

Should standardized tests be used to assess the progress of NCLB?

Yes Throughout history, people have found only five ways to choose leaders: by inheritance, force, chance, political loyalty or achievement. No modern democracy can afford to use any method other than the last. Standardized tests help to measure achievement.

First, we should understand what is meant by the word standardized. Here's an example: If one potential leader is asked to swim across Old Hickory Lake on July 1st and another on January 1st, the test would not be standardized because the lake would be of different temperatures. We might all agree that such a test of leadership would be unfair. Standardized testing simply means that the circumstances in which tests are conducted are made as similar as possible so that our evaluation of a person's achievement will not be unduly biased. Since there is no such thing as a test with zero bias, the question is whether the test is unduly biased.

Further, since there is no such thing as a test which is free of development cost, we must also ask whether we have developed a test which is not unduly biased within the budget allocated. No test developer, policy analyst, or academic should be taken

seriously if they ignore the cost of test development in their recommendations.

There are some who argue that No Child Left Behind (NCLB) is failing because of its use of standardized tests. They say that because stakes are high, teachers help students cheat. They say that tests emphasize only a few of

They are defective. But they are not defective because they are standardized; they are defective because they are of poor quality.

the purposes of schooling. A few teachers do cheat. Bank tellers sometimes steal money. Does that justify banning the management of money by banks? And true, tests do emphasize math and science, but is that any reason not to also have tests which emphasize civics?

There are many problems with NCLB. And there are many problems with the tests used to evaluate NCLB. But the fact that the tests are standardized is not one of them. Many of the tests used to evaluate NCLB are poorly developed. The legislative authorizations for the tests are divorced from the realities of test development costs. Tests which assess the skills of evaluation and synthesis and the wide variety of subject matter in a state's curriculum are much more

expensive to design. Few legislatures know this.

Some of the NCLB tests are the equivalent of demanding that Lindberg fly across the Atlantic with only one wing. They are defective. But they are not defective because they are standardized; they are defective because they are of poor quality.

If someone were to want a non-standardized test to evaluate a nation's youth, let me enter a child of theirs in a school system where such methods have been used. The People's Republic of China during the time of the Cultural Revolution would be one choice. Sitting for an oral exam in contemporary Kazakhstan, where payment by results has a new meaning, might be another example. There are many instances of non-standardized tests around the world; each is riddled with bias. I doubt if any critic would want to enter one of their children in a school system where the tests are not standardized and where judgment is more unfair than they can possibly imagine.



BY Stephen Heyneman,
PROFESSOR OF INTERNATIONAL
EDUCATIONAL POLICY

BY Rich Lehrer,
PROFESSOR OF SCIENCE
EDUCATION

No No Child Left Behind seeks to establish an accounting system for student learning in which standardized tests serve as a metric of progress in domains of reading, mathematics and science. Although standardized tests have well known limitations, the argument is often advanced that they establish thresholds of accountability that any decent educational system should meet. In fact, they are often viewed as tools for equity, and who could possibly be opposed to equity and to accountability?

Despite appearances, this role for standardized tests rests on unexamined assumptions about the goals and roles of assessment. Although we can agree that educators should be accountable for students' education, just as structural engineers are accountable for the soundness of public works projects, it is less obvious what the nature of that accounting should be. Even a comparatively narrow focus on achievement has several implications.

First, assessment should advance learning. Hence, students and teachers should be participating in an assessment system that provides evidence about how a student is reasoning. Knowledge of student reasoning allows

teachers to change the nature of instruction. Assessment assists professional judgment.

Second, assessment should not be a single-shot moment in time (generally, at the end of the school year for NCLB) but rather more continuously embedded within instruction, again with the goal of generating evidence about the nature of student reasoning in a way that can be deployed to improve instruction. Single-shot assessments are rarely of much help in the re-design of instruction.

Third, and aligned with the first two implications, we must ask whose standards

These tests typically employ a model where the quality of learning is indicated by locating an individual along a single dimension of knowledge.

are represented by currently available standardized tests. In mathematics, most standardized tests emphasize procedural competence, but the National Research Council (NRC) suggests that procedural competence is but one of five interacting strands that contribute to mathematical proficiency. This emphasis on procedural competence, uncoupled from other forms of mathematical competence, creates a narrow spectrum for what is considered mathematics. The result is often deleterious, as systems of instruction pursue procedural

knowledge at the expense of mathematical knowledge. One might respond that state standards documents are prospective remedies for narrow spectrum tests, but a recent NRC report, Systems for State Science Assessment, suggest that state standards documents rarely provide much guidance about the nature of the learning performances one wishes to assess.

Finally, one must consider the technical models that guide the construction of most standardized assessments used for purposes of NCLB. These tests typically employ a model where the quality of learning is indicated by locating an individual along a single dimension of knowledge. An analogy to a ruler would be apt. However,

when test items don't fit this model, they are eliminated. Hence, unlike other sciences, where the adequacy of models are tested in light of natural phenomena, most standardized tests assume that the model of learning is correct and simply cast out the phenomena that don't fit. Policies are then erected on the basis of this fabrication. An alternative would be to take the time and expense to create standards and measures anchored firmly to learning, so that testing and learning might again mutually profit from the association.

We invite readers' ideas for future "Versus" topics. If you have ideas or wish to submit commentary, please send it to the Editor, Peabody Reflector, VU Station B #357703, 2301 Vanderbilt Place, Nashville, TN 37235-7703, or email reflector@vanderbilt.edu.