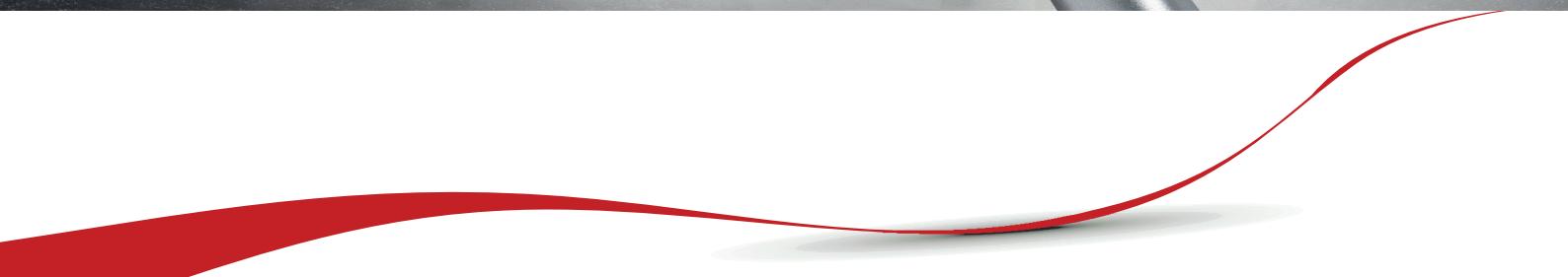




The art of assessment

Psychometric Validation Central Test





The art of Assessment

Psychometric Properties

Periodic assessment of validity and fine-tuning are crucial for long-term survival and effectiveness of any psychometric test. Central Test takes the greatest care to regularly assess the psychometric properties of all its tests. These assessments involve content review, statistical analysis and independent validity studies. Every test at Central Test is assessed against four distinct criteria before being published online.

Reliability

Reliability refers to how dependably or consistently a test measures a characteristic. If a person takes the test again, will he or she get a similar test score, or a much different score? A test that yields similar scores for a person who repeats the test is said to measure the characteristic reliably. It is, however, notable here that the measurement in behavioral sciences is always influenced by certain external variables. These could be:

- **Test taker's temporary psychological or physical state.** Test performance can be influenced by a person's psychological or physical state at the time of testing. For example, differing levels of anxiety, fatigue, or motivation may affect the applicant's test results.
- **Environmental factors.** Differences in the testing environment, such as room temperature, lighting, noise etc. can influence an individual's test performance.

These and other similar factors are sources of chance or random measurement error in the assessment process. The degree to which test scores are unaffected by measurement errors is an indication of the reliability of the test. Reliable assessment tools produce dependable, repeatable, and consistent information about people. There are several types of reliability estimates, each influenced by different sources of measurement error. For the tests developed by us, two kinds of reliabilities are particularly important. These are:

- **Internal consistency reliability** indicates the extent to which items on a test measure the same construct. A high internal consistency reliability coefficient indicates that the items on the test are very similar to each other in content (homogeneous). It is important to note that the length of a test also affects internal consistency. A very long test, therefore, can spuriously have inflated reliability coefficient. Internal consistency is commonly measured as Cronbach Alpha which is between 0 (low) and 1 (high).
- **Test-retest reliability** indicates the repeatability of test scores with the passage of time. These estimates also reflect the stability of the characteristic or construct being measured by the test. Some constructs are more stable than others. For example, an individual's reading ability is more stable over a particular period of time than that individual's anxiety level. Therefore, one would expect a higher test-retest reliability coefficient on a reading test than you would on a test that measures anxiety. For constructs that are expected to vary over time, an acceptable test-retest reliability coefficient may be lower than for constructs that are stable overtime. Test retest reliability is reported as correlation between two administrations of the test.

As quality conscious test developers we report reliability estimates that are relevant to a particular test. The acceptable level of reliability differs depending on the type of test and the reliability estimate used.

Validity

Validity refers to what characteristic the test measures and how well the test measures that characteristic. Validity estimates tell us if the characteristics being measured by a test are related to the requirements of an assessment situation. Validity gives meaning to the test scores. Validity evidence indicates that there is linkage between test performance and job performance. It can tell as to what one 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, one can conclude that people scoring high on the test are more likely to perform well on the job than people who score low on the test, other things being equal. Validity also describes the degree to which one 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. A test's validity is established in reference to a specific purpose. In all our product documents we mention the assessment context and target group the test is validated on.

There are several types of validity conceptualized by researchers and behavioral scientists. They can be grouped into three distinct categories:

- **Criterion-related validity** is assessed by examining correlation or other statistical relationships 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 validity** is explored by examining whether 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. The test content should correspond to the reading level and domain of the target population.
- **Construct-related validity** requires a demonstration that the test measures the construct or characteristic it claims to measure i.e. the test content is actually representing the underlying construct. For example, a test for occupational interest should measure occupational interests and not motivational level. Construct validity is generally examined in one of two ways:
 - By giving the test-item to a panel of experts and asking for their judgment on proximity of the test content with construct of the test. (**Face validity**)
 - By administering the test along with other established tests developed on theoretically similar constructs and examining the correlation between the two (**Convergent validity**) or by administering the test along with theoretically opposite tests and exploring the correlation (**Divergent validity**).

The three types of validity—criterion-related, content, and construct—are 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.

We conduct in-house as well as third party validity studies to examine the validity of our tests. The summary findings of these studies are given in technical documents of the tests and used for further improvement of the test.



The art of Assessment

Desirability and Faking

Social desirability bias is the inclination to present oneself in a manner that will be viewed favorably by others. Being social creatures by nature, people are generally inclined to seek some degree of social acceptance. Social desirability is one of the principle biases of a personality test; it is even more prominent in the context of recruitment. A candidate is usually tempted to answer in a socially acceptable manner in order to impress the recruiter.

We have developed a two-fold strategy to address this issue. First, all items are examined to ensure that they are not susceptible to desirability. Secondly, we include a desirability scale in tests to identify deliberate or unconscious faking depending upon the nature and applicability of the test.

Socio-legal considerations

Another important characteristic of a good psychometric test is that it should not be biased (positively or negatively) towards any socio-cultural groups, particularly those protected by law. A good psychometric test should not show discrimination on the basis of religion, gender, race or culture of the test taker.

While developing the tests all items are examined against appropriateness of content for the target population. In validation studies, the relationship between demographic variables and test scores are examined to identify any potential bias. This enables us to make sensitive but unbiased tests.