Language test reliability

- On defining reliability
- Sources of unreliability
- Methods of estimating reliability
- Standard error of measurement
- Factors affecting reliability

On defining reliability

- Non-technical definitions of reliability

A. Dictionary definition: something that is correct or true or someone who can be trusted

1. The incident cast doubt on her motives and reliability.
2. The reliability of these results has been questioned.

B. Stability or consistency of results or test scores

Technical definition of reliability

The ratio of true score variance to observed score variance

A. Observed score = true score + error score
B. Uncorrelated nature of error scores with true scores

$$r = \frac{Vx - Ve}{Vx}$$

Methods of estimating reliability

I. Some preliminaries

A. Reliability coefficient or reliability estimate

B. Range of reliability coefficient: 00.0 to 1.00

C. Correlation coefficient to estimate reliability

II. Common methods of estimating reliability

A. Test-retest method

B. Parallel-forms method

C. Internal consistency methods

1. Split-half method
2. Cranach alpha
3. Kuder-Richardson methods

Test-retest reliability

- Administration of the same test to the same testees twice
- Obtaining two sets of scores, lining them up, and calculating the correlation coefficient
- Pearson product moment correlation coefficient to estimate test-retest reliability
- Conservative estimate of reliability

$$r_{xy} = \frac{\Sigma(X - Mx)(Y - My)}{NSxSy}$$

Mx = 55.1; My = 52.5; Sx = 15.5; Sy = 16.1

$\Sigma(X - M_x)(Y - M_y) = 895.00$

$$r_{xy} = \frac{895.00}{30(15.5)(16.1)} = \frac{895.00}{7486.5} = 0.11 \approx .12$$

Shortcomings of test-retest reliability

A. Existence of two administrations
B. Learning effect
C. Practice effect

Parallel-(equivalent or alternate), forms reliability

- Definition:

1. Administration of two different versions of the same test two a single group of testees
2. Correlation coefficient to estimate the reliability of forms

- Requirements:

1. Equal mean and standard deviations of two forms
2. Correlation of two forms with a third measure
3. Length of the test
4. Item types
5. Table of specifications

Internal-consistency reliability methods

Split-half method

- Definition:

1. Dividing the same test into two parts and administering it to the same testees only once
2. Spearman-Brown prophesy formula to estimate two parts
3. Measurement of the same trait or ability of the two parts—homogeinity of items
4. Independence of the two parts
5. Importance of length

- Ways of splitting the test

1. Easy-to-difficult method
2. Odd-even method

- Adjustment for full-test reliability

$$r = \frac{2\,(\text{r half})}{1 + (\text{r half})}$$

$$r = \frac{2\,(0.95)}{1 + (0.95)}$$

$$= \frac{1.90}{1.95} = 0.97$$

Advantage and disadvantage

I. Advantages
A. Practicality
1. No twice administration of the same test
2. No two different versions of the same test

II. Disadvantages
A. Insurance of homogeneity
B. Different subsections of the same test

KUDER-RICHARDSON METHODS

I. Kuder-Richardson formula 2o

$$K - R20 = \frac{k}{k-1}\left(1 - \frac{\Sigma SDi}{SDt}\right)$$

K-R20 = Kuder-Richardson formula 20

K = number of items

ΣSDi = sum of item variances

SDt = test score variance

II. Kuder-Richardson formula 21

$$K - R21 = \frac{k}{k-1}\left(1 - \frac{M(k-M)}{kSD}\right)$$

K-R21 =Kuder-Richardson formula 21

K = number of items

M = mean of test scores

SD = standard deviation of test scores

$$K-R21=\frac{k}{k-1}\left(1-\frac{M(k-M)}{kSD}\right)$$

$$K-R21=\frac{60}{60-1}\left(1-\frac{48(60-48)}{60(12.96)}\right)$$

$$=K-R21=\frac{60}{59}\left(1-\frac{576}{777.6}\right)$$

$$=K-R21=1.0169492\ (1-\ 0.7407407)$$

$$=K-R21=1.0169492\ x\ 0.2592593$$

$$=0.2636535$$

$$K - R20 = \frac{k}{k-1}\left(1 - \frac{\Sigma SDi}{SDt}\right)$$

$$K - R21 = \frac{60}{60-1}\left(1 - \frac{1.55}{12.96}\right)$$

$$= K - R21 = \frac{60}{59}\left(1 - 0.1195988\right)$$

$$= K - R21 = 1.0169492 \times 0.8804012$$

$$= 0.8953233$$

Advantages, assumptions and differences

I. Kuder-Richardson formula 21

A. No administration of the same test twice

B. Lack of two different versions of the same test

C. No separate scoring of odd and even numbered items

D. No correlation coefficient calculation

E. No adjustment for length

II. Assumptions

A. Equality of items

B. Independence of items scored

C. Measurement of the same trait

III. Differences

A. K-R21 is simpler to calculate and more common in language testing than K-R20

B. K-R21 is more conservative than K-R20, yielding a lower reliability coefficient

Which method of estimation to choose

Criteria to choose

I. Frequency of appearance—internal consistency methods
II. Function of the method
III. Conceptual clarity—split-half method
IV. Ease of calculation—K-R21
V. Accuracy of results—K-R20 and split-half method
VI. Weighting of items in a test—Cronbach alpha