

A LINE ON LIFE

9/7/83, Updated 6/21/01

Qualities of a Good Test

David A. Gershaw, Ph.D.

For many people, there is no such thing as a "good" test. However, some tests are better than others. What qualities make these tests better?

Tests are better, if they are relatively **objective**. A test is objective, if — using the same scoring key — whoever scores the test will arrive at the same score — assuming no clerical errors. Objective test items are usually multiple choice, matching or true-false. In contrast, essay questions are typically **subjective**. This means that different people — or the same person in a different mood — will tend to score the same essay answers differently. However, with more exact standards of scoring, essay questions can be relatively objective. Scorer bias will be reduced, and essentially the test will be objective — there will be *consistency among scorers*.

A good test should also be relatively **reliable**. As long as the quality being measured has not changed, this means that any person should get about the same score each time they take the test. However, to be reliable, the test must be relatively objective. How can you obtain *consistency among the scores* you earn from one time to the next, if the scorers are inconsistent?

A third quality a good test should have is **validity**. To be valid, a test should *measure what it claims to measure*. Although it needs to be relatively reliable to be valid, merely because it is reliable does not mean that it will be valid.

Suppose I were to give a man an intelligence test by measuring his height. I use a tape measure three different times, and each time, I get a measure of 5'5". His scores are completely consistent. Is my test valid? Probably not. I cannot really measure intelligence with a tape measure. Even though my test is perfectly reliable, it is not necessarily valid.

On the other hand, how can we measure what we claim to measure (validity), if the measurements are not consistent (reliability)? Thus relative reliability is needed for a test to be valid.

**In contrast to absolute measures,
tests only give a relative ranking
in terms of group norms.**

Finally, any good test must have **standardization**. This means that the *same procedures and conditions* are used each time the test is given. Such things as instructions, time limits, lighting and so on are the same for each administration. If this is the case, all those who take the test can be used as part of the standardization **norms**. With any measurement, you can only rate a person as high, low or average in relations to a set of norms. The problem is "Which norm?" If you want to judge yourself in terms of height, you wouldn't want to use basketball players as your norm group.

The question, "Which norm?" causes a big problem with intelligence testing. The most frequently used intelligence tests take "middle-class WASPs" (White, Anglo-Saxon, Protestants) as their norm, assuming that everyone has similar background and learning experiences in our society. However, this does not accurately apply to many minority members, such as African-Americans, Hispanics or American Indians. This is especially true, if they are from different backgrounds, like the ghetto, barrio or reservation, respectively. The assumption of similar backgrounds does not apply in these cases. Because of this, when

members of these groups are compared to general norms, they may be falsely labeled as slow learners or even mentally retarded. However, when compared with norms of others from a similar background, many of these people may earn scores that indicate high potential.

Thus, if you come from a background different from "*middle-class WASP*" and take a "*standardized*" test, before judging yourself from the results, find out what group is being used as a norm.
