Introduction

Validity is arguably the most important criteria for the quality of a test. The term validity refers to whether or not the test measures what it claims to measure. On a test with high validity the items will be closely linked to the test's intended focus. For many certification and licensure tests this means that the items will be highly related to a specific job or occupation. If a test has poor validity then it does not measure the job-related content and competencies it ought to. When this is the case, there is no justification for using the test results for their intended purpose. There are several ways to estimate the validity of a test including content validity, concurrent validity, and predictive validity. The face validity of a test is sometimes also mentioned.

Types of Validity

Content Validity

While there are several types of validity, the most important type for most certification and licensure programs is probably that of content validity. Content validity is a logical process where connections between the test items and the job-related tasks are established. If a thorough test development process was followed, a job analysis was properly conducted, an appropriate set of test specifications were developed, and item writing guidelines were carefully followed, then the content validity of the test is likely to be very high. Content validity is typically estimated by gathering a group of subject matter experts (SMEs) together to review the test items. Specifically, these SMEs are given the list of content areas specified in the test blueprint, along with the test items intended to be based on each content area. The SMEs are then asked to indicate whether or not they agree that each item is appropriately matched to the content area indicated. Any items that the SMEs identify as being inadequately matched to the test blueprint, or flawed in any other way, are either revised or dropped from the test.

Concurrent Validity

Another important method for investigating the validity of a test is concurrent validity. Concurrent validity is a statistical method using correlation, rather than a logical method. Examinees who are known to be either masters or non-masters on the content measured by the test are identified, and the test is administered to them under realistic exam conditions. Once the tests have been scored, the relationship is estimated between the

examinees' known status as either masters or non-masters and their classification as masters or non-masters (i.e., pass or fail) based on the test. This type of validity provides evidence that the test is classifying examinees correctly. The stronger the correlation is, the greater the concurrent validity of the test is.

Predictive Validity

Another statistical approach to validity is predictive validity. This approach is similar to concurrent validity, in that it measures the relationship between examinees' performances on the test and their actual status as masters or non-masters. However, with predictive validity, it is the relationship of test scores to an examinee's *future* performance as a master or non-master that is estimated. In other words, predictive validity considers the question, "How well does the test predict examinees' future status as masters or non-masters?" For this type of validity, the correlation that is computed is between the examinees' classifications as master or non-master based on the test and their later performance, perhaps on the job. This type of validity is especially useful for test purposes such as selection or admissions.

Face Validity

One additional type of validity that you may hear mentioned is face validity. Like content validity, face validity is determined by a review of the items and not through the use of statistical analyses. Unlike content validity, face validity is not investigated through formal procedures and is not determined by subject matter experts. Instead, anyone who looks over the test, including examinees and other stakeholders, may develop an informal opinion as to whether or not the test is measuring what it is supposed to measure. While it is clearly of some value to have the test appear to be valid, face validity alone is insufficient for establishing that the test is measuring what it claims to measure. A well developed exam program will include formal studies into other, more substantive types of validity.

Summary

The validity of a test is critical because, without sufficient validity, test scores have no meaning. The evidence you collect and document about the validity of your test is also your best legal defense should the exam program ever be challenged in a court of law. While there are several ways to estimate validity, for many certification and licensure exam programs the most important type of validity to establish is content validity.