PIONEER INSTITUTE
PUBLIC POLICY RESEARCH

POLICY BRIEF

Setting Academic Performance Standards: MCAS vs. PARCC

by Richard P. Phelps

Common Core proponents have managed to convince most journalists, policymakers, and other opinion leaders that the Common Core standards are higher, deeper, tougher, more challenging, and more rigorous than their antecedents. This is, arguably, their greatest accomplishment.

Ask those journalists, policymakers, and other opinion leaders to identify the aspects of the Common Core standards that make them superior, however, and one is likely to hear only more marketing doublespeak about "problem solving", "deeper learning", "critical thinking", or the like. Most supporters of the Common Core do not understand how the Common Core standards or tests might be better. They simply assume that they must be because they have been told so often that they are.

Large sums from private foundations and the U.S. Education Department have been employed to sell Common Core to the U.S. public.[1] It is unfortunate that funds were not directed toward educating the public about how standards actually work to raise student academic achievement.

Their two-part nature—comprising both content and performance—is most fundamental for such an understanding. The Common Core State Standards (CCSS) document itself comprises only the pretend-content part—listing topics in math, and skills in English language arts that teachers should cover or develop over the course of a student's school career. By themselves, however, these and most other sets of content standards amount to little more than a plan. Indeed, absent any sort of monitoring or evaluation, teachers may feel free to ignore them.

The second part of the structure—the performance standards, or the tests based on the content standards—is essential for standards to be effective. Performance standards tell us how well students master the content via letter grades, test scores, or other types of evaluative feedback.

Richard P. Phelps is editor or author of four books: *Correcting Fallacies about Educational and Psychological Testing* (APA, 2008/2009); *Standardized Testing Primer* (Peter Lang, 2007); *Defending Standardized Testing* (Psychology Press, 2005); and *Kill the Messenger* (Transaction, 2003, 2005), and founder of the Nonpartisan Education Review (http://nonpartisaneducation.org).

Center for School Reform

November 2015

Though standards-induced academic achievement gains require both parts—content and performance—the two parts can be, and usually are, developed quite independently. Moreover, what makes a standards regime "higher", "tougher", or "rigorous" depends on the relationship between the two standards parts.

Content standards should not be considered rigorous, no matter how they read, if they can easily be ignored in the classroom. Performance standards should not be considered rigorous if the boundaries for performance levels, especially the pass/fail score, are set so low that every student passes, no matter how well or how poorly they have mastered the content.

Common Core advocates worked to convince the public that a hugely expensive new set of content standards was necessary to "raise" standards. Not true. Higher performance standards could have been accomplished easily, and virtually for free, in any state simply by raising the test score thresholds—the "cut scores"—that determine test results—who passes, who fails, whose performance is labeled "proficient", "advanced", "basic", or "below basic".

So, if the primary goal was to raise standards, why did we not just raise the cut scores for all the state tests, and avoid the gargantuan disruption effected by the Common Core Initiative? I would argue, that the primary goal was not to raise standards for, if it were, that is exactly what we would have done—simply raise the cut scores on most state tests until they matched the levels of the genuinely most rigorous (e.g., Massachusetts').

That simple solution to the alleged "low standards" problem, however, would not have afforded an opportunity to introduce the constructivist elements now embedded in Common Core instruction and in PARCC and SBAC testing.[2] A wholesale re-do of content and performance standards across the continent has afforded that opportunity.

"Standard Setting" (a.k.a., "cut score" or "passing score") conferences represent the final phase of a new test's development. Despite all the assurances by advocates that Common Core content standards, all by themselves, would raise student achievement, the necessary ingredient of performance standard setting has only in recent weeks begun for the PARCC and SBAC tests.

Despite all the public attention on academic standards, the performance-standard-setting process remains a mystery to many. The primary misconception is that setting cut scores is, or can be, somehow scientifically or empirically determined. It cannot be. The decision as to what will be considered "passing" or "proficient" is entirely a matter of choice.

Typically, "standard setting" (i.e., passing-score or cut-score setting) conferences are held after the first administration of a new test. Participating should be a few dozen current teachers, teacher educators, administrators, and enough content-area experts to outvote the content-clueless participants. They look at each and every test item and, with each, vote individually and by secret ballot. Their vote is an answer to a question that looks something like this: What percentage of students do you believe would be able to answer this question correctly? All the percentage estimates are then averaged for each test item.

