JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

ESSENTIAL FEATURES OF A SOUND TEST

Amna Saleem

Lecturer, Department of Education, The Women University, Multan

Muhammad Ishaq

Subject Specialist, Govt. Higher Secondary School Rohillan Wali,Muzzafar Garh

Muniba khan

Ex-M.Phil. Scholar, Department of Education, The Women University, Multan

ABSTRACT:

Assessment is considered an essential component of education and learning procedure, that controls whether the objectives of schooling are accomplished or not. The learners did not completely acquire what the teachers teach to them. Therefore, the assessment is used to connect the bridge between teaching and learning. The test is an imperative part of the assessment process. The sound test measures the performance of students without setting traps for them. The purpose of this paper is to explain the characteristics that were necessary for appropriate test construction. Validity, reliability, objectivity, and usability are features of a sound test. This article put light on these features deeply.

Keywords: Test, Validity, Reliability, Objectivity, and Usability

INTRODUCTION:

Standardized assessments offer uniform methods to administrate and for scoring the instrument. At the same time, the question may be asked every time where the test is at hand, by giving specific directions that how the test may be utilized. A healthy test is used for the data collection. Among the certain tests, evaluation is mandatory to acquire appropriate data set. One can select any category of the test, considered best suitable intended for a specific purpose. The criteria of the test are the best indication of the test that how this test is providing accurate and relevant information. According to the characteristics of a respectable test to measure a sample of anticipated actions, the test must make sure, how best it measures the essential characteristics of and to what extent it endeavors to measure. Furthermore, below are the features of a sound test (Bryman, & Bell, 2003).

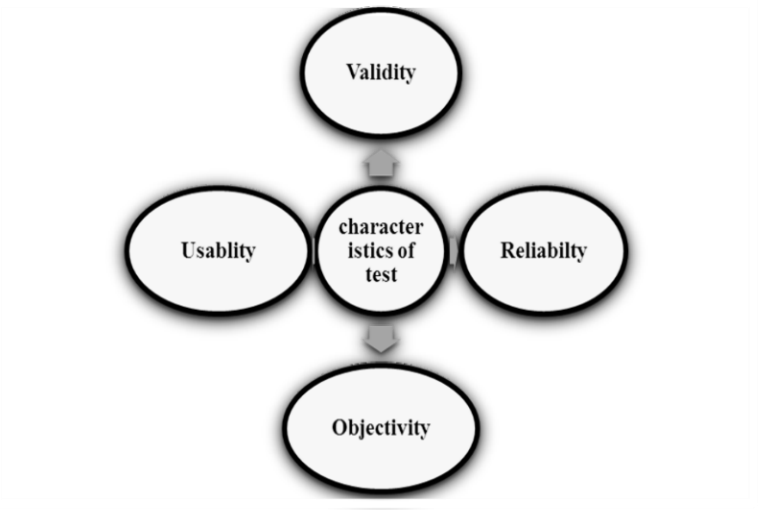


Figure No 01: Characteristics of good test

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

1. RELIABILITY OF THE TEST

Standardized tests are considered always reliable ones. One observes no uncertainties in the outcomes of the test by itself. Any data collection technique is considered reliable when it yields consistent results among the repetitive measurements taken from the same individuals under similar circumstances (Fowler, 2002).

The question of accuracy is addressed under the concept of reliability with that one quantify the 'what'. An instrument can never be considered reliable even it essentially possesses the ability to consistently produce identical conclusions along with the repetitive measures taken from the similar individual having similar characteristics. Reliable tests continuously produce firm results. In determining reliability, importance must be given to the arrangement of the test within itself. Therefore, reliability is described, a height of consistency among distinct measuring methods of anything, i.e. how the results of one test are similar to the other measurement. It clarifies that the application of one test at two different time intervals, how much the results are consistent (www.chfasoa.uni.edu).The stability of the measuring test has referred the reliability (Field, 2005).

Carmines & Zeller (1979) describe:

"Reliability denotes the stability of the conclusions attained and examined by an individual at different time intervals using the same technique, or by using the dissimilar groups of corresponding items, or under additional variable investigating the same conditions."

