



The psychometric properties of the tests for individuals

Personality, Career and Aptitude Tests

career test

recruitment personality test

professional profile

E.Q. test

marketing aptitude test

management skills test

Introduction

Central Test International has more than ten years' experience in the publication of psychometric tests. All our tests are developed using approved scientific methods and are based on established empirical and theory-based research. Unlike other publishers, Central Test subjects all its products to regular reassessment and revision to ensure that they remain relevant to the realities of the employment and careers markets and meet the needs of both job-seekers and graduates.

Whether they're aimed at human resources professionals or the general public, each test is developed according to the same rigorous validation criteria and meets the same stringent quality standards.

We have a number of established partners in the recruitment sector, such as StepStone, Monster, and Success & Career, who are happy to allow us to relay our online tests from their websites, and to endorse our tests in their promotional material.

In this document, we hope to set out the psychometric criteria that underlie the development of a test and show the process of verification it undergoes before being launched on the market. We also hope to give a brief guide to reading and interpreting the test results.

Psychometric properties

Periodic reassessment and ongoing modification are essential to ensure that all our tests remain relevant and applicable. Central Test regularly carries out thorough reviews of the psychometric properties of all its evaluation tools. These revisions involve updating of the questionnaire content, reviewing the statistical analysis to ensure it is still pertinent, and carrying out validity studies. Before any test is published, it is evaluated in terms of the following:

- Reliability and precision
- Validity
- Social desirability
- Socio-legal considerations
- Calibration

Reliability and precision

Reliability refers to how consistently a test measures a particular personality trait or competence. If an individual takes a test more than once, will he or she get the same score each time? A test that gives similar scores for an attribute or skill time after time is said to measure that trait reliably, or precisely. But we should always take into consideration that in behavioural science, measurement is always influenced by a number of external variables such as:

– [The passing psychological and physical state of the person taking the test](#)

Degrees of fatigue and levels of anxiety or motivation vary for all of us from day to day. Such factors influence how we perform on a test and will have a bearing on our results.

– [Environmental factors](#)

Environmental differences in the test location, such as variations in temperature, ambient light and background noise, can all affect performance and frame of mind and influence results.

These and other factors are commonly referred to as “random measurement errors”. If there was no random measurement error, individuals would have the same test score, their “true score” each time they took the test. The degree to which a psychometric test is resistant to random measurement errors is indicative of how reliable it is. Reliable assessment tests produce dependable, consistent information about test-takers’ personality traits and skills. There are several types of reliability measurement that affect psychometric testing, and each is influenced by different categories of measurement error. For tests developed at Central Test, two types are particularly relevant:

- [Internal consistency reliability](#)

This refers to the reliability of the way in which the questions within a test measure a particular characteristic or ability. Internal consistency is usually measured according to Cronbach’s alpha coefficient of reliability, which ranges from 0 (low reliability) to 1 (high reliability). A high coefficient indicates that the questions in the test are similar in content, or uniform. It is important to note that the number of

questions in a test can also affect its internal consistency and a very long test can yield an inflated reliability coefficient.

- **Test-retest reliability**

This refers to how consistently the findings of a test are repeated with the passage of time. Test-retest reliability also depends on the stability of the characteristic or ability being measured. Some characteristics are more stable than others. For example, an individual's reading ability is more stable over a period of time than is his level of anxiety. So the test-retest reliability coefficient is likely to be lower for traits that are assumed to vary with the passage of time.

An acceptable measurement of reliability will depend on the type of test that's being taken in combination with the coefficient that's being applied.

Validity

Validity is the most important criterion to bear in mind when you are choosing a test. It refers to the accuracy with which a test measures the personality traits or abilities it sets out to evaluate. Validity gives meaning to the test scores and gives authority to the link between how an individual performs in the test and the test's stated measurement criteria. It tells us about the degree to which it is possible to draw specific conclusions or predictions, based on an individual's test score; in other words, it provides information on the efficacy of the test. For example, if a test is a valid predictor of performance for a particular job, we can conclude that someone with a high score will be more suitable for that post than a person with a low score – all things being equal.

It is important to understand the difference between reliability and validity. Validity informs us about the extent to which a test accurately evaluates an individual's abilities or personality; reliability tells us the extent to which a test is stable or consistent. Test validity is established in reference to a specific objective.

Researchers refer to many different kinds of validity. However, all can be classified into three broad categories:

- **Criterion-related validity**

This is measured by examining the correlation between test performance and an external criterion such as on-the-job performance. In other words, we should expect individuals who get a high score in the test to perform better on the job than individuals with a low test score. If the external criterion (job performance) takes place at the same time as the test, we call it **concurrent validity**. If the test takes place first and the external criterion is examined later, we call it **predictive validity**.

- **Content-related validity**

This is studied by examining whether the test content is representative of the aspects of job-related behaviour that we want to measure. For example, in a test assessing sales potential, the test questions must be relevant to the requirements of the job, and the results must measure how well these requirements and qualifications have been met. Test content should also reflect the level of understanding of the target population.

- **Construct-related validity**

This requires a demonstration that the test measures the trait or ability (construct) it claims to measure. For example, a test for a specific professional aptitude must

assess an individual's flair for that profession, and not their level of motivation. Construct-related validity is generally assessed in two distinct ways:

- By asking a panel of experts to assess how well the test questions correlate with the particular ability or characteristic that the test sets out to measure (**face validity**)
- By administering the test alongside established tests developed to measure the same constructs, and examining the correlation between them (**convergent validity**); or by administering the test alongside established tests developed to measure converse constructs, and examining their correlation (**divergent validity**)

The three types of validity, criterion-related, content-related, and construct-related, are used in different circumstances to support the validation of tests. In general, the three methods overlap but not all are appropriate to every situation.

At Central Test, validity studies are carried out in-house as well as by external experts.

Social Desirability

A social desirability bias describes people's inclination to present themselves in a more favourable light. Because we are social beings, people tend to seek out social acceptance and the approval of others. The need to appear socially desirable is one of the major biases in personality tests, and it is a factor that is even more pronounced in the context of recruitment, where a candidate is often tempted to give a socially acceptable answer in order to impress the recruiter.

We have developed a strategy to overcome the problem of social desirability bias. All our test questions are reviewed to ensure they are not susceptible to social desirability bias.

Socio-cultural considerations

It's important that a psychometric test should not be biased either towards or against a particular socio-cultural group. It should not discriminate on the basis of the religion, gender, ethnicity or culture of the person who passes the test. When developing our tools, all our questions are pitched so that they do not favour a particular population. During validation studies, the relationships between demographic variables and test scores are reviewed in order to prevent potential bias. This allows us to build sensitive and unbiased tests.

Calibration

In terms of measuring skills and characteristics, it is useful to compare the results of different individuals. In order to do this, we have devised calibration systems that group scores according to how frequently they occur in the reference population. The reference population is composed of a representative sample of people who share the same sociological characteristics (age, education level, socio-professional group, gender, etc.) as the candidates who are taking the test. A score that is calibrated positions the test-takers in relation to the rest of the reference population.

All Central Test's aptitude tests have calibrated scores.