

USER GUIDE

PSYCHOMETRIC

VALIDATION

*This user guide has been created by the psychologists from the
Research and Development Department of Central Test.
No reproduction, modification or adaptation is permitted
without the written consent of the publisher Central Test.*

Central Test - Empowering Talent © 2022

All rights reserved

Table of contents

Introduction.....	2
Reliability	3
Validity	4
Social Desirability	6
Socio-legal Considerations.....	7
Conclusion	9

Introduction

Periodic reviews of psychometric properties are crucial for any psychometric assessment's long-term survival and effectiveness. *Central Test*, therefore, takes the greatest care to assess the psychometric properties of all its assessments regularly. Doing so also ensures that all our assessments comply with bodies such as the Equal Employment Opportunity Commission (EEOC). These reviews are done by regularly updating the norms and statistical analysis on the reports and manuals, as well as training users on the ethical obligations to be fulfilled while administering the assessments and interpreting its results.

Central Test's assessments undergo content review, statistical analyses and independent validity studies. Every assessment is mandatorily assessed against four distinct criteria before publication: *Reliability, Validity, Social Desirability and Socio-legal Considerations*.

Reliability

Reliability refers to how consistently an assessment measures a characteristic. When an assessment yields similar scores for an individual who repeats the assessment, then it is said to measure the characteristic reliably. It is, however, notable here that the measurement in behavioural sciences is always influenced by certain external variables. These could be:

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

These and other similar factors are sources of chance or random measurement errors in the assessment process, and thus the degree to which the assessment scores are unaffected by such measurement errors is an indication of the reliability of the assessment. Therefore, reliable assessment tools produce dependable, repeatable, and consistent information about individuals. However, there are several types of reliability estimates, each influenced by different sources of measurement error. For the assessments developed by *Central Test*, two kinds of reliabilities are particularly important.

- **Internal consistency reliability** of each factor indicates the extent to which items on an assessment measure the same construct. A high internal consistency reliability coefficient indicates that the items on the factor are very similar to each other in content (homogeneous). It is important to note that the length of an assessment also affects internal consistency. A very long assessment, therefore, can spuriously have an 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 assessment scores over time. These estimates also reflect the stability of the characteristic or construct being measured by the assessment, with some constructs being more stable than others. For example, an individual's reading ability is more stable over a particular period than their anxiety level. Therefore, one would expect a higher test-retest reliability coefficient on a reading assessment than on an assessment that measures anxiety. For constructs that are expected to vary over time, an acceptable test-retest reliability coefficient may be lower than those constructs that are stable over time.

Test-retest reliability is reported as the correlation between two administrations of the assessment to the same sample. As quality-conscious assessment developers, *Central Test* reports reliability estimates that are relevant to a particular assessment while taking into consideration that the acceptable level of reliability differs depending on the type of assessment and the reliability estimate used.

Validity

Validity refers to what characteristic an assessment measures and how well it measures that characteristic. Hence, validity studies tell us if the characteristics being measured by an assessment are related to the requirements of an assessment situation. Validity gives meaning to the assessment scores, and validity evidence indicates that there is a linkage between assessment performance and job performance. Therefore, if an assessment has been demonstrated to be a valid predictor of performance on a specific job, one can conclude that individuals scoring high on the assessment are more likely to perform well on the job than those who score low on the assessment, other things being equal. Therefore, validity

describes the degree to which one can make specific conclusions or predictions about individuals based on their assessment scores. In other words, it indicates the usefulness of the assessment.

It is important to understand the differences between reliability and validity. Reliability indicates how trustworthy a score on that assessment will be, while validity indicates how good an assessment is for a particular situation. An assessment's validity is established concerning a specific purpose. In all *Central Test's* assessment documentation, the assessment context and target group on which the assessment is validated are mentioned.

There are several types of validity conceptualised by researchers and behavioural scientists, and they can be grouped into three distinct categories.

