback, 2007). In Illinois, for example, this is apt to mean that some classrooms will place greater emphasis on understanding simple math problems like the one in Math Exhibit 2, while California and Massachusetts students are working with algebraic concepts of much greater sophistication, such as those in Math Exhibits 4 and 5.

### III. Standards for mathematics are generally more difficult to meet than those for reading.

Two sample items (Reading Exhibit 6 and Math Exhibit 6) illustrate the difference in difficulty between the reading and math standards.

**Reading Exhibit 6** – Grade 8 MAP item with difficulty equivalent to Massachusetts’s proficiency cut score (scale score 216, 31st percentile)

#### Read the passage.

Katya’s eyes adjusted to the dimness. She could tell that someone had once inhabited this place. She noticed markings on the walls, and she knew they would be a significant part of her archaeological study. There were jagged lines of lightning and stick figures.

**What story element has the author developed within this passage?**

- A. theme
- B. plot
- C. conflict
- D. setting

This reading item has the same difficulty as the Massachusetts grade 8 reading cut score and is answered correctly by the vast majority of eighth-graders. The passage is not complex, and students who are familiar with the literary concept of setting will answer it correctly.

**Math Exhibit 6** – Grade 8 MAP item with difficulty equivalent to Massachusetts’s proficiency cut score (scale score 242, 67th percentile)

**Maria has \$5.00 more than Joseph. Together they have \$37.50. Which of these equations would you use to find the amount of money Joseph has?**

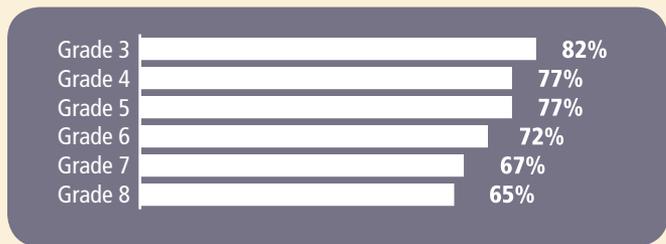
- A.  $j + (5 \times j) = \$37.50$
- B.  $j + (j \div 5) = \$37.50$
- C.  $5 \times j = \$37.50 + j$
- D.  $2 \times (j + 5) = \$37.50$
- E.  $j + j + 5 = \$37.50$

This item has the same difficulty as the Massachusetts mathematics proficiency standard and is missed by the majority of eighth-grade students in the NWEA norm group. The question is a multi-step problem and addresses a concept commonly found in Algebra I. Although the items in these two exhibits come from different disciplines, we know that the mathematics item is empirically more difficult than the reading item because far fewer eighth-graders within the NWEA norm group successfully answer the math item than the reading item.

### IV. Reading and math tests in the upper grades are generally more difficult to pass than those in earlier grades (even after taking into account obvious differences in student development and curriculum content).

The experience of Minnesota illustrates some of the issues that may be encountered when a proficiency standard is not calibrated across grades. Imagine that you are a parent viewing the results of the Minnesota Comprehensive Assessment – series II (MCA-II) in the newspaper. Figure 2 shows the spring 2006 statewide reading results.

**Figure 2** – Proportion of students scoring proficient or better on the Minnesota Comprehensive Assessment in reading (MCA-II), 2006



A parent interpreting these results would probably assume that third-graders in the state were doing far better than their peers in eighth grade. They might be concerned about the “deteriorating” performance in grades 7 and 8. Indeed, newspaper editorials, talk radio, and online discussions might identify a “crisis in the middle grades” and

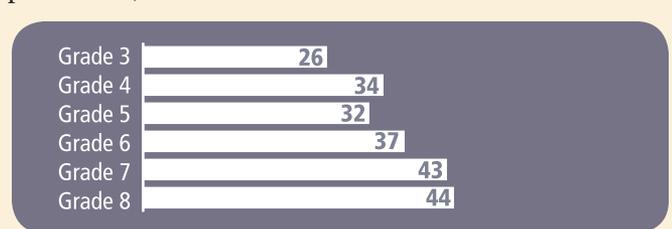
call for radical changes in the curriculum and organization of middle schools. Gradually, Minnesotans might come to believe that the discrepant results are a product of slumping middle school students and their lackluster teachers; meanwhile, they might believe that all is well in their elementary schools. Yet it is not clear that either inference would be warranted. If we look at Minnesota students' performance on the 2005 NAEP test in reading, shown in Table 1, we see that fourth- and eighth-graders perform about the same on their respective tests (albeit far below state-reported performance). Why then the grade-to-grade gap in performance on the Minnesota state assessment?

**Table 1** – Minnesota's performance on the 2005 NAEP in reading

	Grade 4	Grade 8
Percentage performing "proficient" or above	38%	37%

The answer lies in understanding that the difference in reported performance is really a function of differences in the difficulty of the cut scores and not actual differences in student performance. If we look at Figure 3, which shows the NWEA percentile ranks associated with the MCA-II proficiency cut scores for reading, we see that the third-grade cut score was estimated at the 26th percentile, meaning that 26 percent of the NWEA norm group would not pass a standard of this difficulty. By extension, 74 percent of NWEA's norm group would pass this standard. The proficiency cut score for eighth-grade, however, was estimated at the 44th percentile. This more difficult standard would be met by only 56 percent of the NWEA norm population.

**Figure 3** – Reading proficiency cut scores by grade in MAP percentiles, 2006



Now we can see that the difference in reported performance reflects differences in the difficulty of the cut scores rather than any genuine differences in student performance. According to our estimates, because of the difference in difficulty of the standards, about 18 percent fewer students would pass the Minnesota test in eighth grade than passed in third ( $74\% - 56\% = 18\%$ ). And in fact the Minnesota results show that 17 percent fewer eighth-graders passed the MCA-II than third-graders.

