
APPLIED PSYCHOLOGICAL TESTS IN EDUCATION

Abubakar, Hussaini Yoffo

Department of Psychology

Federal College of Education, Yola, Adamawa State, Nigeria.

E-mail: abubakarhussaini121@gmail.com

Abstract: *The paper sought to highlight psychological test and describe it as a systematic procedure for measuring sample of people's behaviour. This paper equally identified various types of psychological tests like intelligence tests, personality tests interest, and attitude achievement tests amongst others. The paper also looked at the strength of psychological test which include prediction selection and classification. The paper identified anxiety and stress amongst others as limitation of psychological tests. The paper recommends that teachers, counselors and parents should be acquaintance with the current psychological tests in order to understand references to the use of such tests in the psychological literature.*

Keywords: Education, Psychological Test, Teachers Reliability.

Introduction

Psychological test is a field characterized by the use of sample of behavior in order to asses' psychological construct(s) such as cognitive and emotional functioning about a given individual. The technical term for the science behind psychological testing is psychometrics. By sample of behavior, one means observation of individual performing tasks that have usually been prescribed before hand, which often means scores on a test. These responses are often compiled into statistical, tables that allow the evaluator to compare the behavior of the individual being tested to the responses of a norm group to achieve this. The trainee should be exposed to a wide range of psychological test and non-test technique. This paper deals with psychological test and testing.

Definition of Psychological Tests

A psychological test is an objective, often standardized, method for measuring individual difference and personality variables of students related to some psychological concept or construct, based on a sample of relevant behavior in scientifically designed and controlled situation (Passer and Smith in Akinade 2009 :86). According to Dreyer (1974) a psychological test is a standardized type of examination that is given to an individual or group with a view to obtaining their cognitive and non- cognitive traits. A psychological test is a stimulus or instrument applied on individuals in order to elicit certain responses that can be assigned scores on the basis of some specified psychological characteristics (Akinade 2009). Thus; score obtained assist the tester to infer the degree to certain traits possessed by the tested individuals. Testing is the act of exposing an individual to a particular test or set of questions in order to obtain scores which can give information needed in making decision on the individual.

Aiken (1979) defines a psychological test as any device used to evaluate the behaviors or performance of an individual. It is any measuring tools or device that is employed to asses, diagnose and evaluate the cognitive and non- cognitive traits, behaviors and performance of an individual or group of individuals (Obidigbo, 1997). Tests are quantitative and qualitative since they help to determine the presence or absence of a particular knowledge, capacity, skills, ability or disability. A useful psychological test must be both valid (i.e. there is evidence to support the specified interpretation of the test result and reliable, (i.e. internally consistent or give consistent result overtime, across raters etc.).

Characteristics of a Good Psychological Test

A Good Test according To Akinade (2009) should have the following characteristics. Reliability, validity, standardization, availability, eases of administration and culture-fairness. Reliability: refers to the consistency of the measuring ability of a test to produce similar result on different occasion from the same testee (Ravid, 2000). According to Anastasin (1988) reliability refers to the consistency of scores obtained by the same persons when re-examined with the same test on different occasions or with different sets of equivalent items, or under other valuable examining conditions. Thus a test is reliable if the student obtains consistent. Scores, which is there are not wide range fluctuation from one test to another section to it. Consistency does not mean that the student should score exactly the same points each time he takes the test slight differences in scores are expected and are acceptable. Since all types of reliability are concerned with the degree of consistency or argument between two independently derived sets of scores, they can all be expressed in terms of a correction coefficient. A correlation coefficient according to Anastasi (1988) expresses the degree of correspondence of relationship between two sets of scores and this is always reported in decimals. She added that this varies from -1 through 0.05 $+1$ with -1 indicating perfect negative corrections, 0 indicating zero correction and $+1$ perfect positive correction. Thus; perfect correction of -1 and $+1$ are however rare to get. The sign $-$ or $+$ of a coefficient indicates, the direction of the relationship, that is whether it is negative or positive. If a positive correction is said to exist between tests x and test y whose scores have been correlated, it means that high scores on test x are associated with high scores on test Y and that low scores on test x are associated with low scores on tests Y . Correlation coefficient may be completed in various way, depending on the nature of the data. The most common is the Pearson product moment correlation coefficient and the spearman rho ranked order correlation.

