

# **The Ways of Assessing Reliability in a Psychological Test**

**Ven. Karandagolle Wijithatissa**

## **Introduction**

Reliability is one of the important characteristics of any psychological test. In its simplest sense reliability refers to precision, or accuracy of the measurement or score. A well made scientific instrument should yield accurate result both at present as well as over time. Reliability refers to consistency of scores of measurement which is reflected in the reproducibility of the scores. Reliable test is one that produces identical or at least highly similar results for an examinee from one occasion to other. A test is said to be consistent over a given period of time when all the examinees retain their

same relative ranks of two separate tastings with the same test. A test also said to be consistent if the examinees who obtain high scores on one set of items also score high on an equivalent set of items and those who obtain a low scores on one set of items also score low on an equivalent set of items.

### Types of Reliability

There are four most common methods of estimating the reliability coefficient of test scores. They are as follows;

01. Test –Retest Reliability
02. Internal Consistency Reliability (Split Half Reliability)
03. Parallel Forms Reliability (Alternative Forms Reliability)
04. Scorer Reliability

### Test – Retest Reliability

In the test –retest reliability the single form of the test is administered twice on the same sample with a reasonable time gap. In this way, two administrations of the same test yield two independent sets of scores. The two sets when correlated give the value of the reliability coefficient. In computing test – retest reliability the investigator often faced with the problem of determining reasonable time gap between two administrations of the test. When the time is too short it is likely to increase reliability coefficient due to the carryover and practice effect. If the time gap on the other hand is too long it is likely to lower the reliability coefficient. The most appropriate and convenient time gap between the two administrations is a fortnight. The test – retest method is the most appropriate method of estimating reliability of both speed test, power test and heterogeneous tests.

#### Disadvantages

Time consuming method.

This method assumes that the examinee's physical and psychological set up remains unchanged in both the testing situations. But in reality this is not so.

Some uncontrolled environmental changes may take place during the administration of the test.

### **Internal Consistency Reliability (Split Half Reliability)**

Internal consistency reliability indicates the homogeneity of the test. If all the items of the test measure the same function or trait, the test is said to be a homogeneous one and its internal consistency reliability would be pretty high. The most common method of estimating internal consistency reliability is the split-half method in which the test is divided into two equal or nearly equal halves. The common way of splitting the test is the odd-even method. In this method all odd numbered items constitute one part of the test and all even numbered items constitute another part of the test. Each examinee thus receives two scores. Product moment ( PM) correlation is computed to obtain the

reliability of the half test. On the basis of the half test reliability the reliability for the whole test is estimated. PM correlation is computed between two sets of scores. When the reliability coefficient of the half test is known, the Spearman – Brown Prophecy is used for estimating the reliability of the whole test.

#### Advantages

All data necessary for the computation of the reliability coefficient are obtained in a single administration of the test.

#### Disadvantages

Fluctuations due to changes in the temporary condition within the examinee as well as due to the temporary changes in the external environment will operate in one direction. That is either favorably or unfavorably.

This should not be used with the speed test.

Having different methods to divide the test into two halves and each method yields a different coefficient of reliability.

## Other Formulas for Computing Consistency Reliability

Rulon and Flanagan Formulas

Kuder- Richardson Formulas and Coefficient Alpha

### Rulon and Flanagan Formulas

Both these formulas provide the reliability of the whole test and not the half test. Both formulas estimate the reliability coefficient on the basis of the proportion of error variance in total variance of the test.

### Kuder - Richardson Formulas and Coefficient Alpha

Popular and well known K-R<sub>20</sub> is the basic formula for computing the reliability coefficient and K-R<sub>21</sub> is the modified form of K-R<sub>20</sub>.

### Coefficient Alpha

The tests which have multiple scored items such as sometimes, Usually, Rarely, Never, for calculating

reliability of such test a generalized formula is known as Coefficient Alpha.

### Parallel Forms Reliability (Alternative - Forms Reliability)

Alternative reliability requires that the test be developed in two forms which should be comparable or equivalent. Two forms of the test are administered to the same sample with the immediately the same day or with the time interval of usually a fortnight. Pearson  $r$  between two sets of scores obtained from two equivalent forms becomes the measure of reliability. Such a coefficient is known as coefficient of equivalence.

#### Disadvantages

Difficulties in making the two forms of test parallel.

It requires much labour and time.

### Scorer Reliability

Scorer reliability is the reliability which can be estimated by having a sample of test independently scored by two or more examiners. The two sets of

scores obtained by each examiner are completed in the usual way and the resulting correlation coefficient is known as scorer reliability.

### Factors Influencing Reliability of the Test Scores

A number of factors that is influencing on the reliability of test scores can be categorized into two such as extrinsic and intrinsic. Extrinsic factors are those factors which lie the outside the test and tend to make the test reliable or unreliable. For example group variability, guessing by the examinees, momentary fluctuations and environmental conditions. Intrinsic factors on the other hand refer to those factors which lie within the test itself and influence the reliability of the test; for example characteristics of items, total score, and length of the test.

### The Way of Improving the Reliability of Test Scores

Reliability of the test scores can be improved by controlling those factors which adversely affect the reliability of the test. For example the group of the examinees should be heterogeneous, items should be

homogeneous, test should preferably be a longer one, and items should be moderate difficulty value and item should be discriminatory one.

### References

- Chadha, N.K. (2009), Applied Psychometry, Sage Publication, India.
- Garfield, J.B. and Gal, I. (1999), Assessment and Statistics Education: Current Challenges and Directions, journal of International Statistical Review, Vol. 67, No. 1, pp. 1-12
- Hohmann, U (2005), Quantitative Methods in Education Research, Retrieved 15<sup>th</sup> April 2013 from <http://www.edu.plymouth.ac.uk/resined/Quantitative/quanthtm.c.htm>
- Kaplan, R.M. & Saccuzzo (2009), Psychological Testing, Wordsworth, U.S.A.
- Moody, (2002) Empirical Research Methods, Sage Publication, India.
- Murphy, J.J. (1972) Current Practices in the use of Psychological Testing by Police Agencies, Journal of Criminal Law, Criminology, and Police Science, Vol. 63, No. 4, pp. 570-576
- Pratt, N (2006), Qualitative research, Retrieved 15<sup>th</sup> April 2013, from

<http://www.edu.plymouth.ac.uk/resined/Qualitative%20methods%20202/qualrshm.htm>

Sparrow, S.S & Davis, S.M, ( 2000), Recent Advances in the Assessment of Intelligence and Cognition, Journal of child Psychology, Vol 41, No 1, pp 117-131

Urbina.S., (2004), Essentials of Psychological Testing, John Wiley & sons, New York.