



Validity & Reliability

CH.8




Reliability: *How consistent is the score?*

- Inter-rater reliability
 - When another person scores it, is the score consistent?
- Test-retest reliability
 - When you take it again, is the score consistent?

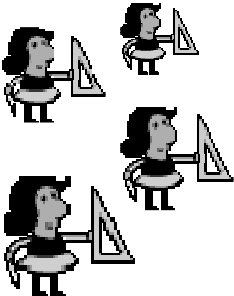



Old Faithful





1. Inter-rater reliability

- Make groups of three
 - Two of you measure the other one's head & write the number down.
 - Compare the numbers to each other
 - Are they similar?
 - Are they reliable?
 - a correlation of at least .90 among scorers as an acceptable level of agreement


2. Test-retest reliability

- Measure the same person's head again.
- Is your number similar to your number last time?
- Is each measurement reliable?
- In Educational Research
 - scores collected over a two-month period: sufficient evidence of test-retest reliability


3. Equivalent-Forms Reliability

- Two different but equivalent (alternate or parallel) forms of an instrument are administered to the same group during the same time period.





4. Estimates of Internal Consistency Reliability

- Inter-item
 - The degree of consistency among the items in a scale
- Item-total
- Split-half
 - Randomly divide items into 2 subsets & examine the consistency in total scores across the 2 subsets
- Cronbach's Alpha
 - Average consistency across all possible split-half reliabilities (Conceptually)



Validity

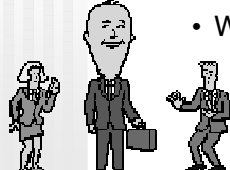

- Measure your intelligence
- Get into groups of two
 - Measure your partner's head

Validity:


Does it measure what it intends to measure?

- Does your head size measure your intelligence?
- Who is the smartest?





1. Content Validity

- Questions
 - Does the test measure the concepts taught during instruction?
 - How well do the test items match the stated objectives?




A panel of social studies **experts** help assess content validity evidence of a national test.




Content Validity with Non-experts: Face Validity

- A measure has face validity
 - if people think it does
 - e.g., *it looks okay.*




- Measures with face validity:
 - Can be invalid
- Measures with lacking face validity:
 - Can be valid




2. Construct Validity

- Measuring hypothetical constructs
 - Assessed by the relationships between the measure of a construct & scores on measures of other constructs
- To have construct validity, a measure should both:
 - Correlate with other measures that it should be related to (**convergent validity**)
 - Not correlate with measures that it should not correlate with (**discriminant validity**)





3. Criterion validity

- Does the test predict to another measure?
 - Presently: **Concurrent validity**
 - How well are the SAT scores related to the ACT scores?
 - Future: **Predictive validity**
 - How well does your high school GPA predict your college GPA?
 - How well does your SAT score predict your college GPA?




Reliability & Validity

- Questions
 - Is your head size reliable?
 - Is it valid for measuring your intelligence?


Reliability & Validity ---Continued

- Discussions
 - Can a test be reliable & not valid?
 - Can a test be valid & not reliable?




Sampling Error/Measurement Error

- **Sampling error** ↔ **External Validity**
 - Reduced by
 - Representative sampling
 - Generalizable sampling
- **Measurement error** ↔ **Reliability**
 - Reduced by
 - Writing items clearly
 - Making instructions easily understood
 - Adhering to proper test administration
 - Providing consistent scoring
 - **Should use highly reliable tests**
 - .80 < reliability coefficient
 - Reliability coefficient of .80= 20% of the variability in test scores is due to measurement error



Factors Influencing Reliability


- 1. Factors in the test itself
 - Unclear directions
 - Reading vocabulary & sentence structure too difficult
 - Ambiguity
 - Inadequate time limits
 - Test items inappropriate for the outcomes being measured
 - Poorly constructed test items
 - Test too short
 - Improper arrangement of items
 - Identifiable pattern of answers
- 2. Factors in administration
 - No standard directions & time limits
 - Unfair aid
 - Cheating



Factors Influencing Reliability ...

Continued


- 3. Factors in scoring & Grading
 - Unreliable scoring or error in scoring
 - Bias in scoring or grading
 - Fatigue while grading
 - Adverse physical & psychological conditions
- 4. Factors in student responses
 - Emotional disturbances
 - Frightened by the assessment situation
 - Not motivated to put their best effort
- 5. Nature of the group & the criterion
 - Age, gender, ability level, educational background, & cultural background



Selection of a Test:

Sources of test information


- 1. Mental Measurement Yearbooks (MMY)
 - Reviews in MMY are most easily accessed through your university library & the services to which they subscribe (e.g., EBSCO)
 - Provides factual information on all known tests
 - Provides objective test reviews
 - Comprehensive bibliography for specific tests
 - Indices: titles, acronyms, subject, publishers, developers
 - Buros Institute



*Selection of a Test:
Sources of test information ---Continued*

2. Tests in Print

- Tests in Print is a subsidiary of the Buros Institute
- Bibliography of all known commercially produced tests currently available
- Very useful to determine availability
- Tests in Print




*Selection of a Test:
Sources of test information ---Continued*

3. ETS Test Collection

- Published & unpublished tests
- Includes test title, author, publication date, target population, publisher, & description of purpose
- Annotated bibliographies on achievement, aptitude, attitude & interests, personality, sensory motor, special populations, vocational/occupational, & miscellaneous
- ETS Test Collection


4. Professional journals

5. Test publishers & distributors



Issues to consider when selecting tests

- Psychometric properties
 - Reliability
 - Validity
 - Length of test
 - Scoring & score interpretation Issues
- Non-psychometric issues
 - Cost
 - Administrative time
 - Objections to content by parents or others



Chapter 8 Quiz

- 1. Why do we need to be concerned about the notions of reliability & validity in educational research? Answer your question with regard to the *instruments* (broadly defined) we use to collect data.
- 2. What type of validity--content-related, criterion-related, or construct-related—do you think is the easiest to obtain? The hardest? Why?
- 3. What is measurement error, how is it different from sampling error, & what can be done to minimize them (measurement errors & sampling errors)?
- 4. The textbook points out that scores from an instrument may be reliable but not valid, yet not the reverse. Why would this be so?