Explains the meaning of reliability, as to how consistently a test measures whatever it measures (Bryman & Bell, 2003). By enhancing the item numbers of the same quality in comparison to the other items, the reliability of the technique used might be enhanced. To administrator a test with no difference among different groups, cautious directions must be directed to groups, as long as an atmosphere is permitted from interferences and any controlling factor that intended to decrease the boredom and fatigue, also helpful in improving the reliability of instrument used for testing (Falvey, Holbrook & Coniam,1994).

Field, A. P. (2005) has the words When the test scores are free from measurement errors the degree of reliability may be enhanced.

Characteristics of Reliability:

Characteristics of reliability are defining as under:

- The consistency of obtained test results is reliability.
- It is the measurement error of variables.
- Reliability is associated with the length of a test.
- Reliability is the consistency of a test for a specific group of individuals.
- Reliability is considered as a self- association.
- It is considered as the degree of stability and associated with interior consistency.
- It is related to the reproducibility of test scores.
- Reliability is the significant feature of determining an instrument.
- The person and accuracy of a measuring instrument s reliability (Field, 2005).

Types of Reliability:

Different processes are used commonly to assess the reliability of a test. These comprise

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

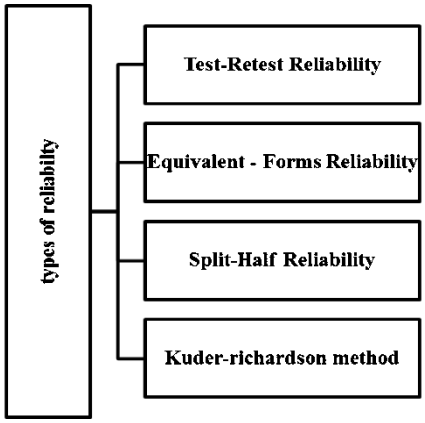


Figure No 02: Types of Reliability

Test-Retest Reliability:

The first one the, Test and re-test describes that reliability is the degree to which the test scores are reliable and consistent over time.

This type of reliability, test-retest, is recognized by determining the association among test score results from controlling the similar test for the identical collection of individuals at diverse time intervals.

Followings are the steps involved:

- Manage the test
- After a specific time interval (e.g. 2 weeks) conduct a similar test again for the same group.
- Have a look at the association of the scores of the two tests.
- If one observes the greater correlation among the results one can say, and the test indicates respectable test-retest reliability (www.opentextbc.ca)

Equivalent - Forms Reliability:

Equivalent forms of a test may be applied, when two completely identical tests to the actual items involved.

The identical test must be specified by considering:

a) Item in numbers

b) Its Structure

c) Its difficulty levels

d) Following the administrator instructions, scoring, and explanation.

e) The measurement of the variable.

It is expected that the two identical tests administered to a similar group, the group may get similar scores in the implemented tests.

Steps:

- Administrate one type of test by using a specific group of individuals.
- Administrate another type of test by using a specific group of individuals as before.
- Correlate the scores of two tests obtained from a different test.
- A reasonably high degree of association designates respectable equivalent form's reliability (www. research-methodology.net).

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

Split-Half Reliability:

This type of reliability i.e., split-half- reliability is determined by establishing the association among the scores obtained from two corresponding splits of test conducted for a group at some point in time. Gronlund, N. E. (1985) was of the view, A reliability coefficient was acquired by associating the score of the first half of the test with the scores of the second half, then by employing the spearman and Brown method to adjust for dual measurement in length for the entire test score. Usually, not essentially, two halves contain the odd and the even-numbered items.

Here, the test items are separated into two equivalent halves. Let begin an experiment with 20 items that may be divided as under:

i. Odd item: $1, 3, 5, 7, 9, 11, 13, 15, 17, 19$

ii. Even items: $2, 4, 6, 8, 10, 12, 14, 16, 18, 20$

To calculate the split-half reliability the scores of odd items are associated with the even items. The Spearman-Brown formula is formulated as under

$$r\ total\ test = \frac{2r\ split\ half}{1 + r\ split\ half}$$

The steps are as follows:

1. Administrate a test for a specific group.
2. Split test-items into two groups as odd items and even items.
3. Calculate the scores for each half group.
4. Correlate scores of the two sets of items.
5. Calculate the Spearman-Brown correction formula.
6. A high correlation coefficient determines good split-half reliability (www. opentextbc.ca).

