Tests' Rigor Varies Plenty State to State

By Sarah D. Sparks

What students are expected to know in order to reach proficiency levels on exams in some states may be as much as four grade levels below the standards set in other states, according to a study by the American Institutes for Research that uses international testing data to gauge states against a common measuring stick.

Released last week, the report by the Washington-based research group makes a case for states, as they collaborate on common standards, to use national and international benchmarking to make cutoff scores more demanding and improve the descriptions of what it means for students to be proficient in reading and mathematics at each grade level.

The researchers compared each state’s standards against the benchmarks for the same subjects used in two international assessments, the Trends in International Mathematics and Science Study, or TIMSS, and the Progress in International Reading Literacy Study, or PIRLS, during 2007, the most recent year all three types of assessments were administered. Researchers then analyzed the percentage of students in each state who would meet minimum proficiency according to their state standards and the common international standards.

Measured against the international benchmarks, the state-to-state gaps were so great, the report notes, that the difference in proficiency between students in states with the most rigorous standards and those with the least rigorous standards was double the national achievement gap between black and white students on the National Assessment of Educational Progress in 2007, which was then about two grade levels. At the 4th grade level, only Massachusetts had more rigorous state standards than the international standards. Its standards for 4th grade math were comparable to those required for a typical student in the highest-performing TIMSS countries and jurisdictions, such as Japan, Taiwan, Singapore, and Hong Kong.

“Short Selling’ Students

Gary W. Phillips, the report’s author and the AIR’s vice president and chief scientist, called state-proficiency standards “the educational equivalent of short-selling.”
“Rather than betting on student success,” he writes, “the educators sell the student short by lowering standards.”

Michael Cohen, the president of Achieve, a Washington-based nonprofit group that works with states to evaluate their academic-content and testing standards, said the study “documents again what we’ve long known, which is on current state tests the bar for proficiency is literally all over the map.”

**Student-Expectation Gap**

A comparison of 4th grade students scoring at the proficient level in math on 2007 state assessments vs. an internationally benchmarked common standard show dramatic differences in what is considered proficient. Of all states, only Massachusetts had more students perform at the proficient level on international standards than on state standards.

The AIR researchers found the percentage of students who reached proficiency in 4th grade math and reading and 8th grade math were strongly inversely proportional to the rigor of the achievement benchmarks. The report suggests low state proficiency bars may account for up to 60 percent of the gains states have reported in student performance in the years since the No Child Left Behind Act was passed by Congress in 2001.

The AIR findings echo a report released last year, in which the National Center for Education Statistics compared states’ standards with those of the National Assessment of Educational Progress. ("NCES Finds States Lowered ‘Proficiency’ Bar," Oct. 29, 2009.) The NCES study found, for instance, that across 2003, 2005, and 2007 assessments, the distance between states with the highest and lowest proficiency bars in 4th grade reading was comparable to the difference between NAEP’s “basic” and “proficient” achievement levels.

**Benchmarking a New Way**

Mr. Phillips said the findings demonstrate a need for states to use a benchmark method to set proficiency levels.

First, the state would reach a consensus on academic-content standards and field-test a representative pool of test questions based on them. It would compile the questions in order from easy to hard, and link the scaled items statistically to equivalent questions in other states and countries. Then content experts would use both the questions and performance descriptions from other states and tests to describe what students should know and be able to
do at each proficiency level. Finally, those descriptions would be used to set cutoff scores for the state content assessments.

Three states—Delaware, Hawaii, and Oregon—have already taken the first step.

In this year’s spring high school math assessments, Oregon embedded sample questions from PISA, which tests the math performance of 15-year-olds in countries in the Organization for Economic Cooperation and Development. While the sample questions did not count for students’ scores, they were used to benchmark the state test against international standards.

From there, with input from educators and researchers, the Oregon education department has recommended changing the proficiency descriptions and cutoff scores for each grade’s assessments, according to Anthony Alpert, the assessment director for the department.

Along with adoption of common core content standards on Thursday, the state board of education approved new proficiency standards for math; the new cutoff scores will increase by half a standard deviation at each grade level, making it more rigorous, Mr. Alpert said. Other subjects are in the works.

Mr. Alpert said he hopes to create “a [testing] system that is better—more consistent with the expectations that other states have for their kids and other countries have for their kids.”