Type of Reliability

- a. **Test-Retest Reliability:** This involves repeating the same or identical test on a second occasion in order to find the reliability of the test score Akinade (2009). This form of reliability according to him allows for the administration of a single test to a particular individual or group on two different occasions a little after the first administration. Thus it is the most direct method of estimating the reliability of a test. The reliability coefficient here is derived from correlating the two set of scores obtained in the two administration of the test by the same individual or group. Test

retest reliability is mostly accompanied with the same intervals upon which it was re-administered. Short intervals say .6 months are better than long intervals and younger children's period should be shorter than older persons.

- b. **Paralleled or Alternate-form Reliability:** In Alternate or parallel-form reliability Anastasin (1988) Asserted that, the same individual is retested with an equivalent test that is an alternative or equivalent to the initial test. Thus using this method, the same persons can be tested with one form of the test on the first occasion. On second occasion they will be tested with another form that is comparable to the first the correlation between the scores obtained on the two forms represents the reliability coefficient of the test. Such a reliability coefficient measures not only temporal stability but also consistency corresponds to different item samples on test forms.
- c. **Inter-scores reliability:** This measures the degree of correspondence among various examiners who score a given test. These types of reliability are usually used when there is expectation that scores bias might lead to inaccurate result (Morah, 1995). Generally, inter-score reliability involves two or more judges (examiners) marking a test or script while their respective score are correlated in the end with a view of determining the score reliability. These types of reliability are mostly used in the scoring school debates and marking essays. One source of error variance here is scorer's variance. One source of error variance here is scores variance. This reliability is mostly used when subjectively scored instrument are employed in research.
- d. **Split Half Reliability:** By this method reliability is measure by obtaining two scores for each person by dividing the test into two halves that can be easily compared for instance in a test of 40 items, all the odd numbered items can be made to be one half test and all the even-numbered items on the other half. That is split half reliability make use of one test which is administered single but the test is divided into two equal halves, say odd and even part. The examiner scores the odd number items separately and also the even numbered items separately. Each testee obtains two set of scores-one for even numbered items and for odd numbered items. The reliability coefficient here, some time called coefficient of internal stability is from correlating the two sets of scores obtained by individual from the two halves of the test.
- e. **Full Scale (Kuder-Richardson) Reliability:** this form of reliability is obtained single administration of a test to an individual on group of examinees. It gives the reliability of all the items in the test and it is also a measure of the internal consistency of a test.

Validity

Refers to the extent to which a test measures what it is supposed to measure, the validity of a test therefore depends on its purpose. Example: an intelligence test is supposed to measure intelligence and nothing else.

Kinds of Validity

Content Validity: every achievement test used in education should have content validity. Content validity is defined as "the adequacy with which an instrument measures a representative sample of behaviors and content domain about which inferences are to be made (Ravid 2000:264). Teachers using tests prepared by others as well as teachers who create their own tests, should ensure that the tests they use have content validity. That is, the items on the test correspond to the content, knowledge, skills, and behavior being tested. According to Ravid (2000), a well-created table of specifications can ensure the content validity of a test. Other less formal approaches, which include a systematic review of the materials to be tested and the construction of corresponding items, can also ensure high content validity of the test.

Face Validity: According to Aiken (1979), test validity refers to how the people taking the test (testees) see the test item, the administrative personnel who decide on its use, and other technically untrained observers who come across the test. The major fundamental question here rests with establishment of rapport and duplicitous reactions. Face validity refers to what the test appears on the surface to measure. Does the test look valid to the students who take it, the school authorities who decide to use it? Face validity is an essential feature of achievement tests. If a test meant for children in primary five (5) is administered to a student in secondary five (5), it will certainly lack face validity. This is so because the test content would be considered irrelevant or even childish by form five students.

Construct Validity: According to Anastasi (1988), Construct Validity involves determining the extent to which an instrument measures a specific trait or construct. In other words, it involves analyzing the meaning of data obtained from the instrument in terms of psychological constructs in order to rule out other unsatisfactory meanings. Methods of establishing construct validity of an instrument include correlation with other valid instruments, appeal to logic, analysis of mental processes required by the instrument, experimental interventions, criteria group studies, the use of factor analysis, and Cronbach's coefficient alpha statistics. According to Aiken (1979), evidence for construct validity of a test can be obtained by the following ways.