The actual passing score for a test can be different depending on which items are used in a particular test administration, because each item has its own passing score. But, they are then adjusted to fit the score scale that is used publicly.

Then the conferees are shown the actual results from the test administration for each test item and asked if they would like to change their percentage estimates. The percentage estimates are again averaged. Typically, the cut scores are lowered, sometimes substantially. Test items that appear, at first glance by adults to be correctly answerable by most students often are correctly answered by only a few. The reasons for the low performance are unknown to the conference members. Was the test question more confusing or ambiguous than was apparent at first glance? Perhaps teachers are at fault because they do not understand the topic or how best to teach it.

Regardless, passing-score conferees inevitably face the decision—and it is an entirely arbitrary and

subjective decision—of how many students to pass, and to fail, based on the evidence provided. They may believe that students should have known the answer to a certain question, but when faced with evidence that they do not, they still must decide the students' fate.

They must choose. And, normally they choose based on the reality of how many will or will not pass, not based on how many they believe should or should not.

To "raise standards" all we as a society needed do was raise the cut scores on the tests aligned to the standards we already had. We didn't need to build from scratch an entirely new set of content standards, at enormous expense.

Some Common Core supporters have countered, however, that we need a new type of standards to buttress ill-defined "21st-century skills". The world is changing so fast that content knowledge gained in school will be outdated by the time students leave school. Instead, they need to "learn how to learn", adapt, think on their feet, etc. The 19th-century factory school model of rote memorization needs to be replaced, and so forth.[3]

Seldom mentioned is that the "19th-century factory model of rote memorization" no longer exists anywhere in North America, or that the dominant instructional model in the many countries killing us on international tests is mid-century modern—very much like what one found in the typical American classroom of the 1950s and 1960s.

Something is not necessarily better just because it is newer, and accumulating content knowledge is hardly frowned upon among our most respected professionals, such as doctors, lawyers, engineers, and scientists. Yes, they need to absorb new knowledge, but new knowledge only makes sense when they have already built a well-organized storehouse of past knowledge in which to place it.

Now that we have PARCC and SBAC, though, what will we do with them? Each testing consortia has only recently hosted its own passing score conference. What if conference attendees were asked to judge what proportion of students should be able to (rather than would be able to), in their judgment, answer the test items correctly? Perhaps then the conferees would prove their "higher level" of "toughness" and "rigor" by raising the cut scores higher than those for the old state tests. Will they be willing to do that, and declare vastly larger proportions of the U.S. student population failures or "below basic"? We shall see.

In practice, past state and local efforts to raise standards simply by raising the performance-standards bar were short-lived. It may seem reasonable to expect all students to reach a certain level of academic achievement at a certain age, especially when their same-age counterparts overseas have done it. But, when a majority of students fail to reach the new threshold, and are held back a grade or denied diplomas, our education system seizes up. The consequences are politically unsustainable, and the bar is lowered back again.

PARCC's primary goal is a single set of performance standards across all states. Its marketers insist that the threshold for all participating states will be at least as high as that for the current highest standard states. But, abundant experience suggests a different outcome: the PARCC performance standards will end up somewhere below the current average for all the participating states. Given Massachusetts's current perch near the top of the performance standards ranks, the Commonwealth's standards have farther to fall.

How far might that be? The U.S. Education Department has mapped state levels for "proficient" performance to the National Assessment of Educational Progress (NAEP) score scale.[4] For 2013, Massachusetts's "proficient" performance standard ranked second, third, fourth, and twenty-third in the nation, respectively, in 4th grade math and reading and 8th grade math and reading. By comparison, the average ranks for 11 PARCC states (as of August 2015) were, respectively, 27.0, 20.5, 25.3, and 25.1. (A rank of 25 lies right in the middle of the range of 50 states.) Massachusetts, then, can expect its current performance standards, among the highest in the country, to sink at least to the middle with PARCC—a regression toward the mean.

In test score terms, the drops from Massachusetts's current performance standards to the PARCC averages equal 26, 21, 24, and -2 score points for, respectively, 4th grade math and reading and 8th grade math and reading. Given the NAEP scale range that amounts to about a half-year drop in performance expectations in both 4th-grade subjects and 8th-grade math.

