What is Validity?

The concept of validity was formulated by Kelly (1927, p. 14) who stated that a test is valid if it measures what it claims to measure.

For example a test of intelligence should measure intelligence and not something else (such as memory).

**Internal and External Validity**

A distinction can be made between internal and external validity. These types of validity are relevant to evaluating the validity of a research study / procedure.

Internal validity refers to whether the effects observed in a study are due to the manipulation of the independent variable and not some other factor. In-other-words there is a causal relationship between the independent and dependent variable.

Internal validity can be improved by controlling extraneous variables, using standardized instructions, counter balancing, and eliminating demand characteristics and investigator effects.

External validity refers to the extent to which the results of a study can be generalized to other settings (ecological validity), other people (population validity) and over time (historical validity).

External validity can be improved by setting experiments in a more natural setting and using random sampling to select participants.

**Assessing the validity of test**

There there are two main categories of validity used to assess the validity of test (i.e. questionnaire, interview, IQ test etc.): Content and criterion.

![Types of Validity](image-url)
What is Validity?

Face Validity

This is the least sophisticated measure of validity. Face validity is simply whether the test appears (at face value) to measure what it claims to.

Tests wherein the purpose is clear, even to naïve respondents, are said to have high face validity. Accordingly, tests wherein the purpose is unclear have low face validity (Nevo, 1985).

A direct measurement of face validity is obtained by asking people to rate the validity of a test as it appears to them. This rater could use a likert scale to assess face validity. For example:

1. the test is extremely suitable for a given purpose
2. the test is very suitable for that purpose;
3. the test is adequate
4. the test is inadequate
5. the test is irrelevant and therefore unsuitable

It is important to select suitable people to rate a test (e.g. questionnaire, interview, IQ test etc.). For example, individuals who actually take the test would be well placed to judge its face validity. Also people who work with the test could offer their opinion (e.g. employers, university administrators, employers). Finally, the researcher could use members of the general public with an interest in the test (e.g. parents of testees, politicians, teachers etc.).

The face validity of a test can be considered a robust construct only if a reasonable level of agreement exists among raters.

It should be noted that the term face validity should be avoided when the rating is done by "expert" as content validity is more appropriate.

Having face validity does not mean that a test really measures what the researcher intends to measure, but only in the judgment of raters that it appears to do so. Consequently it is a crude and basic measure of validity.

A test item such as 'I have recently thought of killing myself' has obvious face validity as an item measuring suicidal cognitions, and may be useful when measuring symptoms of depression.

However, the implications of items on tests with clear face validity is that they are more vulnerable to social desirability bias. Individuals may manipulate their response to deny or hide problems, or exaggerate behaviors to present a positive images of themselves.

It is possible for a test item to lack face validity but still have general validity and measure what it claims to measure. This is good because it reduces demand characteristics and makes it harder for respondents to manipulate their answers.

For example, the test item 'I believe in the second coming of Christ' would lack face validity as a measure of depression (as the purpose of the item is unclear).

This item appeared on the first version of The Minnesota Multiphasic Personality Inventory (MMPI) and loaded on the depression scale. Because most of the original normative sample of the MMPI were good Christians only a depression Christian would think Christ is not coming back. Thus, for this particular religious sample the item does have general validity, but not face validity.
What is Validity?

**Construct Validity**

Construct validity was invented by Cornball and Meehl (1955). This type of validity refers to the extent to which a test captures a specific theoretical construct or trait, and it overlaps with some of the other aspects of validity.

Construct validity does not concern the simple, factual question of whether a test measures an attribute. Instead, it is about the complex question of whether test score interpretations are consistent with a nomological network involving theoretical and observational terms (Cronbach & Meehl, 1955).

To test for construct validity it must be demonstrated that the phenomenon being measured actually exists. So, the construct validity of a test for intelligence, for example, is dependent on a model or theory of intelligence. Construct validity entails demonstrating the power of such a construct to explain a network of research findings and to predict further relationships.

The more evidence a researcher can demonstrate for a test's construct validity the better. However, there is no single method of determining the construct validity of a test. Instead, different methods and approaches are combined to present the overall construct validity of a test. For example, factor analysis and correlational methods can be used.

**Concurrent validity**

This is the degree to which a test corresponds to an external criterion that is known concurrently (i.e. occurring at the same time). If the new test is validated by a comparison with a currently existing criterion, we have concurrent validity. Very often, a new IQ or personality test might be compared with an older but similar test known to have good validity already.

**Predictive validity**

This is the degree to which a test accurately predicts a criterion that will occur in the future. For example, a prediction may be made on the basis of a new intelligence test, that high scorers at age 12 will be more likely to obtain university degrees several years later. If the prediction is born out then the test has predictive validity.

**References**


**How to cite this article:**