

AM Last Page: Reliability and Validity in Educational Measurement

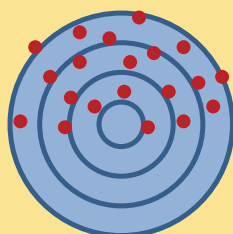
Anthony R. Artino, Jr., PhD, assistant professor of preventive medicine and biometrics, Steven J. Durning, MD, professor of medicine, and Alisha H. Creel, PhD, assistant professor of social and behavioral sciences, Uniformed Services University of the Health Sciences

Reliability is the extent to which the scores produced by a particular measurement tool or procedure are consistent and reproducible.¹ Reliability answers the question, "Does the assessment yield the same scores at different times, from different raters, or from different items?"

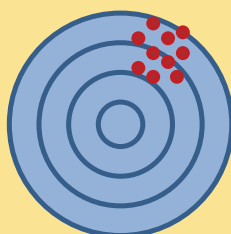
Validity is the degree to which an assessment measures what investigators want to measure, all of what they want to measure, and nothing but what they want to measure.¹ Validity answers the question, "Does the assessment provide information that is relevant to the inferences that are being made from it?" An assessment, such as a test or questionnaire, does not have validity in any absolute sense. Instead, the scores produced are valid for some uses and not valid for others.

A target provides a metaphor for the relationship between reliability and validity.

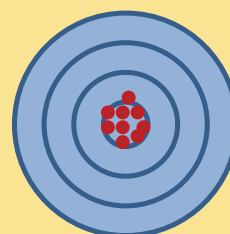
The true score (or value) for the concept the researcher is attempting to measure is at the center of the target, and the observed score the investigator gets from each person assessed is a shot at the target.



Neither reliable
nor valid



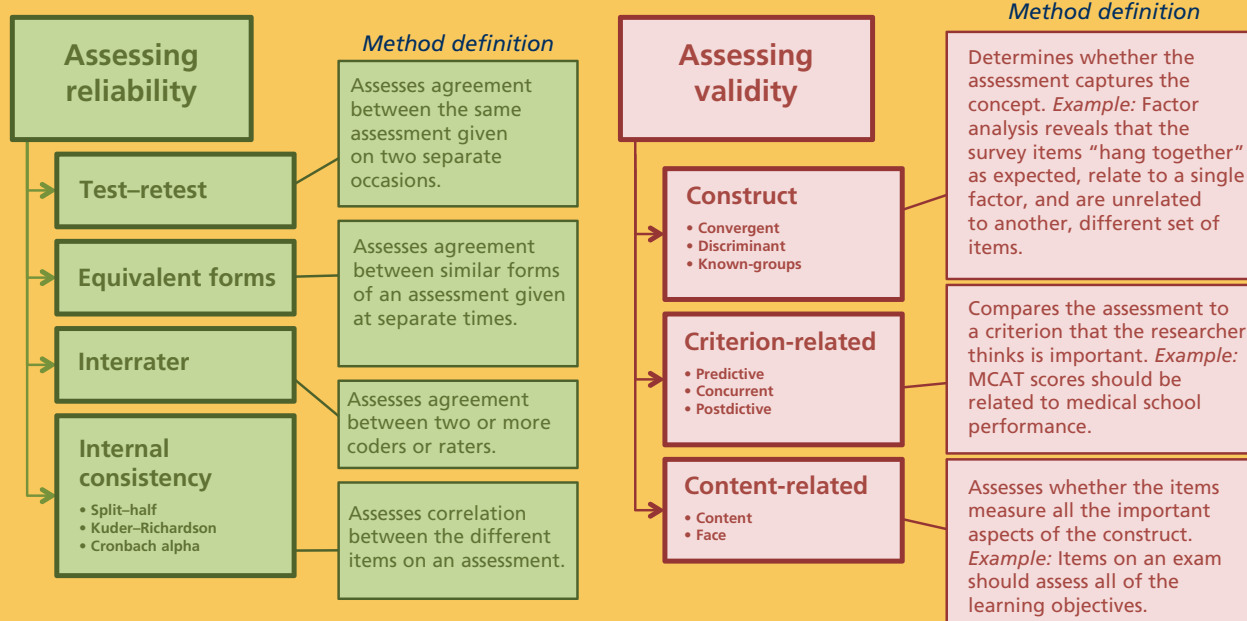
Reliable, but
not valid



Both reliable
and valid

Reliability is a *necessary but insufficient* condition for validity. To be valid, scores must first be at least moderately reliable.¹⁻³ However, scores that are reliable may be devoid of validity for the application the researcher has in mind.¹

Many methods of assessing reliability and validity are available.¹⁻⁴ Each method provides the researcher with slightly different information about the reliability and validity of the assessment.



References

1. Thorndike RM. Measurement and Evaluation in Psychology and Education. 7th ed. Upper Saddle River, NJ: Pearson Education; 2005.
2. American Educational Research Association; American Psychological Association; National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 1999.
3. Downing SM. Validity: On the meaningful interpretation of assessment data. *Med Educ.* 2003;37:830-837.
4. Kane MT. Validation. In: Educational Measurement, 4th ed. Brennan RL, ed. Westport, CT: American Council on Education and Praeger Publishers; 2006:17-64.