A February 2015 report of the pro-Common Core Massachusetts Business Alliance for Education (MBAE) unsurprisingly disagrees with this analysis, claiming that Common Core and PARCC will raise standards from the allegedly low level where MCAS lay. Lacking the time and space to deconstruct all of the report, I concentrate on the first "Summary of Findings" table (p. 5), which purports to answer the question, "Does the test identify students who are college- and career-ready? The table's list of points demeaning the MCAS begins:

> "The Proficient bar on the MCAS high school tests is set very low compared to all other indicators of students' college- and career-readiness."

The "proficient bar" on the MCAS high school test is not, was never intended to be, and should not be interpreted to be an indicator of college readiness. The MCAS high school test is administered to all Bay State students—both those intending to enroll in college and the many with no such intention. The MCAS high school test is a retrospectively focused standards-based achievement test, designed to measure how well students have mastered the material in the MCAS standards.[5] Retrospective achievement tests with stakes (e.g., one must pass to obtain a diploma) are legally limited to coverage of the subject matter taught. Courts have ruled that it is not fair to deny a student a diploma based on subject matter to which they were not exposed in school.[6]

Organizations that calculate college-readiness measures are typically those that develop a very different type of test that is labeled an "aptitude", "admission", or "readiness" test. Unlike retrospective achievement tests, these tests are designed to be predictive, and typically contain content that ranges widely, well beyond the bounds of any school's curriculum. Moreover, they are typically administered to self-selected samples of people seeking admission to a program, job, or occupation.[7]

The MBAE report continues:

> "The percentage of students performing at the Proficient level or higher on the MCAS English Language Arts and Mathematics tests is much higher than the percentage of students meeting the college readiness benchmarks on other tests such as the SAT or NAEP."

Yet, Massachusetts' "proficient bar" is one of the highest among the fifty states. The aforementioned U.S. Education Department mapping study found only three states with their 8th-grade math test proficiency levels set at a higher level of difficulty than the NAEP's in 2013 (Massachusetts' was fourth highest in the nation, just under the NAEP's level). Only one state's (New York's) 8th-grade reading proficiency level exceeded the NAEP's (Massachusetts was in the middle of the pack). In 4th-grade math and reading, three and two states, respectively, adopted proficiency levels exceeding the NAEP's. Massachusetts ranked second and third in the nation, just below and just above the NAEP levels in reading and math, respectively.[8]

One might argue, as some have, that the NAEP proficiency levels, set during the George H.W. Bush administration, are unrealistically high; indeed intentionally so.[9] Others have described the NAEP levels as deliberately "aspirational".[10] The NAEP levels can be set high because NAEP scores have no consequences for students. Administered to a matrix sample of classrooms in each state, most students do not take the NAEP, and those who do only complete a small section of it. Student-level NAEP scores do not exist.

The MBAE report continues in its MCAS criticism:

> "More than one-third of Massachusetts high school graduates who enroll at one of the state's public colleges or universities place into one or more noncredit-bearing, remedial courses."

Meanwhile, on the other side of the table, where only laudatory praise resides for PARCC:

> "Students receiving the PARCC college- and career-ready determination may be exempt from having to take and pass placement tests in two- and four-year public institutions of higher education."

That is, the powers-that-be behind the Common Core and PARCC have either browbeaten or hoodwinked some higher-education institutions into guaranteeing entry into credit-bearing courses with above-proficient scores on PARCC exams. PARCC above-proficient test scores will mean a student is college-ready simply because they are defined in advance to be so. And, given that those students will be exempted from taking placement tests, the empirical evidence that would verify that those students are, indeed, college-ready will be unavailable.

PARCC-favorable federal legislation—backed by gargantuan quantities of revenue-sharing dollars—defines in advance—by *fiat*—that PARCC test scores are superior to the scores of any other test, even if the other test's scores were calculated by the Platonic assembly of all of the most highly-regarded psychometricians on earth, or by God.

If PARCC supporters were genuinely confident that their test was a good indicator of college readiness, they should have been willing to let it prove itself through an accumulation of evidence over time, instead of hiding from it through a sneaky legislative *fiat*.

PARCC praises continue in the MBAE report table:

> "PARCC intends to establish a college- and career ready bar that ensures that students who meet it 'are academically prepared to engage successfully in entry-level, credit-bearing courses' in English and mathematics in college."
>
> "PARCC plans to conduct studies with colleges to ensure that students who are designated as college- and career-ready have a high probability of passing entry-level, credit-bearing English and mathematics courses."

So, they are telling us that PARCC's plans and intentions are superior to MCAS's current practical realities.

