

Validity

The extent to which an assessment instrument measures what it is supposed to measure

Reliability

The consistency of assessment results

Reliability is a necessary but not sufficient condition for validity.

About validity . . .

- **Refers to the appropriateness of the interpretation and use made of the results not the procedure itself.**
- **It is a matter of degree--not all or nothing.**
- **It is specific to some particular use or interpretation--no assessment is valid for all purposes.**

Can have a reliable, but invalid measure.

If measure is valid, then necessarily reliable.

Sources of Validity Evidence

- **Does the assessment tap into a representative sample of the content domain being assessed (content validity)**
- **Does the instrument measure a particular psychological or educational characteristic (construct validity)**
- **Do students' scores predict their success on a later task (predictive validity/criterion-related validity)**
- **What are the effects of using a particular assessment (consequential validity)**
- **Does the assessment appear to be a reasonable measure? (face validity)**

Content Validity

- **Important for both classroom tests and when selecting standardized tests.**
- **On a classroom test this takes much forethought and developed skill to ensure this.**

EG: SAY YOUR TEXTBOOK HAD 12 CHAPTERS ON PSYCHOMETRICS AND YOUR EXAM ONLY COVERED 2 – NOT A VALID EXAMINATION OF THE ENTIRE DOMAIN.

Construct Validity

- **Whether or not an abstract, hypothetical concept exists as postulated**
- **Examples of Constructs:**
 - **Intelligence**
 - **Happiness**
 - **Impulsivity**

Based on:

- **Convergence = different measures that purport to measure the same construct should be highly correlated (similar) with one another**
- **Divergence = tests measuring one construct should not be highly correlated (similar) to tests purporting to measure other constructs**

Construct Underrepresentation -- the extent to which aspects of the construct are underrepresented in the assessment

Construct Irrelevant Variance -- performance is influenced by irrelevant factors (e.g., reading level on a math test)

Methods for Construct Validation

- **Defining the domain or tasks to be measured**
- **Analyzing the response process required by the assessment tasks**
- **Comparing the scores of known groups**
- **Comparing scores before and after a particular learning experience**
- **Correlating the scores with other measures**

Criterion Related Validity

DOES THE TEST RELATE TO ANOTHER ESTABLISHED MEASURE OF THE SAME PHENOMENON?

2 TYPES:

1) PREDICTIVE – DOES TEST PREDICT FUTURE PERFORMANCE ON RELATED MEASURE?

EG DO SAT SCORES PREDICT COLLEGE GPA?

2) CONCURRENT – COMPARE TEST TO EXISTING ONE – DO SCORES RELATE?

EG: NEW IQ TEST SCORES AND WAIS SCORES.

Prediction Design

- **Predictor Variable:** a variable that is used to make a forecast about an outcome in the correlational study.
- **Criterion Variable:** the outcome being predicted

E.g.,

Predictor = High school grades, self-efficacy, SATs

Criterion = Grades in first year of college, dropout rate

Prediction Design: Characteristics

- The authors typically include the word “prediction” in the title
- The researchers typically measure the predictor variables at one point in time and the criterion variable at a later point in time.
- The authors are interested in forecasting future performance

Criterion Validity Coefficients & Linear Regression

- **Linear regression is used to predict scores on one variable from scores on another.**
- **We can use linear regression to predict criterion scores given predictor (i.e., test) scores.**

Criterion Validity Coefficients & Linear Regression

- **If we have a strong correlation between the test and the criterion (r_{xy}), the linear regression model will be relatively accurate in predicting criterion scores.**
- **If the correlation is weak, prediction will be less accurate.**

Considering Consequences

- **Must be aware of “teaching to the test” and narrowing the focus of instruction to fit one particular test**

Threats to Validity

- **Numerous ways validity can be threatened**
 - **Factors in the test or assessment itself**
 - **Functioning of tasks and teaching procedures**
 - **Factors in administration and scoring**
 - **Factors in student responses**
 - **Nature of the group and the criterion**

Face Validity

- **Not technically a facet of validity, but refers to a test “appearing” to measure what it is designed to measure.**
- **Content-based evidence of validity is acquired through a systematic and technical analysis of the test content, face validity only involves the superficial appearance of a test.**
- **Often desirable as it makes the test more acceptable to those taking the test and the general public.**