KUDER-RICHARDSON METHOD:

Kuder-Richardson procedure is the estimations of reliability. This method estimates the reliability with a formula by a single time administration of the test. It allows estimating the internal consistency of the test. This formula Kuder-Richardson 20 and 21 is used (Field, 2005).

FACTORS THAT INFLUENCING RELIABILITY:

The following factors affect the reliability.

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

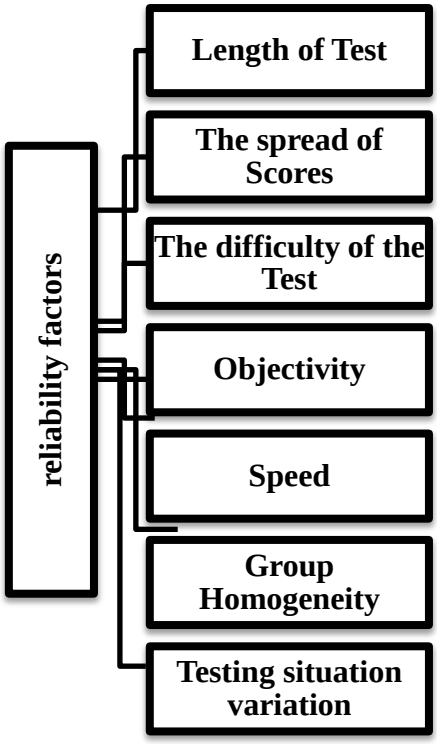


Figure No 03: FACTORS THAT INFLUENCING THE RELIABILITY

Length of Test:

The longer the test is, the higher its reliability will be. This is because a longer test will measure a more adequate sample of behavior and scores are less affected by chance factor or guessing.

The spread of Scores:

The larger the spread of the score is, the higher the estimate of reliability will be. This is because a larger reliability coefficient results when an individual's position remaining the same from one testing to another. The greater differences among scores reduce the possibility of shifting positions.

The difficulty of the Test:

Too easy or too difficult tests have low reliability. This is because both easy and difficult tests result in a restricted spread of scores. For the easy test, the scores are close together at the top. For the difficult test, the scores are grouped at the bottom. The differences among individuals are small, hence tend to be unreliable.

Objectivity:

The standardized tests which are high in objectivity have high reliability. A test is said to be objective if scores obtain the same results. The objective type tests have high reliability and essay type tests have low reliability. This is because scoring is not affected by the personal opinion of the scorer in objective type test. In essay tests scoring is affected by the personal opinion of the scorer. Essay tests have low objectivity and hence low reliability.

Speed:

Reliability would be problematic if one employs the speed test. Every student may not compile completely the all items in a speed test. Alternatively, a power test may enable all the students to complete all items.

Group Homogeneity:

A test may be considered supplementary reliable only if one deliberately uses a more heterogeneous group of students.

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

Testing situation variation:

Deviation throughout the assessment of test, for **i**nstance, noisy atmosphere and interruption that may lead to biased test scores, and may disturb consistency of the test. (www.changingminds.org).

2. VALIDITY:

In this discipline, say, education, no test is accurately valid, because this assessment is not-direct. It is impossible to ascertain that a test that precisely measures and has the capability for the purpose it is designed. Validity is about the degree, such as low, moderate, and higher. Validity relates to the outcomes of any examination but not within instruments. The validity of an instrument can be attained relative to the precise practice of the test (Falvey, Holbrook & Coniam, 1994).

The validity delivers an uninterrupted pattern of, how best it accomplishes its purposes. Frequently, it suggests the degree to which any test fulfills its purpose. It essential to be well-known that every test for evaluation has its particular purpose. Therefore, the selection of a test must be based, and we must look what is the purpose for which it is being used and it is considered an indispensable characteristic of the test. A measurement purpose must be met within a chosen instrument (www.scribbr.com).

Huck (2007) said, "Validity index demonstrates the degree for which a test produces what the test is intended to measure"

Furthermore, the idea of validity is connected with personnel honesty. Honesty is defined in a test when an individual potential to do with no biases. Briefly, validity mentions, how good any test accesses what it endeavors to measure (Oluwatayo,2012).

