

FairTest

National Center for Fair & Open Testing

Criterion- and Standards-Referenced Tests

Criterion-referenced tests (CRTs) are intended to measure how well a person has learned a specific body of knowledge and skills. Multiple-choice tests most people take to get a driver's license and on-the-road driving tests are both examples of criterion-referenced tests. As on most other CRTs, it is possible for everyone to earn a passing score if they know about driving rules and if they drive reasonably well.

In contrast, norm-referenced tests (NRTs) are made to compare test takers to each other. On an NRT driving test, test-takers would be compared as to who knew most or least about driving rules or who drove better or worse. Scores would be reported as a percentage rank with half scoring above and half below the mid-point (see NRT fact sheet).

In education, CRTs usually are made to determine whether a student has learned the material taught in a specific grade or course. An algebra CRT would include questions based on what was supposed to be taught in algebra classes. It would not include geometry questions or more advanced algebra than was in the curriculum. Most all students who took algebra could pass this test if they were taught well and they studied enough and the test was well-made.

On a standardized CRT (one taken by students in many schools), the passing or "cut-off" score is usually set by a committee of experts, while in a classroom the teacher sets the passing score. In both cases, deciding the passing score is subjective, not objective. Sometimes cut scores have been set in a way that maximizes the number of low income or minority students who fail the test. A small change in the cut score would not change the test's meaning but would greatly increase minority pass rates.

Some CRT's, such as many state tests, are not based on a specific curriculum, but on a more general idea of what students might be taught. Therefore, they may not match the curriculum. For example, a state grade 10 math test might include areas of math which some students have not studied.

Standards-Referenced Tests

A recent variation of criterion-referenced testing is "standards-referenced testing" or "standards based assessment." Many states and districts have adopted *content standards* (or "curriculum frameworks") which describe what students should know and be able to do in different subjects at various grade levels. They also have *performance standards* that define how much of the content standards students should know to reach the "basic" or "proficient" or "advanced" level in the subject area. Tests are then based on the standards and the results are reported in terms of these "levels," which, of course, represent human judgement. In some states, performance standards have been steadily increased, so that students continually have to know more to meet the same level.

Educators often disagree about the quality of a given set of standards. Standards are supposed to cover the important knowledge and skills students should learn -- they define the "big picture." State standards should be well-written and reasonable. Some state standards have been criticized for including too much, for being too vague, for being ridiculously difficult, for undermining higher quality local curriculum and instruction, and for taking sides in educational and political controversies. If the standards are flawed or limited, tests based on them also will be. In any event, standards enforced by state tests will have -- and are meant to have -- a strong impact on local curriculum and instruction.

Even if standards are of high quality, it is important to know how well a particular test actually matches the standards. In particular, are all the important parts of the standards measured by the test? Often, many important topics or skills are not assessed.

A major reason for this is that most state exams still rely almost entirely on multiple-choice and short-answer questions. Such tests cannot measure many important kinds of learning, such as the ability to conduct and report on a science experiment, to analyze and interpret information to present a reasonable explanation of the causes of the Civil War, to do an art project or a research paper, or to engage in serious discussion or make a public presentation (see fact sheet on multiple-choice tests). A few standards-based exams have gone beyond multiple-choice and short-answer, but even then they may not be balanced or complete measures of the standards.

CRTs and NRTs

Sometimes one kind of test is used for two purposes at the same time. *In addition* to ranking test takers in relation to a national sample of students, a NRT might be used to decide if students have learned the content they were taught. A CRT might be used to assess mastery *and* to rank students or schools based on their scores. In many states, students have to pass *either* an NRT or a CRT to obtain a diploma or be promoted. This is a serious misuse of tests. Because schools serving wealthier students usually score higher than other schools, ranking often just compares schools based on community wealth. This practice offers no real help for schools to improve.

NRTs are designed to sort and rank students "on the curve," not to see if they met a standard or criterion. Therefore, NRTs should not be used to assess whether students have met standards. However, in some states or districts a NRT is used to measure student learning in relation to standards. Specific cut-off scores on the NRT are then chosen (usually by a committee) to separate levels of achievement on the standards. In some cases, a CRT is made using technical procedures developed for NRTs, causing the CRT to sort students in ways that are inappropriate for standards-based decisions.

Sometimes the NRT is changed to more closely fit the state standards and to report standards-referenced scores. As a result, a state could report that 35 percent of its students were proficient according to state standards (depending, of course, on where the cut-off score is set), but that 60 percent of its students were above the national average score on the norm-referenced test. Adapting an NRT also means that while everything on the test is in the standards, much of what is in the standards is not in the tests.

Conclusion

If standardized tests are used at all, CRTs make more sense for schools than do NRTs. However, they should be based on relevant, high-quality standards and curriculum and should make the least possible use of multiple-choice and short-answer questions. As with all tests, CRTs and NRTs, no matter what they are called, should not control curriculum and instruction, and important decisions about students, teachers or schools should not be based solely or automatically on test scores.

For a sample newsletter, a publications catalog, or a list of other fact sheets, please write to us at 342 Broadway, Cambridge, MA 02139, or email to FairTest@aol.com. You also can find fact sheets and other materials on our website at <http://www.fairtest.org>. We welcome your feedback.