- i. Expert judgments that the content of the tests pertains to the construct of the interest.
- ii. An analysis of the internal consistency of the test.
- iii. Factor analysis of the correlations of the test with other tests and variables.

Criterion-Related Validity: This type of validity indicates the effectiveness of a test in prediction or estimating an individual's behavior in specified situations. This serves as a

direct measure of the trait or ability which the test is intended to predict, for instance grade or mark may be the reiteration for a scholastic aptitude test while subsequent job performance as a machinist may be a reiteration for mechanical aptitude test. Criterion related validity can be achieved by giving a test to a group of people who are either entering some job or training programme. After the test the correlation between the test score and criteria measure of success is computed. The test is an effective predictor to the extent the correlation is high. However, the following factors can affect the validity of test namely unclear instructions, poor construction of items, ambiguous statement, inappropriateness of the items and sentence structure (Ezeh, 1992). Correct for this need to standardize our tests, use simple language and give clear instructions.

Standardization: According to Akinade (2009), Standardization refers to both standard items and procedures for obtaining score and the development of norms. It involves establishment of well-controlled environment for administering the test. Norms refers to collection and treatment to test result obtained from a large sample that represents particular age segments of the population. The normative scores provide a basis for interpreting given individual scores.

Availability: - a test should be easily available in good quantity in order for it to be useful. Those who are qualified to use them should know where and how to find and procure them. If a test is not readily available it will not be used. Good test too should not be too costly otherwise they may not be used with ease.

Usability: This refers to the extent to which an instrument technique (test) can be used, easily used to obtain interpretable information that can not be of very little value to the school system. Usability of assessment instrument/technique is associated with the following factors, simplicity, time required for instructions, ease of application) ease of scoring, ease of interpretation of its scores and availability of equivalent forms Universal Basic Education Programme. (UBEP, 2000:28).

Ease of Administration: Each test must have manuals that will assist qualified users in administering the tests. The test should be easy to score, analysis and interpreted. They should not be cumbersome or too lengthy.

Culture-Fairness: - A good test should be culture- fair or unbiased. If is culturally unfair, the result and interpretation may also be biased, unfair and subjective. Its contents should not contain unfamiliar vocabulary or ideas that are strange to student taking the test. Examples of culture-fair test include draw- a test and convent's stand and progressive matrices.

Uses of Psychological Test:-At present school are among the largest test users. The classification of children with references to their ability to profit from different types of school instruction, the identification of the intellectually retarded on the one hand and

the gifted on other, the diagnosis of academic failures, the educational and vocational counseling of high school and college students, and the selection of an applicant for professional and other special schools are among the many of educational uses of test (Anastasi, 1988). Thus, a test is needed whenever information is needed to help make decision about an individual. Some form of test (educational, achievement or placement test) is needed whenever we want to decide on courses of action affecting individual's lives and testing is invaluable in hospitals, clinic, counseling session, psychotherapy centers selection, placement, promotion and terminating of workers etc. Test are used extensively in school for curriculum planning and behavior modification. Aiken (1979) affirms that test is used to describe the interest of an individual such as his job oriented character. Test help to describe the interest one has on a given construct and test when administered shows the quantitative and qualitative nature of his interest. It can also be used to make statement about the potential life career of an individual's job satisfaction.

According to Ipaye (1980) in Akinade (2009:88) test could be used for the following purposes: as achievement tests to measure academic attainment and diagnosing academic strength or weakness in curriculum based subject or course; as aptitude tests, intelligent, tests and personality and vocational inventories, attitude scales, personality tests and projective techniques. Olayinka (1980) added that test, can be used for motivating learners, identifying those who need remedial support or referral, appraising individual skills and placement of student in appropriate programme. Generally, tests yield setting types of information more accurately, meaningfully and economically than subjective techniques e.g. observation or interview or projective techniques. Traxier (1960) listed ten (10) ways can use test result from school testing programme on their summarized below:

- Assessing abilities and aptitude.
- Comparing achievements with scholastic aptitude.
- Judging the level of ability or achievement in a class.
- Improving self-understanding in a counseling or psychotherapeutic encounter.
- Diagnosing strengths and weakness.
- Identifying student interest.
- Helping students with adjustment problems.
- Guiding student towards educational and vocational goals.
- Conducting researchers.
- Discovering specific abilities

Limitation of Tests

Some Psychologist has used tests as if tests have all the answers. Tests should not be seen and used as such. Some tests are technically suspects and they should be avoided, some also are likely prone to misinterpretation:

Not many people can analyze and interpret tests properly. In summary, the following are some of the limitation of psychological tests;

- i. Testing leads to test anxiety, especially among students.