Characteristics of Validity:

In the following, the characteristics are defined about validity,

- It is a significant characteristic of any measuring test.
- Reliability measures the variable error while validity measures constant error.
- It is considered as an index of peripheral associates. The outcomes are associated along with an external score.
- The measure must be established according to the set of processes as a predictor for the future course of test scores.
- It is related to the objectivity of the measurement of the score.
- Ensuring the validity ensures the reliability of any instrument. Though, a test is considered valid when it is reliable.
- It denotes the honesty of test scores.
- Validity changes with the change in length.
- It controls the performance of an individual with varying situations (www.explorable.com)

Nature of Validity:

Certain attentions are as under, and one should keep in observance by utilizing the validity in assessment:

- It denotes the consequences of tests for a specified group not by itself.
- It is a comparative measure and a conducted test must be valid for a certain situation.
- A particular measurement is considered not valid for every circumstance.
- There are various methods to measure validity (www.socialresearchmethods.net).

Types of validity:

Commonly utilized types of validity in education discipline are described as under.

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

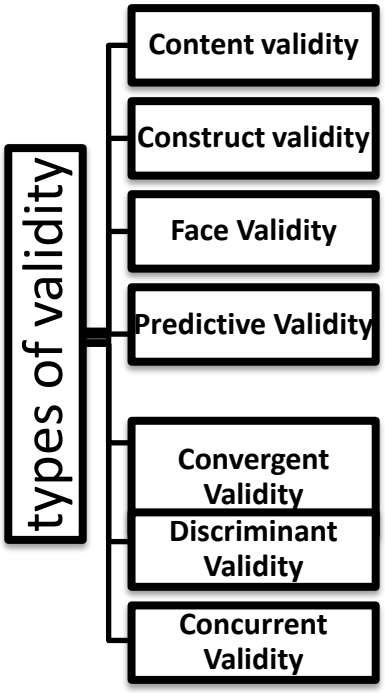


Figure no 04: Types of validity

Construct validity:

The degree to which the outcomes of any procedure of observation and assessment lead to the theoretical constructs for which the method is based. This category of validity, the construct, is correlated with an external type of validity, and both of them move from specific to general. But there is a little alteration that is specified as external validity simplifies in context of study or individual, places, time. The other type of construct validity encompasses the generalization of the procedures. These procedures are used for measurement that precisely reflects the situation construct. It undertakes that the investigator must have an unambiguous description of the construct. One can similarly check the method one has used against this one. Suppose, one method is used must occupy in predictable manners in comparison to the additional method used are constructed upon the theory of construct (www.acadstuff.blogspot.com).

Construct validity is a type of valuation of how well an individual interprets ideas and theories into definite measures. It is observed as truth in labeling. This is an illustration of the concept. The idea of construct validity is regarded from a descriptive perspective. It is also the type of rationalist perspective. Agreeing with the viewpoint of the individual that describes constructing validity that guarantees that it ought to describe the construct precisely. This standpoint, states that an individual whichever operationalizes the construct appropriately or does not (www.study.com).

Content Validity:

This is a procedure of corresponding the test items along with instructional aims. This type of validity is non-statistical. Content validity is distinct in extent where test procedures an illustrative sample, behavioral, and subject matter changes are taken into consideration. Content validity of any test may continuously be observed concerning specific aims to measured (Falvey, Holbrook& Coniam,1994).

Face Validity:

The Face validity demonstrates to measure for which is intended to measure. The question at hand for face validity is not unique in the traditional sense but relatively one in the relationship and in dealing with public relations. To test the content validity of any examination remains a vigilant examination concerning test items and the objectives must be formulated. It is said, face validity, what the test procedures are not, but whatever a test 'seems to measure'. Content of test may not perform an incorrect test. It is fundamentally created on the decision of quite a lot of specialists. In the study at hand, we will give a test to diverse experts. They resolved items included in

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

the test then give recommendations. The researcher made essential changes according to the proposals given by the experts. (www.thewisdomthatworks.com).

Predictive Validity:

The ability to predict something is evaluated in predictive validity. It hypothetically enables to predict and assess. The validity of any test is evaluated on a recognized criterion. If the test's predictive validity is considered high if the scores predict upcoming performance. The test is conducted and scores may attain based on a certain criterion and the outcomes are associated with another criterion (Fowler, 2002).