If PARCC is successful, by the time it completes its studies "to ensure that students [with above-proficient PARCC scores] have a high probability of passing entry-level, credit-bearing English and mathematics courses" they may find that they actually do. But, the "high probability" will more likely be a result of lowered standards in entry-level college courses than raised standards at the high school level. Without placement tests and remedial courses, the standards of entry-level college courses will be forced down. Entry-level college courses will acquire the same content as today's remedial courses. The content of today's entry-level credit-bearing college courses will become the content for second- or third-year college courses.

The MBAE report suggests that the MCAS high school test scale is not robust enough at the high end of the scale to validly measure college readiness.[11] But, more high-end content could easily be added to the test.

By contrast, if the PARCC is as high this, rich that, and deep something else as its proponents claim, it will not be robust enough at the low end of the scale to validly measure high school diploma achievement for the whole of the Bay State's high school students. That is, it will not meet the most basic requirements of the Massachusetts Education Reform Act (MERA), a still binding set of laws that was considered and passed by the entirety of the Great and General Court (i.e., both houses of the Massachusetts legislature). (Not that PARCC could meet the requirements of MERA anyway; with the limitation to just the two subject areas of ELA and math.)

What will Massachusetts do then, if the Bay State adopts PARCC as its high school exit examination, even though it is neither designed to be, can be, nor is advertised to be a high school exit exam, and it covers only two of the several subject areas required by law to be covered for a high school exit exam? Will Massachusetts leverage the single vote it now possesses among the seven or eight states remaining in the PARCC coalition to lobby for changes? Good luck with that.

Endnotes

1. For example, from the federal government alone, PARCC received $185,862,832 on August 13, 2013. https://www2.ed.gov/programs/racetothetop-assessment/parcc-budget-summary-tables.pdf; SBAC received $175,849,539 to cover expenses to September 30, 2014. https://www2.ed.gov/programs/racetothetop-assessment/sbac-budget-summary-tables.pdf. A complete accounting, of course, would include vast sums from the Bill and Melinda Gates Foundation, other foundations, the CCSSO, NGA, Achieve, and state governments.
2. "Constructivism is basically a theory — based on observation and scientific study — about how people learn. It says that people construct their own understanding and knowledge of the world, through experiencing things and reflecting on those experiences." Here are two descriptions of constructivism: one supportive, http://www.thirteen.org/edonline/concept2class/constructivism/ and one critical, http://epaa.asu.edu/ojs/article/view/631.
3. See, for example, Julia Steiny. (2015, October 21). "The Long Overdue Death of 19th-Century Education". *Education News*. http://www.educationnews.org/k-12-schools/julia-steiny-the-long-overdue-death-of-19th-century-education/
4. Bandeira de Melo, V., Bohrnstedt, G., Blankenship, C., and Sherman, D. (2015). *Mapping State Proficiency Standards onto the NAEP Scales: Results from the 2013 NAEP Reading and Mathematics Assessments* (NCES 2015-046). U.S. Department of Education, Washington, DC: National Center for Education Statistics.
5. Pioneer Institute White Paper No. 122 explains the differences between (retrospective) achievement tests and (predictive) aptitude/readiness tests in some detail, on pages 15-21. See R.P. Phelps & R.J. Milgram. (2014, September). *The Revenge of K-12: How Common Core and the new SAT lower college standards in the U.S.* Boston: Pioneer Institute. http://pioneerinstitute.org/featured/common-core-math-will-reduce-enrollment-in-high-level-high-school-courses/
6. C. Buckendahl & R. Hunt. (2007). Whose Rules? The Relationship Between the "Rules" and "Law" of Testing, chapter 7 in R.P. Phelps, Ed. *Defending Standardized Testing*. Mahwah, NJ: Psychology Press.
7. See Phelps & Milgram, pp. 15-21.
8. See Bandeira de Melo, et al., pp. 7-18.
9. See, for example, James Harvey. (2011, October 25). NAEP's odd definition of proficiency. *Education Week*.
10. See, for example, "Methods for NAEP standard setting" at the National Center for Educational Statistics web site, and link from there.
11. Scott Marion, et al., p. 6. http://www.mbae.org/wp-content/uploads/2015/02/MBAE-MCAS-PARCC-Report-Web.pdf

PIONEER INSTITUTE
PUBLIC POLICY RESEARCH

185 Devonshire Street, Suite 1101, Boston, MA 02110
T: 617.723.2277 | F: 617.723.1880
www.pioneerinstitute.org