

Language test validity

- A. Defining validity
- B. Types of validity
- C. Factors affecting validity
- D. Validity and washback
- E. Reliability and validity

Defining validity

- “A test is said to be valid if it measures accurately what it is intended to measure” (Hughes, 2003, p. 26).
- “The validity of a test is the extent to which it measures what it is supposed to measure and nothing else” (Heaton, 1988, p. 159).
- “If the test is found to be based upon a sound analysis of the skill or skills we wish to measure [what precisely does the test measure?], and if there is sufficient evidence that test scores correlate fairly highly with actual ability in the skills area being tested [how well does the test measure?], then we may feel reasonably safe in assuming that the test is valid for our purposes” (Harris, 1969, p. 19)

- “Test validity [is] the degree to which a test measures what it claims , or purports, to be measuring” (Brown, 2005, p. 220).
- “Validity in general refers to the appropriateness of a given test or any of its component parts as a measure of what it is purported to measure. A test is said to be valid to the extent that it measures what it is supposed to measure” (Henning, 1988, p. 89)
- “Validity is the adequacy and appropriateness of the interpretations and uses of assessment results” (Miller, Linn, & Gronlund, 2010, p. 70).

Points to keep in mind about validity

- Validity is a property of a test.
- A test is valid for some purposes, but not for others.
- A test must be appropriate in terms of objectives we have set.
- Validity refers to the inferences made from test scores.
- Validity is as intuitive as it is scientific.

Types of validity

- I. Nonempirical validity: no collection of data or use of formulae and no involvement of coefficient or mathematical computations
 - A. Content validity
 - B. Face validity
 - C. Response validity
- II. Empirical validity: collection of data and involvement of mathematical computations
 - A. Criterion-related or statistical validity
 - 1. Concurrent or status validity
 - 2. Predictive validity
 - B. Construct validity

Content validity

- A. The correspondence between test content and the contents of materials to be tested
- B. Strategies to establish content validity
 - 1. Defining a domain
 - 2. Drawing a table of specifications
 - a. Representativeness
 - b. Appropriateness of ability
- C. Having experts comment on the test content
- D. Subjective nature of test validity

Face validity

- A. Definition: The degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers.
- B. Features:
 1. Superficial features of a test—the way the test looks or the appearance of a test
 2. The appeal of the test to the lay judgment
 3. Content of a test irrespective of its correspondence to the to-be-tested materials and level of testees
 4. Motivational nature of face validity

- C. High face validity in the eyes of learners
 - 1. A well-constructed, expected format with familiar tasks
 - 2. A clearly doable test within allotted time limit
 - 3. Clear and uncomplicated items
 - 4. Crystal clear directions
 - 5. Course-related tasks
 - 6. Reasonably challenging test

Response validity

- A. The correspondence between testees's responses and test developers' expectations
 - 1. Haphazard or nonreflective responses violate response validity
 - 2. Unclear instructions and unfamiliar test formats violate response validity.

Concurrent validity

- A. The simultaneous administration of the newly made test and an already well-established validated test and estimation of the correlation coefficient (known as validity coefficient) between the scores obtained from the two tests
- B. TOLIMO and MCHE as general proficiency tests in Iran

Predictive validity

- A. The administration of the newly made test and an already well-established validated test and estimation of the correlation coefficient (known as validity coefficient) between the scores obtained from the two tests
- B. Many tests administered in Iran before MA exam

Construct validity

- A. Psychological construct: an attribute, proficiency, ability, or skill defined in psychological terms
- B. Examples of language constructs are intelligence, language aptitude, overall general proficiency
- C. Constructs are abstract and unobservable
- D. Constructs should be demonstrated indirectly through experiments

Factors affecting validity

- I. Directions
 - A. The way the test items to be marked
 - B. Permission to guess
 - C. Freedom to ask questions
- II. Difficulty level of the test—neither too easy nor too difficult
- III. Structure of the items—poorly constructed or ambiguous items
- IV. Arrangement of items and correct responses

- V. Invalid application of tests
- VI. Inappropriate selection of content—lack of representativeness and mismatch between test content and objectives
- VII. Imperfect cooperation of the examinees—lack of response validity
- VIII. Poor criterion selection
- IX. Sample truncation—homogeneous groups

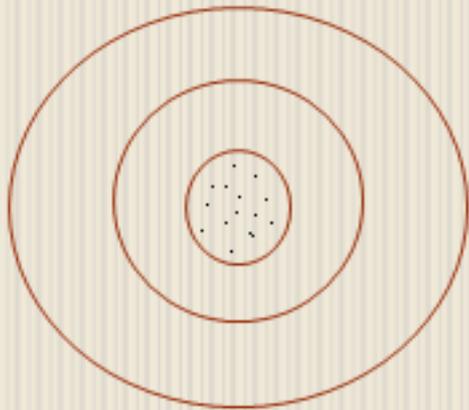
Reliability and validity

- I. Reliability is testee dependent, but validity is context dependent.
- II. Reliability is a purely statistical concept, but validity needs expert judgment.
- III. Reliability is a necessary, but not sufficient condition for a test.
- IV. Validity is more important than reliability.

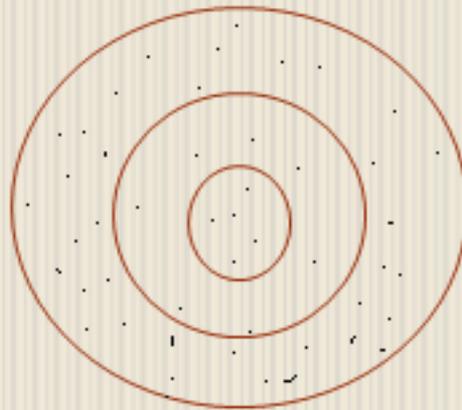
How high should reliability and validity be?

- I. Purpose of the test
- II. Importance of decisions
- III. 0.50 = low; between 0.50 and 0.75 = moderate; 0.75 or above = high

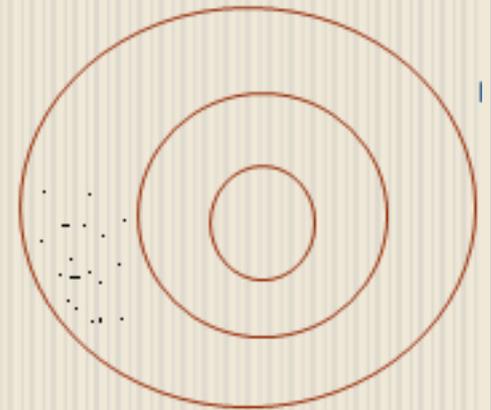
- I. Target-shooting
- II. Sweet-buying
- III. Time-showing



Reliable and
valid shooting



Unreliable and
invalid shooting



Reliable but
invalid shooting