Convergent Validity:

Convergent validity expresses the degree how which the observation test is comparable to other tests. It illuminates the practical facts. An example may provide clarity of the concept: The scores of achievement may be correlated with scores on some additional attainment. And the test is important to access success. The high degree of association indicates convergent validity (Gronlund,1985).

Discriminant Validity:

"Discriminant" explains the concept of differentiation. This validity measures the extent to which one test provides different results in comparison with the other test. It is a validity measure that clarifies, the observed fact, must not be comparable with each other. For example, discriminant validity with any training program. As anyone collects proofs that express the program's similarity to some other development programs, may not be a training program. Then the discriminant validity one can be correlated motor skills with scores of the test along with scores with the cognitive skill test. Low-slung correlations indicated discriminant validity (Kubiszyn & Borich, 2003).

Concurrent Validity:

It is the degree, to which the marks on a test are associated with the marks on another already recognized test conducted at the same time (Field, 2005).

In concurrent validity, the observation approaches to differentiate amongst groups measured academically and it would be capable to differentiate among groups. When an examination provides an estimate of specific performance then it is considered a test that possesses concurrent validity. The method administrates a test and gets a score. Then scores after obtaining from some other performance are associated. This provides a pre-established database (Oluwatayo,2012).

Steps:

- Manage the novel test.
- Manage already settled valid tests to a similar group at the same time or soon thereafter.
- Relate the two sets of scores.
- A high relationship coefficient will designate good concurrent validity (Falvey, Holbrook & Coniam, 1994).

Factors affecting the Validity:

Any validity may be subjective and varied by a great number of aspects. Hughes (2003) suggested the factor that stimulates validity of the test:

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

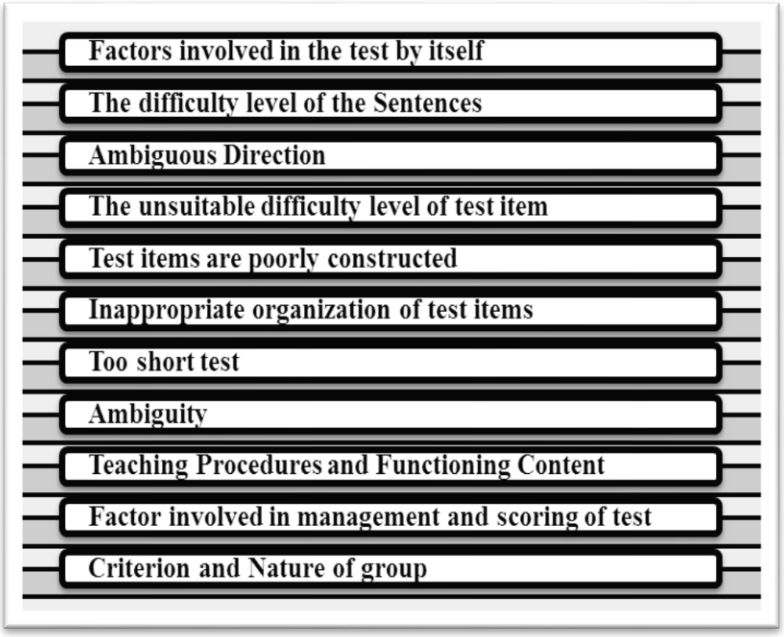


Figure No 05: Factors affecting the Validity

Factors involved in the test by itself:

The test comprises the items, we would establish, whether the test seems to measure the content or subject matter or the behavioral function are established.

Ambiguous Direction:

Correct and perfect instructions must be given to the investigators concerning the test items, besides these instructions the validity of the test may be affected.

The difficulty level of the Sentences:

The difficult and complicated sentences that that is above the level of understanding of the individuals, but a limit on the validity of the test.

The unsuitable difficulty level of test item:

Unsuitable exertion (difficulty) level test item may disturb test validity.

Test items are poorly constructed:

The test items that offer unplanned pieces of evidence to response may correspondingly vary the test validity.

Inappropriate organization of test items:

The easy to the difficulty level of test item's arrangement may as the respondents may fail to an extent easy test items, at last, they may leave the test unfinished. Furthermore, this organization will disturb the validity of the test.

Too short test:

Even when the test is conducted too short, to turn out not to be demonstrative of, then validity will be at risk.

