

## Reliability & Dependability in Language Assessment

Internal-consistency reliability: estimate the consistency of a test using only information internal to a test, that is available in ONE administration of a single test

### Split-half method

- Easiest to understand conceptually
- Test splits odd and even numbered items, calculates a correlation coefficient for the two sets of scores
- The coefficient gives reliability for either the odd or even numbered items
- To calculate the **full-test** reliability is accomplished using **Spearman-Brown prophecy formula**

### Cronbach Alpha / Coefficient Alpha

- Easy to calculate and offers similar reliability as the split-half reliability
- Can use it for objectively scored items and inter-rater reliability
- Based on the concept of variance NOT coefficient
  - Variance = difference in scores

### Kuder-Richardson Formulas

- K-R20
  - Avoids the problem of underestimating the reliability. It is the most accurate and flexible of estimates
  - Difficult to calculate
  - Can only be applied when items are scored correct/incorrect
  - If accuracy is the main concern then this is the formula to use
- K-R21
  - Relatively easy to use
  - Tester only needs to know the number of items, the mean and standard deviation on the test
  - Tester doesn't have to calculate a correlation coefficient
  - Produces reliability results that are lower than Cronbach alpha and split half method.
  - Provides a conservative estimate of reliability
  - For a quick, rough estimate

### Inter-rater reliability

- Must calculate a correlation coefficient between the two sets of scores from the two raters
- Calculates how closely the different rater's ratings correlate

### Intra-rater reliability

- Two sets of scores produced by the same rater for the same group of students and calculating a coefficient between those two sets of scores

## 3 Most Commonly Used Statistics for Calculating Reliability

### Pearson's R

- Most commonly used and most powerful
- Independent and dependent variables are measured on an interval scale (points on a scale is equidistant)
- Works better with numbers over 30
- Does not like restricted ranges (there isn't a distribution of scores)
- Variables must be normally distributed
- Each observation is independent than the others

### Spearman's Rank Order Correlation Coefficient

- Second most frequently used

- Uses rank-**ordered** data or data from an **ordinal** scale
- Tells strengths of relationship of rank orderings on X and Y
- Represented by symbols =  $r_{ho}$  or  $R_s$

**Point Biserial Correlation Coefficient**

- Least used
- One variable is continuous and the other is dichotomous
  - Dichotomous = Two. The comparison ranked side by side
- Important in item analysis in language testing procedures
- Used for calculating item discriminability (high, low scores)
- Flannagan's shortcut is a way to calculate the item discrimination
- Symbol =  $R_{pbi}$