

Validity & Reliability in Language Assessment Con't

Classical Test Theory

- Every score consists of a part we want (true score variance) and a part we don't want (error variance).
- We want to MAXIMIZE true score variance and MINIMIZE error variance
 - True score variance
 - Reflects a students' true ability
 - To maximize this - clear directions, compute reliability
 - Error variance
 - Stress, anxiety that lower a students' scores
 - The error variance works against the students
 - Poor instructions

Gathering Data

- Test-retest: Administer a test. Wait a little time and then re-administer the test.... 1 test administered 2 times
- Alternate Forms reliability: Two tests with form A and form B given to the same group of students, one time.... 2 tests, administered 1 time
- Split-half: 1 test administered 1 time
 - Used for **objectively scored** items

Spearman-Brown Prophecy Formula

- Makes a prophecy about the reliability of the whole test based on the reliability of the split halves.
 - Used for **objectively scored** items

Standard Error of the Measurement (SEM)

- Reliability in this formula can be from KR-21, Spearman-Brown Prophecy reliability etc etc
- You want the SEM to be as little as possible based on the length of the test

There can be no validity without reliability

Trait = construct

Method = item format. How we go about testing the things we are testing

Multi trait, multi method = at least 2 traits are measured by at least 2 methods