

EDUCATION WEEK

Published Online: May 10, 2011

Published in Print: May 11, 2011, as **Testing the Test**

Testing the Test

By **Marcia Kastner**

Even well into this era of standardized testing, it is clear that many important questions remain. Most importantly: How good are the tests themselves? Do they actually measure what they were designed to measure, or do they give misleading information about student achievement?

These questions are urgent and timely.

In efforts to improve the **No Child Left Behind Act** and implement the federal Race to the Top initiative, education policymakers are placing ever-greater emphasis on using test scores for teacher, school, and student accountability. However, it is inappropriate, counterproductive, and unfair to penalize teachers and students for low test performance (or to praise them for high test performance) when the tests themselves are flawed. We need to make sure tests are valid, that is, they measure what they were designed to measure. Only then can test scores accurately reflect what students know—and don't know—about the material being tested.

How do we ensure the validity of standardized tests? The answer is to have more oversight of test developers and more careful scrutiny of the tests. State governments and the federal government need to make sure that the individuals overseeing the development of state and national standardized tests are trained to recognize and fix flawed test questions. This is crucial for developing valid tests.

My own experience in this regard may be instructive. I was responsible for holding standardized tests accountable as the mathematics-assessment lead at the **Massachusetts Department of Elementary and Secondary Education** from 2003 to 2005. In that position, I oversaw the development of the state's standardized math tests required under NCLB. I reviewed thousands of math questions, analyzed their field-test data, and read samples of student work. This experience taught me how to recognize when a math question is flawed.

The reality is that most teachers, states, and national test developers believe that their math tests are valid, even though they often are not. What is troubling is that after a quick review of recent state and national standardized tests released to the Web, I easily discovered numerous examples of flawed math questions. This serious problem needs action, but is too often ignored.

"Most teachers, states, and national test developers believe that their math tests are valid,

[← Back to Story](#)



What are some of the types of flaws I have seen? The examples include assessments that: allow students to get the right answer for the wrong reason; present options in a multiple-choice question that give away the answer; allow use of calculators in cases where the calculators rather than the students can solve the problem; and contain language that is confusing and imprecise (which particularly disadvantages special education students and English-language learners).

even though they often are not."

Why are so many standardized math tests not valid? The answer may lie with how states develop their tests. For example, in Massachusetts, the state education department contracts with a testing company to write the initial versions of the test questions and then submit them to the department for review. When I worked at the department, I rejected questions that I determined were not valid, and I instructed those I supervised to do the same. Since rejected questions do not count toward the number of questions the testing company is contracted to produce, the company has an incentive to produce valid questions. The questions that state education officials do not reject are then reviewed by experienced teachers, content experts, and a bias-and-sensitivity committee. After this thorough review, state officials decide on the final set of questions to field-test. The questions accepted by the education department after field-testing go into a bank of questions from which future test questions are drawn. As is clear from this process, the state of Massachusetts is ultimately responsible for the quality of its tests. The education department decides which questions to field-test and which questions to put into the bank.

But what happens if a state does not have a strong review process and a well-trained education department staff to recognize flawed questions and reject them? What happens is that flawed questions may end up on that state's tests.

Given education reform efforts to change how states implement standardized testing, the need for a strong review process is even more urgent. Under NCLB, each state is responsible for writing its own learning standards and its own standardized tests. However, many states, including the recipients of Race to the Top funding, have adopted the new national **common-core state standards**. These states eventually will use new common assessments that are aligned with the common standards. Since these assessments are currently under development, now is the time to make sure they are valid.

Information provided by valid tests not only ensures proper accountability, but also helps educators target instruction to meet students' needs, thereby improving the quality of education.

In short, we need to hold our tests accountable. Nothing less than the future of education is at stake.

Marcia Kastner was the administrator for mathematics test development at the Massachusetts Department of Elementary and Secondary Education from 2003 to 2005 and a part-time contractor overseeing the state's high school graduation tests from 2005 to 2008. She is also the author of Testing The Test: How to Recognize When Math Tests Are Flawed, How to Fix Them, Why We Should Care (Marcia Kastner, 2010). Examples of flawed math questions from state and national standardized tests can be found on her website, www.marciakastner.com.

Vol. 30, Issue 30, Page 40