- ii. Testing lead to over emphasize on certification especially educational and achievement tests.
- iii. Educationally, testing restricts teaching for it leads to teaching to the test that is narrowing teaching to the specific area that would be included in the examination.
- iv. Tests are faked at times: This is because some individuals give society desirable responses to test items in an attempt to sound "holy" or normal affective tests are mostly abused.
- v. Testing can be misleading and erroneous especially when the measuring devices are wrongly worded or interpreted.
- vi. Test stigmatizes clients and hands to alienation of such clients.

Classification of Tests

Psychological and educational tests are variously classified. Most psychometricians dichotomised tests in to two broad categories namely standardized and non standardized test or cognitive and non-cognitive test.

Standardized Tests

Standardized test refers to a test that has fixed directions administration, scoring and it is normally constructed by a professional under standardized conditions (Anastasi, 1988). Thus, such tests are normally administered to a representative sample of examinees from the population for whom the test is intended. This type of test has manual and norms which save as basis for interpreting the score of subsequent examinees.

Non-Standardized Test: are opposed to standardized test and teachers, normally construct them in an informal manner for a single administration. Teacher made test (as they are called) are distinguished from standardized test because the teacher made test are constructed by the classroom teacher for the purpose of measuring the achievement of his own students. Most tests used in our educational system commonly referred as classroom test are not standardized.

Cognitive Vs Non-Cognitive

Cognitive test try to find out how much the learner or testee knows. They are primarily concerned with knowledge hence the name intelligence test, achievement test and aptitude test are good example of this type of test. Non-Cognitive Test (affective tests) are test which try to find out personal qualities and behavior characteristic of learners (testees). These tests are not concerned with knowledge and are very useful in counseling, psychotheraphytic services and curriculum planning. They asses the affective behavior and personality of individuals since they measure how we feel that is, they measure our interest, values, preference, attitudes etc. They include adjective checking, personality inventories and attitude scale etc.

Some Major Cognitive Tests

The following are major cognitive test in psychology;

- a. **Achievement Tests:** are for measuring attainment after exposure to a specific course of study, their objectives include identification and or subject for diagnosis of discrepancies in the learner's and instructional methodology. Achievement tests measure what a person has atrophy learned. This achievement test are cognitive ability test which measure how much the individual has been able to learn from what he has been taught formally such as in a classroom situation. Examples of achievement test are senior school certificate, west- African school certificate, Nigeria certificate in education (N.C.E) etc.
- b. **Intelligence Tests:** the mental ability tests measure intellectual status of an individual. They measure a wide range of cognitive skills such as motor co-ordination, perception, conception, memory, language, reasoning and creativity. Intelligence test can be administered on individual like the Simon-Binet or Standfort- Binet intelligence scale or the Wechsler adult intelligence scale (WAIS). It can also be administered to a group (Group intelligence test) like the Otis-Lennon mental ability tests.
- c. **Aptitude Test:** According to Anastasi, (1988), an aptitude test is a test designed to measure potential success in learning and performance. It often measures aspects of intelligence and achievement abilities. Thus it is seem to be fairer than achievement test since it depends less on prior knowledge. But it is difficult to construct aptitude test that is free from prior learning. Some of the aptitudes that one can test for include- the Differential Aptitude tests (DAT), Verbal, scientific, numerical, clerical, spatial musical, mechanical artistic and scholastic aptitudes. Usually, aptitude tests appear in batteries group of separate tests which measure various aptitudes. Foreign examples include: the scholastic aptitude tests (SAT), Law school Aptitude test (LSAT), Armed services vocational Aptitude Battery (ASVAB), General Aptitude test Battery (GATB), The joint Admission and Matriculation Examination (JAMB).

