What is validity?

It is the accuracy of the measure in reflecting the concept it is supposed to measure.

In simple English, the validity of a test concerns

- what the test measures and
- how well it measures

Types of validity

- Content-description
- Criterion-description
- Construct-identification

Content validity

- non-statistical in nature
- involves determining whether the sample used for the measure is representative of the aspect to be measured

Content validity

It adopts a subjective approach whereby we have recourse, for e.g., to expert opinion for the evaluation of items during the test construction phase.
Content validity

Relevant for evaluating
- Achievement
- Educational
- Occupational
measures.

Content validity

Basic requirement for
- Criterion-referenced
- Job sample
measures, which are essential for
- Employee selection
- Employee classification

Content validity

Measures are interpreted in terms of
- Mastery of knowledge
- Skills
for a specific job.

Content validity

Not appropriate for
- Aptitude
- Personality
measures since validation has to be made through criterion-prediction procedures.

Face validity

It is not validity in psychometric terms!
It just refers to what the test appears to measure and not to what it measures in fact.

Face validity

It is not useless since the aim may be achieved by using appropriate phrasing only!
In a sense, it ensures relevance to the context by employing correct expressions.
**Criterion-related validity**

**Definition**

A criterion variable is one with (or against) which psychological measures are compared or evaluated.

A criterion must be reliable!

**Criterion-related validity**

It is a quantitative procedure which involves calculating the correlation coefficient between one or more predictor variables and a criterion variable.

**Criterion-related validity**

The validity of the measure is also determined by its ability to predict performance on the criterion.

**Warning!**

Sometimes a factor may affect the criterion such that it is no longer a valid measure. This is known as criterion contamination.

**Criterion-related validity**

- **Concurrent validity**
  Accuracy to identify the current status regarding skills and characteristics

- **Predictive validity**
  Accuracy to forecast future behaviour. It implicitly contains the concept of decision-making

**Criterion-related validity**

**Warning!**

For example, the rater might:

- Be too lenient
- Commit halo error, i.e., rely on impressions
**Criterion-related validity**

**Warning!**
Therefore, we must make sure that the criterion is free from bias (prejudice).
Bias definitely influences the correlation coefficient.

---

**Common criterion measures**

- Academic achievement
- Performance in specialised training
- Job performance
- Contrasted groups
- Psychiatric diagnoses

---

**Common criterion measures**

- **Academic achievement**
  Used for validation of intelligence, multiple aptitude and personality measures.
  Indices include
  - School, college or university grades
  - Achievement test scores
  - Special awards

---

**Common criterion measures**

- **Performance in specialised training**
  Used for specific aptitude measures
  Indices include training outcomes for
  - Technical courses
  - Academic courses

---

**Common criterion measures**

- **Job performance**
  Used for validating intelligence, special aptitude and personality measures
  Indices include jobs in industry, business, armed services, government
  Tests describe duties performed and the ways that they are measured

---

**Common criterion measures**

- **Contrasted groups**
  Sometimes used for validating personality measures
  Relevant when distinguishing the nature of occupations (e.g., social and non-social – Public Relations Officer and Clerk)
Common criterion measures

Psychiatric diagnoses
Used mainly for validating personality measures
Based on
- Prolonged observation
- Case history

Ratings

- Suitable for almost any type of measure
- Very subjective in nature
- Often the only source available

Criterion-related validity

We can always validate a new measure by correlating it with another valid test (obviously, reliable as well!)

Common criterion measures

<table>
<thead>
<tr>
<th></th>
<th>Academic achievement</th>
<th>Performance in specialised training</th>
<th>Job performance</th>
<th>Contrast groups</th>
<th>Psychiatric diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligence</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aptitude</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Personality</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Ratings

These are given by teachers, lecturers, instructors, supervisors, officers, etc,…

Raters may be trained to avoid common errors like
- Halo error
- Ambiguity
- Error of central tendency
- Leniency

Criterion-related validity

Some modern and popular criterion-prediction procedures which are now widely used are
- Validity generalisation
- Meta-analysis
- Cross-validation
Criterion-related validity

Validity generalisation
Schmidt, Hunter et al. showed that the validity of tests measuring verbal, numeric and reasoning aptitudes can be generalised widely across occupations (these require common cognitive skills).

Meta-analysis
- Method of reviewing research literature
- Statistical integration and analysis of previous and current findings on a topic.
- Validation by correlation

Criterion-related validity

- Cross-validation
  - Refinement of initial measure
  - Application to another representative normative sample
  - Recalculation of validity coefficients
  - Lowering of coefficient expected after minimisation of chance differences and sampling errors (spuriousness)

Construct-identification validity

- Construct validity is the sensitivity of the instrument to pick up minor variations in the concept being measured.
  - Can an instrument (questionnaire) to measure anxiety pick up different levels of anxiety or just its presence or absence?