Ambiguity:

Ambiguity indicates misperception and misunderstanding that will lessen the test validity. A good learner can attain a response while a poor can acquire a wrong answer.

Teaching Procedures and Functioning Content:

The test is valid only if the test items function as intended.

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

Factor involved in management and scoring of test:

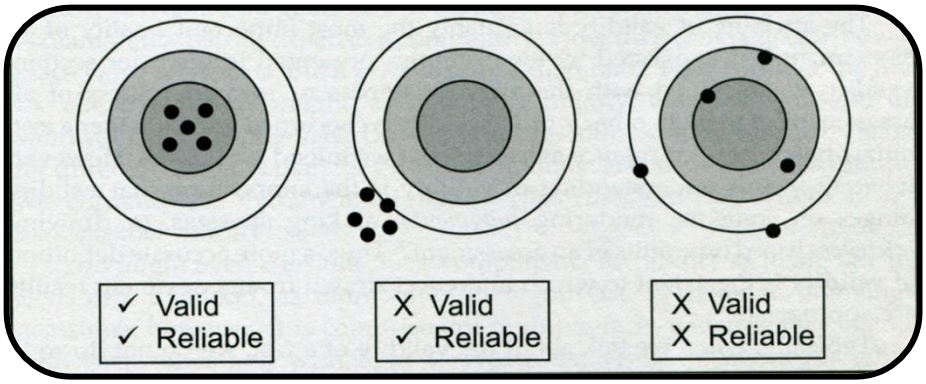
There are many factors influence teachers who have created the test. For instance, elements just like biased help to the respondent, inadequate time to accomplish the test, and cheating and unfair scoring might lead to low validity of the test.

Criterion and Nature of group:

Specific influences foe instance, age, capability, sex, educational context, culture, influences that encourage the test validity. Henceforth, the nature of the criterion utilized may be considered significant though assessing validity coefficient

Association among reliability and Validity:

Validity and reliability are closely related to each other. These aspects are the two Measurements of similar objects entitled the test efficiency. Reliability is associated with consistency of any test scores self-correlation of the test, while validity is the association of test with some external criteria. Consequently, a test that owns poor consistency cannot expect to have high validity. Therefore, reliability is the precondition for validity. A good degree of reliability is valid at all times. This describes that reliability covers validity. Furthermore, a test is theoretically valid, on the other hand essentially unacceptable and judged with the association of some other criteria (Viswanathan, 2005).



3. OBJECTIVITY:

The objectivity of a test mentions the degree that correspondingly produces competent scores with similar results. Most standardized aptitude tests are high in objectivity. The test items are always of objective type (e.g. multiple choice) and resulting scores must not be of subjective nature. Such examinations are usually assembled so that they can be precisely scored by

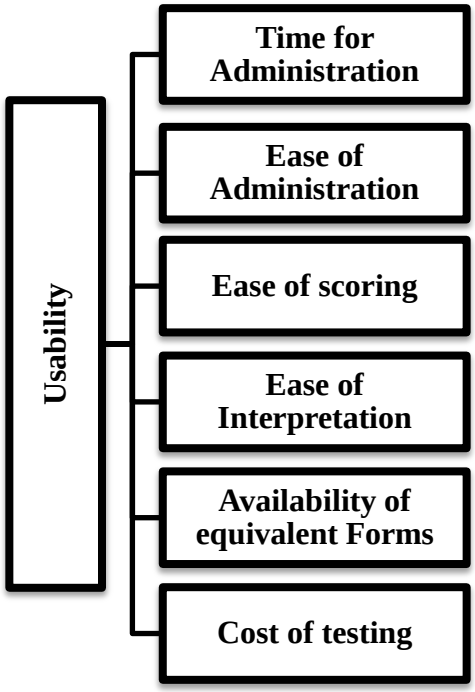
qualified assistants and scoring technologies. When such casually objective procedures are used, the reliability of the test consequences is not affected by the scoring procedures (Kubiszyn & Borich, 2003).

For classroom tests created by teachers, objectivity may perform a significant role in obtaining reliable measures of attainment. In essay testing, the results are influenced by a large amount of the individual doing the scoring. Different persons get not the same results, and even the same person, may get different results at different periods. Such inconsistency in scoring harms the reliability of the measure obtained, for the test scores now reflect the opinions and biases of the scorer as well as the differences among pupils in the characteristics being measured. In the subjective test, for instance, objectivity can be increased by vigilant phrasing of the queries and by a standard set of rules for scoring (Huck, 2007).

