

---

# TIP SHEET: PSYCHOMETRICS BASICS

Reliability and Validity Considerations in Survey Research

Psychometrics is the field of psychology that is concerned with the measurement of attitudes, behaviors, skills, and beliefs. This document summarizes the psychometric properties to consider when developing, refining, or choosing survey tools.

## RELIABILITY

Reliability is defined as the ability of a measurement tool to produce consistent findings. There are several types of reliability, including internal reliability, separation reliability, test-retest reliability, and inter-rater reliability.

- **Internal reliability:** For survey scales (multiple items intended to measure the same construct), internal consistency measures show the extent to which individuals provide consistent responses to the items. The most common measure of internal consistency is a statistic known as Cronbach's alpha ( $\alpha$ ). Cronbach's  $\alpha$  values range between 0 and 1, with higher values indicating greater internal consistency.
- **Separation reliability:** Person separation reliability refers to the ability of a measurement tool to distinguish between high and low performers in a given skill area. It is possible for a measurement tool to have very high internal consistency alongside very low person-separation reliability if all people are scoring similarly on the measure.
- **Test-retest reliability:** The extent to which a measurement tool produces consistent results when there has been no change in the construct being measured. Typically test-retest reliability is assessed by administering the same survey to the same individual at two (close together) time intervals and analyzing the correlation between the results. The shorter the time interval that has elapsed, the higher the correlation should be.
- **Inter-rater reliability:** This is a consideration when using observational assessments. It is defined as the extent to which two or more observers of the same phenomenon will rate it similarly on a given measurement tool.

## VALIDITY

Validity is another key consideration in measurement. It is generally defined as the extent to which a measurement tool is capturing what it is intended to capture. Validity is less straightforward than reliability but is similar in that it has several relevant dimensions. Two important ones are:

- **Face validity:** A subjective interpretation of whether a measurement tool makes sense "on its face" (i.e., in light of theoretical and empirical evidence). For a survey scale to have face validity, items would need to appear to practitioners to be related to each other, and to the construct they are intended to measure.
- **Criterion-related validity:** The validity of a measurement tool relative to a related, external criterion that is known to be valid. The relationship can be predictive (e.g., does a students' self-report on their study skills correlate with their grades later in the year?) or concurrent (do youth who report higher levels of engagement also have better attendance?).

## OTHER CONSIDERATIONS

There are other types of reliability and validity, and other factors to consider, such as responsiveness (or sensitivity) to meaningful change and susceptibility to learning effects over time. Typically when we describe something as "validated," this refers to a particular purpose, population, and point in time. Ideally, evidence of validity accrues over time and repeated investigation. Just because something is validated does not mean it is useful in all circumstances.