Construct-identification validity

Any data throwing light on the nature of the trait and the conditions affecting its development and manifestations represent appropriate evidence for this validation.

Example
I have designed a program to lower girls’ Math phobia. The girls who complete my program should have lower scores on the Math Phobia Measure compared to their scores before the program and compared to the scores of girls who have not completed the program.

Construct validity methods

- Correlational validity
- Factor analysis
- Convergent and discriminant validity
Construct validity methods

Correlational validity
This involves correlating a new measure with similar previous measures of the same name.
Warning!
High correlation may indicate duplication of measures.

Factor analysis (FA)
It is a multivariate statistical technique which is used to group multiple variables into a few factors.
In doing FA you hope to find clusters of variables that can be identified as new hypothetical factors.

Convergent and discriminant validity
The idea is that a test should correlate highly with other similar tests and the test should correlate poorly with tests that are very dissimilar.

Example
A newly developed test of motor coordination should correlate highly with other tests of motor coordination. It should also have low correlation with tests that measure attitudes.

Indices and interpretation of validity
- Validity coefficient
  - Magnitude of coefficient
  - Factors affecting validity
- Coefficient of determination
- Standard error of estimation
- Regression analysis (prediction)

Validity coefficient
Definition
It is a correlation coefficient between the criterion and the predictor(s) variables.
Indices and interpretation of validity

Validity coefficient

Differential validity refers to differences in the magnitude of the correlation coefficients for different groups of test-takers.

Indices and interpretation of validity

Magnitude of validity coefficient

Treated in the same way as the Pearson correlation coefficient!

Indices and interpretation of validity

Factors affecting validity

Nature of the group

Sample heterogeneity

Criterion-predictor relationship

Validity-reliability proportionality

Criterion contamination

Moderator variables

Indices and interpretation of validity

Factors affecting validity

Nature of the group

Consistency of the validity coefficient for subgroups which differ in any characteristic (e.g. age, gender, educational level, etc, …)

Indices and interpretation of validity

Factors affecting validity

Sample heterogeneity

A wider range of scores results in a higher validity coefficient (range restriction phenomenon)

Indices and interpretation of validity

Factors affecting validity

Criterion-predictor relationship

There must be a linear relationship between predictor and criterion. Otherwise, the Pearson correlation coefficient would be of no use!
Factors affecting validity

Validity-reliability proportionality
Reliability has a limiting influence on validity – we simply cannot validate an unreliable measure!

Factors affecting validity

Criterion contamination
Get rid of bias by measuring contaminated influences. Then correct this influence statistically by use of partial correlation.

Moderator variables
Variables like age, gender, personality characteristics may help to predict performance for particular variables only – keep them in mind!

Coefficient of determination
Indicates the proportion of variance in the criterion variable explained by the predictor.
E.g. If \( r = 0.9 \), \( r^2 = 0.81 \). 81% of the changes in the criterion is accounted for by the predictor.

Regression analysis
Mainly used to predict values of the criterion variable.
If \( r \) is high, prediction is more accurate.
Predicted values are obtained from the line of best fit.

Standard error of estimation (SE)

\[
SE_{est} = s_y \sqrt{1 - r_{xy}^2}
\]
Treated just like the standard deviation. (True and predicted values for the criterion should differ by at most 1.96SE at a 95% confidence level.)
Indices and interpretation of validity

Regression analysis
Linear regression
It involves one criterion variable but may involve one (simple regression) or more than one predictor variable (multiple regression).

Reliability and Validity

<table>
<thead>
<tr>
<th>IF</th>
<th>THEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unreliable</td>
<td>Test validity is undermined.</td>
</tr>
<tr>
<td>Reliable, but not valid</td>
<td>Test is not useful.</td>
</tr>
<tr>
<td>Unreliable and invalid</td>
<td>Test is definitely NOT useful!</td>
</tr>
<tr>
<td>Reliable and valid</td>
<td>Test can be used with good results.</td>
</tr>
</tbody>
</table>

Optimising reliability and Validity

- The more questions the better (the number of test items)
- Ask questions several times in slightly different ways (homogeneity)
- Get as many people as you can in your program (sample size n)
- Get different kinds of people in your program (sample heterogeneity)
- Linear relationship between the test and the criterion (Pearson correlation coefficient)

Selecting and creating measures

- Define the construct(s) that you want to measure clearly
- Identify existing measures, particularly those with established reliability and validity
- Determine whether those measures will work for your purpose and identify any areas where you may need to create a new measure or add new questions
- Create additional questions/measures
- Identify criteria that your measure should correlate with or predict, and develop procedures for assessing those criteria