A N O V E R I E W
VALIDITY FOR TEACHERS
AN OVERVIEW
DEFINITION

- Validity refers to how appropriate the interpretations of a test score are for the purpose intended.
- For instance, does a test that is supposed to measure intelligence actually measure intelligence? If it does, the interpretation of the test is valid.
- Generally speaking, tests themselves are not referred to as valid or not; only the interpretation/use of scores. For instance, using a test administered in English to determine academic achievement of a Spanish-speaking first-year immigrant.
THREATS TO VALIDITY

- Construct underrepresentation = When a test measures less than the construct it’s supposed to measure
- Construct-irrelevant variance = When a test measures more than the construct it’s supposed to measure
- External factors
  - Instructional procedures
  - Test administration and scoring procedures
  - Student characteristics
RELIABILITY AND VALIDITY

• Reliability and validity are closely related measures
• A test that does not produce reliable scores cannot produce valid interpretations
• However, reliability does not guarantee validity
TYPES OF VALIDITY

• Content validity = How well the test samples the content area of the identified construct (experts may help determine this)

• Criterion-related validity = Involves the relationships between the test and the external variables that are thought to be direct measures of the construct (e.g., a departmental test is considered to have criterion validity if it is correlated with the standardized test in that subject and grade)

• Construct validity = Involves an integration of evidence that relates to the meaning or interpretation of test scores (e.g., establishing that a test of “attitude toward science careers” yields a single score that is meaningful)
In the 1970s and 1980s, researchers began describing validity as a unitary concept; as a result, the focus shifted from describing different “types” of validity to looking at different ways to collect evidence to support validity.
EVIDENCE BASED ON TEST CONTENT

• Definition= Evidence derived from an analysis of test content
• Content= Themes, wording, and format of a test
• "Does the test cover the content it’s supposed to cover?"
• The development of a test blueprint, or table of specifications, helps to guide test creation in relation to content
• After the test is written, test writers look at item relevance and content coverage to ensure the individual items and overall content cover the relevant domain
FACE VALIDITY

- Technically not a form of validity
- Refers to a test “appearing” to measure what it is meant to measure
- Tests with good face validity are better received by the general public, but may not hold up under close scrutiny from experts
- In general, face validity is desirable in order to persuade the general public to view a test as meaningful
EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

• Definition= Evidence based on an examination of the relationships between test performance and external variables

• There are two distinct application of this approach:
  • Examination of text-criterion evidence
  • Examination of other convergent and divergent evidence
EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

• Test-criterion evidence = When a test can predict performance on a given criterion
• Criterion = A measure of some attribute or outcome that is of primary interest
• Researchers use a correlation coefficient to determine the relationship between a predictor and criterion, which in this case is referred to as the validity coefficient
EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

• There are two main types of studies to collect test-criterion evidence: predictive and concurrent
  • In predictive studies the test is administered, time passes, and the criterion is measured (e.g., career aptitude survey administered, followed by career choices in the future)
  • In concurrent studies the test is administered and the criterion is measured at about the same time (e.g., correlation between career aptitude scores and career choices as indicated in a survey)
• Although neither study is objectively “better,” predictive studies may be preferable from a technical perspective, while concurrent ones may save time and money
EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

• Validity coefficient = The correlation coefficient when referring to the relationship between a predictor (ex. SAT) and criterion (ex. success in college)
• It is important that the criterion itself should be reliable and valid, or the best existing measure of the construct of interest
• Criterion contamination occurs when predictor scores influence criterion scores
EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

• Interpreting Validity Coefficients
  • The validity coefficient should be large enough to indicate that the information from the test will help predict how students will perform on the criterion measure.
  • We can use a linear regression to estimate this, however, calculating a standard error of estimate is necessary due to the fact that a test’s results cannot have perfect validity.
EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

- Factors like a **selection ratio** (the proportion of applicants needed to fill positions) and **base rate** (proportion of applicants who can be successful on the criterion) can influence the usefulness of test scores.

- Decision-theory models= Models designed to help the examiner decide how much information a test result can contribute when making classification decisions.
EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

• Convergent and discriminant evidence = Obtained when you correlate a test with existing tests that measure similar constructs.

• Discriminant evidence = Obtained when you correlate a test with existing tests that measure dissimilar constructs.

• Multitrait-multimethod matrix = An approach that requires an examination of two or more traits using two or more measurement methods.

• Method variance = When two measures show an unexpected similarity in their method of measurement.
EVIDENCE BASED ON RELATIONS TO OTHER VARIABLES

• Contrasted group studies= When validity evidence is gained by examining two different groups who are expected to differ on the construct the test is designed to measure

• Each group is administered the test, and validity would be supported if each group differed in performance as predicted
EVIDENCE BASED ON INTERNAL STRUCTURE

• Definition = Evidence regarding relationships among test items and components
• Tests may measure single or multiple dimensions
• Factor analysis = Sophisticated statistical procedure used to determine the number of distinct factors that underlie a test
  • Used to confirm or refute that the structure of a test is consistent with that of the construct
EVIDENCE BASED ON RESPONSE PROCESS

• Definition: Evidence based on an analysis of the processes the examiner and test-taker engage in.

• Recording evidence on the processes test-takers are engaging in can be obtained through interviews, observation, or analyzing errors.
EVIDENCE BASED ON CONSEQUENCES OF TESTING

- Definition: Evidence based on an examination of the intended and unintended consequences of testing
- In many cases, tests are administered on the assumption that the use of the results will have some benefit
- This line of validity evidence asks, “are these benefits being achieved?”
- Most applicable to tests designed for selection and promotion
- A special note: If the consequence of testing is that it results in groups of individuals not being placed in advanced classes/certain career tracks, placement could be deemed an invalid purpose for the test scores
INTEGRATING EVIDENCE OF VALIDITY

• Validity argument= Validation can be viewed as a developing scientifically sound validity argument to support the intended interpretation of test scores and their relevance to proposed use.
PRACTICAL STRATEGIES FOR TEACHERS TO PROMOTE VALIDITY

• Examine the test content; does it cover the content it is supposed to cover? Frequently tests that are departmental may not match with content coverage. In such cases it is not fair to use student scores in grades (you might need to remove scores based on the content not covered, and re-test for it at a later date)
• Examine the types of cognitive processes the test asks the students to engage in; are they ones specified in the learning objectives? If you have taught for “understanding” then, it is inappropriate to ask for “evaluation”
• Examine other results of student assessment; are they consistent with the students’ results on the current assessment?
• Have you made sure that the test is reliable? That is, it is free of spelling errors, difficult vocabulary, etc.?
• Have you made sure that the test is fair to students of all backgrounds?
• Is the assessment practical and efficient?
• Consider having multiple assessments instead of just one to make a decision about the student’s learning