RELIABILITY OF THE ASSESSMENT TOOLS

Reliability

What does the term reliability mean? Reliability means Trustworthy. A test score is called reliable when we have reasons for believing the test score to be stable and objective. For example if the same test is given to two classes and is marked by different teachers even then it produced the similar results, it may be considered as reliable. Stability and trustworthiness depends upon the degree to which score is free of chance error. We must first build a conceptual bridge between the question asked by the individual (i.e. are my scores reliable?) and how reliability is measured scientifically. This bridge is not as simple as it may first appear. When a person thinks of reliability, many things may come into his mind – my friend is very reliable, my car is very reliable, my internet bill-paying process is very reliable, my client's performance is very reliable, and so on. The characteristics being addressed are the concepts such as consistency, dependability, predictability, variability etc. Note that implicit, reliability statements, is the behaviour, machine performance, data processes, and work performance may sometimes not reliable. The question is "how much the scores of tests vary over different observations?"

Some Definitions of Reliability:

According to Merriam Webster Dictionary:

"Reliability is the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials."

According to Hopkins & Antes (2000):

"Reliability is the consistency of observations yielded over repeated recordings either for one subject or a set of subjects."

Types of Reliability

Reliability is one of the most important elements of test quality. It has to do with the consistency, or reproducibility, of an examinee's performance in the test. It's not possible to calculate reliability exactly. Instead, we have to estimate reliability, and this is always an imperfect attempt. Here, we introduce the major reliability estimators and talk about their strengths and weaknesses. There are six general classes of reliability estimates, each of which estimates reliability in a different way. They are:

i)   Inter-Rater or Inter-Observer

Reliability To assess the degree to which different raters/observers give consistent estimates of the same phenomenon. That is if two teachers mark same test and the results are similar, so it indicates the inter-rater or inter-observer reliability.

ii)   Test-Retest Reliability

To assess the consistency of a measure from one time to another, when a same test is administered twice and the results of both administrations are similar, this constitutes the test-retest reliability. Students may remember and may be mature after the first administration creates a problem for test-retest reliability.

iii)   Parallel-Form Reliability

To assess the consistency of the results of two tests constructed in the same way from the same content domain. Here the test designer tries to develop two tests of the similar kinds and after administration the results are similar then it will indicate the parallel form reliability.

iv)   Internal Consistency Reliability

In internal consistency reliability estimation, we use our single test. The test is administered to a group of students on one occasion to estimate reliability. In effect we judge the reliability of the instrument by estimating how well the items that reflect the same content give similar results. We are looking at how consistent the results are for different items for the same construct within the measure. There are a wide variety of internal consistency measures that can be used.

v)   Split half Reliability

To assess the consistency of results comparing two halves of single test, these halves may be even odd items on the single test. Suppose you have to develop a test of 30 items and you want to know that how reliable the test is? What you have to do is to administer the test, mark it and

divide it in to two parts, in such a way that place all the even numbered items (2,4,6…………) in one half and the odd numbered items (1,3,5…………..) in the second. Calculate the reliability by using the Spearman-Brown prophecy formula given below. Actually in split-half reliability we randomly divide all items that claim to measure the same contents into two sets. We administer the entire instrument to a sample of students and calculate the total score for each randomly divided half. The split-half reliability estimate is simply the correlation between these two total scores.