

STANDARDIZED TESTS

(Author's Introduction: In early 2006 I was invited to contribute a 750-word description of "standardized tests" for the upcoming publication of the Routledge International Encyclopedia of Education. My essay was to be a general introduction to the topic that would be suitable for readers in nations other than the United States. Well, three weeks before my September 1, 2006 deadline, I dutifully submitted the brief essay to the editor and publisher. Immediately thereafter, however, I received an e-mail notice from the publisher indicating that, based on a recent market appraisal, Routledge had decided not to continue with the project. I will never be certain that the publisher's receipt of the following essay was not the final straw that fractured the camel's back. But it's possible.)

Standardized tests are assessment instruments that are administered, scored, and interpreted in a standard, predetermined manner. Although standardized tests in the field of education are employed worldwide, their use has been particularly prevalent in the USA. Indeed, during the final years of the twentieth century and the early part of the twenty-first century, almost all educational accountability programs in the USA were based dominantly on students' standardized test performances. Although it is possible for standardized tests to be employed for small-scale assessment tasks, such as those involved in the routine classroom testing of students, almost all standardized tests are used in settings where large-scale assessments are required. A typical use of a standardized test, therefore, is to monitor the quality of tax-supported schools by annually requiring all students in those schools to complete one or more standardized tests.

Standardized tests in education are customarily classified into one of two categories, namely, aptitude tests or achievement tests. Educational aptitude tests are intended to predict how well students will perform in a subsequent academic setting. For example, students nearing the late stages of their pre-collegiate schooling might be required to take a college entrance examination whose sole purpose is to predict the grades likely to be earned by those test-takers after arriving at college. Although, for well over a half-century, such assessments have been described as aptitude tests, during the 1990s critics objected to "aptitude" as a modifier because this descriptor tended to connote the measurement of a set of innate, unalterable capacities in students. However, even though it might be more accurate, and more politically correct, to refer to such assessments as "predictor tests," the label "aptitude tests" is still widely used.

In contrast, educational achievement tests are intended to assess the degree to which students possess certain skills and knowledge. Achievement tests attempt to supply answers to questions about what it is that students know and can do. In most situations where a standardized test is used as the cornerstone of an

educational accountability strategy, such tests are, understandably, achievement tests rather than aptitude tests.

Standardized achievement testing in the USA commenced shortly after World War I. It was during this conflict that the *Army Alpha* was employed by military officials to predict which army recruits were likely to succeed in officer training programs. Scores of test-takers were compared with scores of a set of previous test-takers (known as the “norm group”) so that Army recruits with the highest scores could be sent to officer training programs. The *Army Alpha*, administered to 1,750,000 men during World War I, embodied the basic assessment strategy adopted by subsequent developers not only of standardized aptitude tests such as the *Alpha*, but also by developers of standardized achievement tests. Because today’s standardized tests yield scores whose interpretations must be referenced to the performance of the examinees constituting a test’s norm group, these tests are often described as “norm-referenced” tests.

During the 1980s and 1990s, a new genre of standardized achievement tests became increasingly popular, not only in the USA, but also in many other nations. These newer achievement tests were designed to measure the degree to which students’ had mastered official collections of curricular aims known as “content standards.” Such assessments, typically referred to as “standards-based” tests, are aimed less at providing comparative interpretations and more at determining the extent to which students have mastered particular content standards. Because students’ score-interpretations are typically referenced to a set of targeted criterion outcomes, that is, the content standards being measured, standards-based examinations are usually characterized as “criterion-referenced” tests.

Because of the significant role standardized achievement tests play in today’s educational accountability programs, increasing concerns have been recently registered regarding the degree to which some standardized tests are sensitive to instruction, that is, are capable of accurately discerning differences in instructional quality. Because of the way in which standardized tests are typically constructed, many of today’s standardized achievement tests are so closely linked to students’ socioeconomic status that these tests tend to measure what students bring to school rather than what they’ve learned there. Clearly, an instructionally insensitive standardized test is altogether inappropriate for evaluating the quality of schooling.

Standardized tests, whether focused on aptitude or achievement, have been with us for many years. Their continued existence, however, does not necessarily signify that such tests are appropriate for the measurement missions they are currently being asked to accomplish.

W. JAMES POPHAM
August 2006