

EDUCATION WEEK

Published Online: March 22, 2016

Published in Print: March 23, 2016, as **Purposeful Assessment Is an Antidote to Harmful School Testing**

COMMENTARY

The Fatal Flaw of Educational Assessment

By **W. James Popham**

America's students are not being educated as well these days as they should be. A key reason for this calamity is that we currently use the wrong tests to make our most important educational decisions. The effectiveness of both teachers and schools is now evaluated largely using students' scores on annually administered standardized tests, but most of these tests are simply unsuitable for this intended purpose.

When we use the wrong tests to evaluate instructional quality, many strong teachers are regarded as ineffective and directed by administrators to abandon teaching procedures that actually work well. Conversely, the wrong test scores often fail to identify truly weak teachers—those in serious need of instructional assistance who don't receive help because they are thought to be teaching satisfactorily. In both these instances, it is the students who are shortchanged.

What's most dismaying about this widespread misuse of educational tests is that many educators, most policymakers, and almost all parents of school-age children do not realize how these tests contribute to diminished educational quality.

Today's educational tests are intended to satisfy three primary purposes, all of which can play a constructive role in students' education: *to compare*, *to instruct*, and *to evaluate*.

Comparison-focused educational tests permit us to identify score-based differences among individual students or among groups of students. The resulting comparisons often lead to classifications of students' scores on a student-by-student basis (such as by using percentiles) or on a group-by-group basis (such as by distinguishing between "proficient" and "nonproficient" students).

A second purpose of educational testing is instructional—that is, to elicit ongoing evidence regarding

[← Back to Story](#)



**Blended
by Design**

Diagnostic and monitoring,
whole class and small group instruction,
and personalized learning.

i-Ready[™]

Learn more

students' levels of achievement so that better decisions can be made about how to teach those students. Test-based evidence can also help students themselves decide whether to modify how they are trying to learn.

A third purpose of educational testing is evaluation—that is, determining the quality of a completed set of instructional activities provided by one or more teachers. These evaluations often focus on a lengthy segment of instruction, such as an entire school year.

All three of these purposes, if implemented by using appropriate tests, can benefit students. The trouble is that one of those purposes—comparison—has completely dominated America's educational testing for almost a century.

Our preoccupation with comparative testing can be traced back to World War I when, in order to identify the best candidates for officer-training programs, a group-administered intelligence test called the Army Alpha was developed for more than **1.5 million U.S. Army recruits**. The test, whose comparative purpose was to spot the strongest officer candidates, worked well. As a consequence, for nearly 100 years, almost all our nation's educational tests have been built and evaluated on the basis of a test's comparative capabilities.

However, tests built chiefly for comparisons are not suitable for purposes of instruction or evaluation of instructional quality in education. These tests provide teachers with few instructional insights and typically lead to inaccurate evaluations of a teacher's instructional quality.

In 2014, the three national associations most concerned with U.S. educational testing—the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education—published a long-awaited **new edition of guidelines** for building and evaluating educational tests. The revised standards emphatically call for construction and evaluation of educational tests according to the specific purpose for which a test will be used. In a very direct manner, these revised standards advocate intentional educational testing, in which purpose-specific tactics dominate test development and purpose-specific evidence governs test evaluation.

The time has come for us to abandon the naive belief that an educational test created for Purpose X can be cavalierly used for Purpose Z. Too many children in our schools are harmed by these methods because educators are basing their decisions on inaccurate information supplied by the wrong tests. We must follow the up-to-date advice of the measurement community and demand the use of purposeful educational testing.



—Getty

"Tests built chiefly for comparisons are not suitable for purposes of instruction or evaluation of instructional quality."

MORE OPINION



[Visit Opinion.](#)

W. James Popham is a professor emeritus at the graduate school of education and information studies at the University of California, Los Angeles. He is a past president of the American Educational Research Association and a recipient of the 2015 Robert L. Linn Distinguished Address Award. This Commentary is based on the address he will deliver at the AERA annual meeting next month.

Follow the Education Week Commentary section on [Facebook](#) and [Twitter](#).

Vol. 35, Issue 25, Page 22