Major Non Cognitive Tests (Affective Tests)

These include;

Personality Tests: these measure non-academic characteristics of testees, they could be aptitude test projective tests, motivational scales or interest inventories. Personality inventories are used to ascertain personality traits, for personal counseling, gives information on personal problems, dimensions, depth and emotional issues. They often measure the more stable and enduring characteristic of a person. They can measure degree e.g. anxiety shyness or need for love. These- are tests that are designed to reveal the extent to which an individualist adjusting to the conditions facing him. They are also used to indicate whether the individual is developing normally. Personality test can come inform of

personality inventories and for projective tests. Personality inventories include the Minnesota multiphase personality inventory (MMPI), eyeneck personality inventory (EPI), cartel and ember's 16 personality factors questionnaire (16PF), Minnesota multiphase inventory (MMPJ), Taylor manifest Anxiety scale saran son's test Anxiety scale STAS. (General and situational) Akinade's Shyness Personality Scale, SPS Bakare's 5tUE1t problem inventory, SPI, study habit inventory SHJ (Bakare), adolescent personality data inventory APDI (Akinboye) and, stress behaviors scale, SBS (Akinboye).

Interest Tests: An interest test inventory measures activities one will enjoy for its own sake; it is useful for ascertaining areas of interest and learning. It could relate to different areas or group. Example include: the strong vocational interest blank (SVIB)Vocational interest inventory, VII (bakare) vocational interest survey, Kuder Preference Inventory, OI-iIO, strong vocational interest inventory (SVJ)Blank.

Attitude Scales: Attitude scales are measured with instrument called attitude scales a self report inventory designed to measure the degree of favorableness or unfavorableness with regard to an individual's opinion towards an attitudinal object. Attitude test usual has three component cognitive, affective and behavioral attitudinal behaviors. Examples of these scales are the Thurstone scaling method and the like method which is most popular.

Check List: this is an instrument for assessing affective behavior. It determines the presence absence of certain personality traits. The examiner (observer) simply recorder can tick a good () or yes (y) against the action or behavior to indicate that it occurred. Because this method is time consuming, the rating techniques are highly preferred. This rating technique used rating scales which indicate the degree or frequency of the occurrence of a behavior in the individual.

Test Administration Guidelines

According to Akinade (2009:89) test administration guideline include the following:

1. The testees should be adequately trained to do so.
2. The testees must know and understand several local and international tests.
3. The testees must follow meticulously the manuals (guidelines) of test constructor e.g. style of administration, timing, location, space and light.
4. Employ proctor when it is group test e.g. a class or group of 25 or more testees.
5. The order of test administration should be establishment of rapport, presentation of answer sheet the test(s) collect the test(s) collect the answer sheets. Ensure the number of copies of test retained is the same as the number given out especially in aptitude test.
6. Collect and store the test scores very properly.

Conclusion

Psychological test should be regarded as tools. Like all tools their effectiveness depends on the knowledge, skill and integrity of the user. Since psychological test are measures of

behavior, the interpretation of test result requires knowledge about human behavior familiarity with relevant behavioral research is needed not only by the test constructor but also by test user.

References

- Aiken, L. (1979). *Psychological Testing and Assessment*. Boston Allu and Bacon.
- Akinade, E.A. (2009). *Psychological Testing in Counseling in the Counselor* Special Edition Official Publication of the Counseling Association of Nigeria (asson) Volume 25 no.1 p.86-95.
- Anastasi .A. (1988).*Psychological Testing*. New York Macmillan.
- Drever, J. (1974). *A Dictionary of Psychology*, London Penguin Books.
- Ezeh, D.N. (1992).*Reliability and Validity in Nwogu (ed) Educational Measurement and Evaluation*. Awka: Hallman.
- Morah. E. & Obidigbo. C.T.C. (1995). Interview in E.I. Morah (ed) *Psychological Testing and Assessment*. Enugu: Dulacs.
- Obidigbo, G.C.E. (1997). *Recidivision Psychological a Paper Delivered at the 2nd National Conference of N.S.P. Held at Calabar*.
- Ravid. R. (2000). *Practical Statistics for Education (2 ed)*. Lanham, M.D.: University Press of America.
- Ubep (2000). *A Training Handbook for Nigerian Primary School Teachers*.

Reference to this paper should be made as follows: Abubakar, Hussaini Yoffo (2013), *Applied Psychological Tests in Education. J. of Education and Policy Review*, Vol. 5, No. 2, Pp. 81 – 90.