- **Criterion-related validity** is assessed by examining correlation or other statistical relationships between assessment performance and job performance. In other words, individuals who score high on the assessment tend to perform better on the job than those who score low on the assessment. If the criterion is obtained at the same time the assessment 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 assessment represents important job-related behaviours. In other words, assessment items should be relevant to and measure directly important requirements and qualifications for the job. The assessment content should correspond to the reading level and domain of the target population.
- **Construct-related validity** requires a demonstration that the assessment measures the construct or characteristic it claims to measure, i.e. the assessment content represents the underlying construct. For example, an assessment for occupational interest should measure occupational interests

and not motivations. Construct validity is generally examined in one of three ways:

- By giving the items of the assessment to a panel of experts and asking for their judgement on the proximity of the assessment content with the construct of the assessment (**face validity**).
- By administering the assessment along with other established assessments developed on theoretically similar constructs and examining the correlation between the two (**convergent validity**) or by administering the test along with a theoretically opposite assessment and exploring the correlation (**divergent validity**).
- By reviewing that the factors of the assessment are in accordance with the theoretical model proposed for the construction of the assessment (**confirmatory factor analysis**).

The three types of validity – criterion, 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.

Central Test conducts in-house as well as third-party validity studies to examine the validity of the assessments. The summary findings of these studies are given in technical documents of the assessments and used for further improvement of the assessment.

Social Desirability

Social desirability bias is the inclination to present oneself in a manner that will be viewed favourably by others and is one of the principal biases of an assessment. Being social creatures by nature, people are generally inclined to seek some degree of social acceptance. When done unintentionally, it is referred to as self-deception. When done intentionally, it is referred to as impression management (Tett & Simonet, 2011) ¹.

¹ Tett, R., & Simonet, D. (2021). *Applicant Faking on Personality Tests: Good or Bad and Why Should We Care? Personnel Assessment and Decisions*, 7(1).

To minimise the risk of faking the results (done intentionally or not), social desirability is controlled in psychometric assessments, where relevant, by means of a social desirability scale or within the format of the assessment.

- **Ipsative assessments** actively guard against social desirability with the use of a questionnaire format which includes questions that are presented in a forced-choice format, whereby each question is composed of general statements that require the individual to choose the option that best represents their preference. Research has found ipsative measures to be more resistant to distorted responding compared to more commonly used Likert scales (Bowen, Martin, & Hunt ², 2002; Kluger, Reilly, & Russell, 1991)³.
- In **normative assessments**, social desirability is measured by adding questions to the assessments as part of the social desirability scale. They are constructed in a way that is similar to the other items in the assessment. This is done to ensure that they are not easily detected, and the assessment will include a social desirability score as part of the report.

Socio-legal Considerations

Another important characteristic of a good psychometric assessment is that it should not be biased (positively or negatively) towards any socio-cultural group. Assessments can be discriminatory on four major aspects - age, disability, country of origin and gender. Therefore, when developing assessments, all of the items are examined against the appropriateness of content via statistical measures like -

- **ANOVA and t-tests**, that is used to compare the means of groups.

<https://doi.org/10.25035/pad.2021.01.002>

² Bowen, C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *The International Journal of Organizational Analysis*, 10(3), 240-259.

³ Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology*, 76(6), 889-896. <http://doi.org/10.1037/0021-9010.76.6.889>

- **Differential Item Functioning**, which provides information on whether an item is perceived in a similar way across different groups.

Countermeasures are also undertaken during development to ensure that the assessments are not biased toward the target population concerning demographic criteria like –

- **Age** – Norms are calculated, and statistical analyses are run to ensure that the items of the assessment do not favour a certain age group over the other.
- **Disability** – The assessments offer accessibility while incorporating dyslexic-friendly fonts.
- **Country of origin** – Norms are calculated, and statistical analyses are run to ensure that there is no cultural bias.
- **Gender** – Norms are calculated, and statistical analyses are run to negate any bias based on gender.

All details of the norms and statistical analyses carried out are mentioned in the technical manuals of the assessments for further understanding.

Conclusion

The degree to which an assessment is theoretically, empirically and statistically analysed determines how good its psychometric properties will be and, ultimately, how well-developed a psychological assessment will be. Therefore, *Central Test* measures these criteria before an assessment is published to ensure its relevancy and fitment to the evolving times.

CONTACT US

Client Service: customer@centraltest.com

Training and Consulting: consulting@centraltest.com



www.centraltest.com

© 2022 by Central Test – Empowering Talent.

All rights reserved.