

Reliability and Validity

Reliability is the relationship among items within an assessment or between different administrations of the assessment. It is an estimate of how closely an observed score is related to what the “true” score would be if measured perfectly.

Type of reliability	Description
Test-retest reliability	The degree to which an individual’s scores on an assessment are related to their scores on a subsequent administration of that assessment.
Interrater reliability	The degree to which different raters produce the same rating on an assessment or assessment domain.
Split-Half	The relationship between half the items on a test to the rest of the items.
Internal reliability (Cronbach’s alpha)	The relationship among items within an assessment or within assessment domains. Estimates the expected correlation between two tests that measure the same construct
Generalizability theory	<p>A way of thinking about reliability, which leads to procedures for choosing the reliability coefficient or error variance most appropriate to the situation at hand.</p> <ul style="list-style-type: none"> • G-Studies are used to test the extent to which a sample of measurements generalizes to a universe of measurements • D-Studies are studies in which data are collected for the specific purpose of making a decision.
Alternate forms (or parallel) reliability	The relationship between scores on two versions of the same assessment.

Validity is the degree to which evidence and theory support the use of a measure for a specific purpose. It gives an estimate of the usefulness of inferences drawn from test scores.

Type of validity	Description
Predictive validity	The degree to which assessment scores predict related measures scores at a later time. Predictive validity can be one form of test-criterion validity.
Construct validity: Evidence based on internal structure	The degree to which the relationships among assessment items and domains (for example, components of the assessment) align with the intended construct(s) proposed for scoring and use. This category of analyses may also examine the extent to which there is potential bias in items or domain scores across subgroups.
Content validity	The relationship between the content of an assessment and the construct(s), or in other words the concept(s) or characteristic(s), the assessment is intended to measure.
Response process validity	The extent to which the actions and thought processes of test takers or survey responders demonstrate that they understand the construct in the same way it is defined by the researchers.
Criterion-related validity	The degree to which test scores and criterion performance results are related.
Concurrent validity	The degree to which the assessment correlates with other independent, reliable, and valid assessments of the same construct. Concurrent validity is established by examining convergent and discriminant relationships. Concurrent validity evidence can be one form of test-criterion validity.
Convergent validity	Convergent validity is a type of concurrent validity evidenced by two assessments of the same construct having a strong relationship.
Divergent validity	Discriminant validity is a type of concurrent validity evidenced by a weak relationship between two assessments of purportedly different constructs.