4. USABILITY:

Kubiszyn & Borich (2003) has explained the usability of a test in this way:

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021



Time for Administration:

The smaller test is favored but how much they are reliable. A safe procedure is to allocate as considerable time as is essential to obtain valid and reliable results. Everywhere among twenty to sixty minutes of the testing period is perhaps a respectable guide.

Ease of Administration:

A test will be easy to manage when one considers the following

- Simple and clear direction.
- A few subsets.
- Appropriate testing time

The complex instructions, some sub-tests, and short time unfavorably affect the rationality and reliability of a test.

Ease of scoring:

Those tests are favored that offer comfort and economy of scoring lacking sacrificing scoring accuracy.

Ease of Interpretation:

When the test results are obtainable to the individuals, ease of interpretation and presentation are particularly significant. If results are appropriately interpreted, they contribute to imperative educational conclusions.

Availability of equivalent Forms:

Equivalent forms of a test quantity the same aspect of behavior by using test items that are the same in contents, difficulty level, and other appearances. The equivalent procedures of the test are often desirable for instance, e.g. if a teacher may feel that score of a student in an achievement test is very low, he may easily check by managing the equivalent form.

Cost of testing:

ISSN- 2394-5125 VOL 8, ISSUE 01, 2021

The test would be economical but sacrificing valid and reliable tests of being high cost and selecting cheaper tests is a false economy (www.interaction-design.org).

CONCLUSION:

The test is an integral part of the assessment process without appropriate testing of students we cannot make the valid decisions (placement, promotion, grading, and classification, etc).And for reliable and valid testing of students teachers must keep in mind these features of sound test so that they made an appropriate test for their students. Validity and reliability are closely related to each other. Reliability is associated with consistency of any test scores self-correlation of the test, while validity is the association of test with some external criteria. Consequently, a test that owns poor consistency cannot expect to have high validity. Therefore, reliability is the precondition for validity. The process of testing must be error-free and biased free of the scorer.

REFERENCES

Bryman, A. & Bell, E. (2003). *Business research methods,* Oxford, Oxford University Press.

Carmines, E. G. & Zeller, R. A.(1979). *Reliability and Validity Assessment,* Newbury Park, CA, SAGE.

Falvey, P., Holbrook, J., & Coniam, D. (1994). *Assessing students*. Hong Kong, China: Longman.

Field, A. P. (2005). *Discovering Statistics Using SPSS*, Sage Publications Inc.

Fowler, F. J. (2002). *Survey research methods,* Newbury Park, CA, SAGE.

Gronlund, N. E. (1985). *Measurement and evaluation in teaching* (5th ed.). New York, NY: Macmillan.

Huck, S. W.(2007). *Reading Statistics and Research,* United States of America, Allyn & Bacon.

Hughes, A. (2003). *Testing for language teachers (2nd ed.)*. Cambridge, England: Cambridge University Press.

Kubiszyn, T., & Borich, G. (2003). *Educational testing and measurement: Classroom application and practice*. New York, NY: Harper Collins.

Oluwatayo, J. 2012. Validity and reliability issues in educational research. *Journal of Educational and Social Research* 2**,** 391-400.

Viswanathan, M. (2005). *Measurement error and research design*, Thousand Oaks, CA: Sage.

https://socialresearchmethods.net/kb/types-of-reliability/

https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/

https://chfasoa.uni.edu/reliabilityandvalidity.htm

https://research-methodology.net/research-methodology/reliability-validity-and-repeatability/research-reliability/

https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/

https://www.scribbr.com/methodology/types-of-validity/

https://explorable.com/types-of-validity

http://changingminds.org/explanations/research/design/types_validity.htm

https://prezi.com/rud4qgzrxjis/what-is-a-good-test-reliability-objectivity-and-feasibility/

https://acadstuff.blogspot.com/2017/06/objectivity-characteristic-of-good-test.html

https://www.thewisdomthatworks.com/targeting-objectivity-reliability-validity/

https://study.com/academy/lesson/what-is-usability-definition-tools.html

https://www.interaction-design.org/literature/topics/usability