These data make the problem obvious. Poorly calibrated standards create misleading perceptions about the performance of schools and children. They can lead parents, educators, and others to conclude that younger

Poorly calibrated standards create misleading perceptions. Younger students who might need help do not get resources because they have passed the state tests, while schools serving older students may make drastic changes in their instructional programs to fix deficiencies that may not actually exist.

pupils are safely on track to meet standards when that is not the case. They can also lead policymakers to conclude that programs serving older students have failed because proficiency rates are lower for these students, when in reality, those students may be performing no worse than their younger peers. And conclusions of this sort can encourage unfortunate misallocations of resources. Younger students who might need help now if they are to reach more difficult standards in the upper grades do not get those resources because they have passed the state tests, while schools serving older students may make drastic changes in their instructional programs in an effort to fix deficiencies that may not actually exist.

Bringing coherence to the standards by setting initial standards that are calibrated to the same level of difficulty can help avoid these problems. If states begin with calibrated standards, then they know that between-grade differences in performance represent changes in the effectiveness of instruction, rather than in the difficulty of the standard. Armed with this knowledge, schools can make better use of resources to address weaknesses in their programs and can build on strengths.

*John Cronin is a research specialist with the Northwest Evaluation Association (NWEA), where Michael Dahlin and Deborah Adkins are both research associates. G. Gage Kingsbury is NWEA's chief research and development officer. This sidebar is adapted from the "National Findings" section of The Proficiency Illusion, published by the Thomas B. Fordham Institute and the NWEA.*

(Continued from page 22)

domly, erratically, from place to place and grade to grade and year to year in ways that have little or nothing to do with true differences in pupil achievement. America is awash in achievement “data,” yet the truth about our educational performance is far from transparent and trustworthy. It may be smoke and mirrors. Gains (and slip-pages) may be illusory. Comparisons may be misleading. Apparent problems may be nonexistent or, at least, mis-stated. The testing infrastructure on which so many school reform efforts rest, and in which so much confidence has been vested, is unreliable—at best. We believe in results-based, test-measured, standards-aligned accountability systems. They’re the core of NCLB, not to mention earlier (and concurrent) systems devised by individual states. But it turns out that there’s far less to trust here than we, and you, and lawmakers have assumed. Indeed, the policy implications are sobering. First, we see that Congress erred big-time when NCLB assigned each state to set its own standards and devise and score its own tests. No matter what one thinks of America’s history of state primacy in K-12 education, this study underscores the folly of a big modern nation, worried about its global competitiveness, nodding with approval as Colorado sets its eighth-grade reading passing level at the 14th percentile while South Carolina sets its at the 71st percentile. A youngster moving from middle school in Boulder to high school in Charleston would be grievously unprepared for what lies ahead. So would a child moving from third grade in Detroit to fourth grade in Albuquerque.

Moreover, many states are internally inconsistent, with more demanding expectations in math than in reading and with higher bars in seventh and eighth grade than in third and fourth (though occasionally it goes the other way), differences that are far greater than could be explained by conscious curricular decisions and children’s levels of intellectual development. This means that millions of parents are being told that their 8- and 9-year-olds are doing fine in relation to state standards, only to discover later that (assuming normal academic progress) they are nowhere near being prepared to succeed at the end of middle school. It means that too little is being expected of millions of younger kids and/or that states may erroneously think their middle schools are underperforming. And it means that Americans may wrongly think their children are doing better in reading than in math—when in fact, less is expected in the former subject.

NCLB does not seem to be fueling a broad “race to the bottom” in the sense of many states lowering their cut scores in order to be able to claim that more youngsters are proficient. But, this study reveals that, in several instances, gains on state tests are not being matched by gains on the Northwest Evaluation Association (NWEA) test, raising questions about whether the state tests are becoming easier for students to pass. The NWEA’s experts describe this as a “walk to the middle,” as states with the highest standards were the ones whose estimated passing scores dropped the most.

---

If the desired outcomes of schooling aren’t well stated, if teachers, textbook writers, and curriculum planners don’t get decent guidance from state education leaders, and parents have no clarity regarding what their daughters and sons are expected to learn, the odds are slim that school results will be strong.

---

NCLB aside, what is the meaning of a “standard” if it changes from year to year? What is the meaning of measurable academic gains—and “adequate yearly progress”—if the yardstick is elastic?

Standards-based reform hinges on the assumption that one can trust the standards, that they are stable anchors to which the educational accountability vessel is moored. If the anchor doesn’t hold firm, the vessel moves—and if the anchor really slips, the vessel can crash against the rocks or be lost at sea.

That, we now see clearly, is the dire plight of standards-based reform in the U.S. today.

What to do? It’s crazy not to have some form of national standards for educational achievement—stable, reliable, cumulative, and comparable. That doesn’t mean Uncle Sam should set them, but if Uncle Sam is going to push successfully for standards-based reform, he cannot avoid the responsibility of ensuring that they get set. NCLB edition 1.0 didn’t do that and, so far as one can read the policy tea-leaves today, version 2.0 won’t either. If the feds won’t act, the states should, by coming together to agree to common, rational, workable standards (as most states have been doing with regard to high-school graduation rates). □

#### Article and Sidebar References

Finn, Jr., C.E., Petrilli, M.J., and Julian, L. (2006). *The State of State Standards 2006*. Washington, D.C.: Thomas B. Fordham Institute.

Neal, D. and Whitmore-Schanzenbach, D. (2007). *Left Behind by Design: Proficiency Counts and Test-Based Accountability*. NBER Working Paper No. W132930.