

TESTING VALIDITY OF RESEARCH

Validity is the most important issue in selecting a test. Validity refers to what characteristic the test measures and how well the test measures that characteristic.

1. Validity tells you if the characteristic being measured by a test is related to job qualifications and requirements.
2. Validity gives meaning to the test scores. Validity evidence indicates that there is linkage between test performance and job performance. It can tell you what you may conclude or predict about someone from his or her score on the test. If a test has been demonstrated to be a valid predictor of performance on a specific job, you can conclude that persons scoring high on the test are more likely to perform well on the job than persons who score low on the test, all else being equal.
3. Validity also describes the degree to which you can make specific conclusions or predictions about people based on their test scores. In other words, it indicates the usefulness of the test.

It is important to understand the differences between reliability and validity. Validity will tell you how good a test is for a particular situation; reliability will tell you how trustworthy a score on that test will be. You cannot draw valid conclusions from a test score unless you are sure that the test is reliable. Even when a test is reliable, it may not be valid. You should be careful that any test you select is both reliable and valid for your situation.

A test's validity is established in reference to a specific purpose; the test may not be valid for different purposes. For example, the test you use to make valid predictions about someone's technical proficiency on the job may not be valid for predicting his or her leadership skills or absenteeism rate. This leads to the next principle of assessment.

Similarly, a test's validity is established in reference to specific groups. These groups are called the reference groups. The test may not be valid for different groups. For example, a test designed to predict the performance of managers in situations requiring problem solving may not allow you to make valid or meaningful predictions about the performance of clerical employees. If, for example, the kind of problem-solving ability required for the two positions is different, or the reading level of the test is not suitable for clerical applicants, the test results may be valid for managers, but not for clerical employees.

Test developers have the responsibility of describing the reference groups used to develop the test. The manual should describe the groups for whom the test is valid, and the interpretation of scores for individuals belonging to each of these groups. You must determine if the test can be used appropriately with the particular type of

people you want to test. This group of people is called your target population or target group.

Methods for conducting validation studies

The Uniform Guidelines discuss the following three methods of conducting validation studies. The Guidelines describe conditions under which each type of validation strategy is appropriate. They do not express a preference for any one strategy to demonstrate the job-relatedness of a test.

Criterion-related validation requires demonstration of a correlation or other statistical relationship between test performance and job performance. In other words, individuals who score high on the test tend to perform better on the job than those who score low on the test. If the criterion is obtained at the same time the test is given, it is called concurrent validity; if the criterion is obtained at a later time, it is called predictive validity.

Content-related validation requires a demonstration that the content of the test represents important job-related behaviors. In other words, test items should be relevant to and measure directly important requirements and qualifications for the job.

Construct-related validation requires a demonstration that the test measures the construct or characteristic it claims to measure, and that this characteristic is important to successful performance on the job.

The three methods of validity—criterion-related, content, and construct—should be used to provide validation support depending on the situation. These three general methods often overlap, and, depending on the situation, one or more may be appropriate. French (1990) offers situational examples of when each method of validity may be applied.

First, as an example of criterion-related validity, take the position of millwright. Employees' scores (predictors) on a test designed to measure mechanical skill could be correlated with their performance in servicing machines (criterion) in the mill. If the correlation is high, it can be said that the test has a high degree of validation support, and its use as a selection tool would be appropriate.

Second, the content validation method may be used when you want to determine if there is a relationship between behaviors measured by a test and behaviors involved in the job. For example, a typing test would be high validation support for a secretarial position, assuming much typing is required each day. If, however, the job required only minimal typing, then the same test would have little content validity. Content validity does not apply to tests measuring learning ability or general problem-solving skills (French, 1990).

Finally, the third method is construct validity. This method often pertains to tests that may measure abstract traits of an applicant. For example, construct validity may be used when a bank desires to test its applicants for "numerical aptitude." In this case, an aptitude is not an observable behavior, but a concept created to explain possible future behaviors. To demonstrate that the test possesses construct validation support, ". . . the bank would need to show (1) that the test did indeed measure the desired trait and (2) that this trait corresponded to success on the job" (French, 1990, p. 260).