

Reliability and Validity of Measurement

Basuki Nath Jha (B.Ed)

Sem– 2, TC- 204

Godda College Godda

Measurement involves assigning scores to individuals so that they represent some characteristic of the individuals. But to know that the scores actually represent the characteristics the researchers conduct research using the measure to confirm that the scores make sense based on their understanding of the construct being measured. This is an extremely important point. Psychologists do not simply **assume** that their measures work. Instead, they collect data to **demonstrate** that they work. If their research does not demonstrate that a measure works, they stop using it. In evaluating a measurement method, psychologists consider two general dimensions: **reliability and validity**.

Reliability

In the research, reliability is the degree to which the results of the research are consistent and repeatable. Researchers repeat research again and again in different settings to compare the reliability of the research. Theories are developed from the research inferences when it proves to be highly reliable.

In experiments, the question of reliability can be overcome by repeating the experiments again and again. In social sciences, the researcher uses logic to achieve more reliable results. However, in social sciences it is difficult to achieve reliability in the data collection, because, human behaviours are difficult to repeat even in similar situations. There are several external factors that influence the human behaviours and it is very important to know the effect of any such factor other than the independent variable.

Types of reliability

1. Test-retest reliability

Test-retest reliability measures the reliability of the measure over a period of time. In social sciences, a test is administered more than one time over a period of time to check or retest its reliability. In natural sciences, the researcher conducts experiment more than one time to ascertain its reliability. The results of the tests and the inferences drawn have to be applied to natural settings, they should be reliable. This method of testing the reliability of the test is time-consuming, since the researcher has to wait for some time to re-administer the test.

2. Parallel forms reliability

It measures the reliability of the test by administering it in two different forms. Both forms of the test measure the same variables under study, but the format of the measure is different. The researcher must be able to formulate two different tests that measure the same variables. The difficulty arises in formulating two tests that are similar in nature and measurement level. The researcher may also find it difficult to administer it to two similar populations. In social sciences, using parallel forms of the same test is difficult and subjectivity is highly involved.

3. Inter-rater reliability

Inter-rater reliability check is used to measure the test by more than one rater or judge. The researcher asks more than one people to rate the reliability of his test. This type of reliability test is useful for subjective measures where more than one rater can best describe the reliability of the test.

4. Internal consistency reliability

The idea behind internal consistency reliability is that items measuring the same phenomenon should produce similar results. This means that the items, that test the attitude or behaviour, are divided into half. Each half is tested separately and then their

scores are correlated. For example, the researcher has developed questionnaire to test the attitude of people towards any state program. The researcher might divide the questions on the questionnaire in half and administer both questionnaires separately. The resultant scores of the test will be correlated to know the internal consistency reliability.

Interpretation of reliability

The reliability of a test is indicated by the reliability coefficient. It is denoted by the letter "r," and is expressed as a number ranging between 0 and 1.00, with $r = 0$ indicating no reliability, and $r = 1.00$ indicating perfect reliability. Generally, we will see the reliability of a test as a decimal, for example, $r = .80$ or $r = .93$. The larger the reliability coefficient, the more repeatable or reliable the test scores. However, selection or rejection of a test not solely based on the size of its reliability coefficient. To evaluate a test's reliability, we should consider the type of test, the type of reliability estimate reported, and the context in which the test will be used.

Validity

Validity refers to the accurateness of the research as a whole and the accuracy of each step independently. It is the highest aim every researcher wants to achieve. When we measure what we have intended to measure we reach a conclusion that is valid and verifiable.

According to **Kerlinger**, 'the commonest definition of the validity is epitomized by the question: Are we measuring what we think we are measuring'. The first step to achieving validity in the research is to develop research objectives that really target the research questions that a researcher has formulated. To establish the validity in research, the researcher should use logic and statistical evidence. The concept of validity was formulated by **Kelly (1927)** who stated that a test is valid if it measures what it claims to measure.

Types of validity

Face validity

Face validity as the name suggests shows the face-off value of the research or the measures used in the research. Face validity is not an authentic way to check the validity of the research. It is a general validity measure for the common people. Face validity can be tested by people who are taking the test because they can better decide whether the measure is appropriate or not. The researcher can also ask some experts in his field to check the measure and its validity.

Construct validity

To check the validity of the construct a panel of experts are hired. They check whether the construct measures what it needs to measure. The measure should measure what it is intended for and not some external factor.

Criterion-related validity

Sometimes the instrument is developed to observe some criteria. The validity of the criteria can be judged by comparing it with another future assessment, if the future assessment proves to be successful it shows that the criteria or the test devised to test a behaviour was valid and should be used again.

Interpretation of Validity

In evaluating validity information, it is important to determine whether the test can be used in the specific way intended, and whether the target group is similar to the test reference group.

The Validity of a test is measured by Validity Coefficient. It is reported as a number between 0 and 1.00 that indicates the magnitude of the relationship, " r ," between the test and a measure of job performance (criterion). The larger the validity coefficient, the more confidence you can have in predictions made from the test scores. However, a single test can

never fully predict job performance because success on the job depends on so many varied factors. Therefore, validity coefficients, unlike reliability coefficients, rarely exceed $r = .40$.

Relation of Validity to Reliability

Reliability and validity are two dimensions of the same thing, that is, test efficiency. Validity is the correlation of the test with some outside independent criteria and reliability of the test is the self-correlation of the test. A test which is not correlating with self is not expected to correlate with outside independent criteria. In other words, a test which has poor reliability is not expected to yield high validity. Thus, validity is dependent upon reliability. This is true in the case of homogeneous test but in heterogeneous test each part measures an independent function. Thus, validity may be high without the underlying high reliability.

A test constructor should not always aim at having high reliability and high validity in the same test. If he does so, he is said to be working at cross purpose because sometimes the goals of reliability and validity are incompatible. This is because the requirements for high reliability and high validity are opposite to each other. High reliability requires items of equal difficulty and high intercorrelations between the items, whereas high validity requires items of different difficulty values and low in intercorrelations among items. Obviously, attempting high validity as well as high reliability would imply working at cross purposes. In any test if inter items correlations range from 0.10 to 0.60, one can expect to have both reliability and validity to a satisfactory degree.

The End