

FairTest

National Center for Fair & Open Testing

The Limits of Standardized Tests for Diagnosing and Assisting Student Learning

Standardized tests have historically been used as measures of how students compare with each other (norm-referenced) or how much of a particular curriculum they have learned (criterion-referenced). Increasingly, standardized tests are being used to make major decisions about students, such as grade promotion or high school graduation, and schools. More and more often, they also are intended to shape curriculum and instruction.

Proponents of the expanded uses and consequences of tests claim that newer exams are superior to the flawed exams of the past, measure what is important, and are worth teaching to. These arguments ignore the real-world limits to what standardized tests can usefully do. Repeating such false claims perpetuates test misuse and the dangerous belief that what is worth teaching is that which can be assessed by a standardized test.

Under a new federal law, state assessments of reading and math must be administered for accountability annually in grades 3-8 and once in high schools. The assessments must be based on state content and performance standards; measure higher order thinking; provide useful diagnostic information; and be valid and reliable. While the law does not mandate the use of standardized tests, many states will be inclined administer them to meet the federal law. An examination of each requirement, however, reveals the limits of standardized tests.

Tests are to be based on state standards

State standards are often too long and detailed to ever be taught. Many fail to distinguish what is important from what is unimportant or to separate what all students ought to learn in a subject from what only the most interested might learn. In part because of the level of detail, much of the content in state standards is not assessed by state tests.

Moreover, much of value in state standards cannot be tested with any paper-and-pencil test of a few hours duration. In a high quality education, students conduct science experiments, solve real-world math problems, write research papers, read novels and stories and analyze them, make oral presentations, evaluate and synthesize information from a variety of fields, and apply their learning to new and ill-defined situations. Standardized tests are poor tools for evaluating these important kinds of learning. If instruction focuses on the test, students will not learn these skills, which are needed for success in college and often in life.

Measure higher-order thinking

Standardized exams offer few opportunities to display the attributes of higher-order thinking, such as analysis, synthesis, evaluation, and creativity. Higher order thinking is encouraged and revealed by in-depth and extended work, not by one-shot tests.

Provide useful diagnostic information

Assessments of educational strengths and weaknesses can be useful at the individual, classroom, school or district levels. However, information needs to be sufficiently timely, accurate, meaningful, detailed and comprehensive for the kind of diagnosis being made. The lengthy turn-around time for scoring most standardized tests makes them nearly useless for helping a particular individual, though the information might be of some value to teachers and schools for longer-range planning.

In addition, standardized tests usually include only a few questions on any particular topic. This is too little information to produce accurate, comprehensive or detailed results. Many topics in state standards are not addressed at all in state exams, so the tests provide no diagnostic information about them.

Diagnosis suggests the use of “formative” assessment – assessments that can help a teacher and student know what to do next. Standardized tests administered at the end of the year – “summative assessment” – cannot possibly meet this need. Sound diagnostic practices also include understanding why a student is having difficulty or success and determining appropriate action. As snapshots with limited information, standardized tests provide neither an answer to “why” nor little guidance for successful instruction.

Be valid and reliable

Test validity, experts explain, resides in the inferences drawn from assessment results and the consequences of their uses. Relying solely on scores from one test to determine success or progress in broad areas such as reading or math is likely to lead to incorrect inferences and then to actions that are ineffective or even harmful. For these and other reasons, the standards of the testing profession call for using multiple measures for informing major decisions – as does the ESEA legislation.

Reliability, or consistency of information, is sometimes treated as the most important aspect of testing. However, consistent information about too narrow a range of topics, skills or knowledge cannot provide adequate information for credible decisions: a doctor needs more than just reliable blood pressure results to treat a patient. Well-designed classroom-based assessments can provide richer, consistent information that enhances validity, diagnostic capacity, and the ability to assess progress toward meaningful standards.

Conclusion

When standardized tests are the primary factor in accountability, the temptation is to use the tests to define curriculum and focus instruction. What is not tested is not taught, and what is taught does not include higher-order learning. How the subject is tested becomes a model for how to teach the subject. At the extreme, school becomes a test prep program – and this extreme already exists.

It is of course possible to use a standardized test and not let its limits control curriculum and instruction. However, this can result in a school putting itself at risk for producing lower test scores. It also means parents and the community are not informed systematically about the non-tested areas, unless the school or district makes a great effort.

To improve learning and provide meaningful accountability, schools and districts cannot rely solely on standardized tests. The inherent limits of the instruments allow them only to generate information that is inadequate in both breadth and depth. Thus, states, districts and schools must find ways to strengthen classroom assessments and to use the information that comes from these richer measures to inform the public.