STANDARDIZED ACHIEVEMENT TESTS: MISNAMED AND MISLEADING

W. James Popham

During the last decade, the educators who operate America's public schools have been increasingly preoccupied with students' scores on standardized achievement tests. That's because schools whose students score high on such tests are thought to be effective, while schools whose students score low are thought to be ineffective. Both thoughts about school effectiveness, however, are often mistaken. The wrong tests are being used.

Concerns about students' test scores will escalate dramatically, of course, if Congress requires states to give standardized achievement tests each year to all students in grades 3-8.

To accurately evaluate a school staff's instructional success, it is almost self-evident that we should determine how much students have learned in that school. Most Americans believe this is what's being measured by standardized achievement tests such as the Iowa Tests of Basic Skills or the Stanford Achievement Tests. That belief, however, is mistaken.

Many of the misperceptions that Americans have about traditional standardized achievement tests stems from the misleading label pinned on those tests. *Achievement* conveys the idea that these tests, as Webster's Dictionary puts it, measure "knowledge or proficiency in something that can be learned or taught." In other words, an *achievement* test would seem to be measuring what students have *learned* in school.

But this is not the measurement function of traditional standardized achievement tests. Ever since such tests arrived on the scene in the early 1920s, their overriding function has been to permit comparisons among test-takers. Indeed, today's standardized achievement tests are patterned directly after the Army Alpha, a group-administered intelligence test used in World War I to identify potential officers.

To provide accurate comparisons among test-takers, the Army Alpha and its descendants must make sure that examinees' scores are widely spread out. If there's plenty of score-spread, then a test can determine (in relationship to a norm group of previous test-takers), for example, that Chris scored at the 87th percentile and Lee scored at the 83rd percentile. Score-spread is imperative if these tests are going to do their job, namely, identifying an examinee's *relative* performance.

But to make certain that standardized achievement tests provide accurate comparisons, the developers of these tests include many items that have nothing to do with what's supposed to be taught in school. Remember, that's not the measurement mission of these tests.

---

* A version of this essay appeared on page 46 of the September 19, 2001 *Education Week*, Vol. 21, No. 3.

So, if you were to review the actual items in a typical standardized achievement test, you'd find many items whose correct answer depends heavily on the socioeconomic status of a child's family. There are also many items that measure the verbal, quantitative, or spatial aptitudes that children *inherit at birth*. Such items are better suited for intelligence tests. Clearly, items dependent either on the affluence of a student's family or on a child's genetic inheritance are not suitable for evaluating schools.

In short, "achievement" tests really aren't. And, because many of their items measure what students bring to school, not what they learn there, traditional standardized achievement tests should have no role in evaluating our schools.

Is it possible to build standardized tests that accurately measure what's been taught in school? Absolutely! But those tests must be built with that specific role in mind. We need to evaluate a school based on how much students have learned in that school. But we'll never do so if, because of misunderstandings about the role of traditional standardized achievement tests, we continue to use the wrong tests when judging our schools.