



# **JOURNAL <sup>OF</sup> FUNCTIONAL MANAGEMENT**

**LEAD CITY UNIVERSITY  
I B A D A N**

**Volume 4 No. 1 April, 2011**

**ISSN 2141-3711**

# CHALLENGES ASSOCIATED WITH THE DEVELOPMENT OF LOCALLY MADE PSYCHOLOGICAL TESTS IN NIGERIA

By

**Olowookere, Elizabeth. I**

Department of Psychology

Covenant University, Ota, Ogun State, Nigeria

[bettymii@yahoo.com](mailto:bettymii@yahoo.com)

## Abstract

*Psychological tests are essential to the study and practice of psychology and are used for recruitment, selection, placement, and classification in commerce, industries, and the military. They are scientifically constructed following careful and detailed procedures based on sound theoretical approach. Psychologists in Nigeria have constructed culturally relevant tests to assess personality in various settings. However, the number of psychological tests in Nigeria is grossly inadequate. Most of the tests used by psychologists in Nigeria are foreign and not suitable for African cultural background. The scarcity of local tests in Nigeria may be as a result of the challenges psychologists in Nigeria face in the process of test construction. This paper therefore discusses some of these challenges which include the shortage of manpower in the field of psychometrics; financial and time constraints; work overload; lack of mentoring and training programmes; and the lack of attraction to psychometrics as a career in psychology. It was recommended among others that Universities in Nigeria should offer psychometrics as an area of specialization in psychology at the postgraduate level and also, that Psychometricians in Nigeria should encourage young psychologists in the field through mentoring, coaching, and by forming and financing research clusters towards the development of psychological tests.*

**Key Words:** *Psychological Tests*

## Introduction

Crucial to the study of psychology is the strict adherence to scientific principles. Psychology may therefore be defined as the scientific study of human and animal behaviour and mental processes with the aim of understanding, organizing, directing, predicting, and modifying behaviour. Gary (2010) defined psychology as the science of mind and behaviour. The scientific component involves the strict adherence to the basic scientific principles of determinism, empiricism, objectivity, and operational definition of terms. Behaviour can be overt or covert, overt behaviours are evident and can be studied through direct observation, while the covert behaviours are clandestine and can be studied through the use and application of psychological testing methods.



Psychology is characterised by both research and application, with the immediate goal of understanding humanity through the discovery of general principles and exploration of specific cases, it is therefore imperative to employ the use of objective and standardised measure of behaviour. Fernald (2008). The branch of psychology concerned with the theory and technique of psychological and educational measurement is called psychometrics; this includes the measurement of intelligence, attitudes, abilities, and traits. Psychometrics primarily entails the task of constructing and validating measurement instruments such as psychological tests and inventories.

Apart from direct observation of the behaviour of interest, unobserved constructs (latent variables) are measured through empirically designed instruments called psychological tests. Anastasi and Urbina (1997) defined psychological tests as an objective and standardised measure of a sample of behaviour. Crucial to this definition are the concepts of validity and reliability. A test is said to be valid when it measures what it purports to measure, on the other hand a test is described as reliable when it yields the same or similar results when measured at different times by the same or different observers.

### **Importance of psychological tests**

Psychological tests are essential to the study and practice of psychology, they represent measures of individual differences on criterion variables and aid in decision making in applied psychology and basic research. Psychological tests are used for recruitment, selection, placement, and classification in commerce, industries, and the military. They aid decisions on transfer, promotion, demotion, or termination of employment, and are also used for selecting employees for training and other personnel development programmes.

Psychological tests are employed to measure individual differences as it relates to job aptitude and abilities, and the results are used as guides to personnel management decisions. In clinical settings, psychological tests are used for diagnosis of various forms of psychological impairment and to monitor the effect of treatments and interventions. They are also used for the examination of the mental health of individuals. Within the educational locale, academic progress is evaluated using psychological tests. These tests help to evaluate the effectiveness of educational programmes by checking whether or not the academic progress of students improved as a consequence of the educational instruction. Tests are also used for placement of students into career options, for diagnosis of areas of academic difficulty and for vocational counselling of students. Furthermore, psychological tests are used as a tool to solve a wide spectrum of practical problems in basic research.

### **Development of psychological tests**

Psychological tests are scientifically constructed following careful and detailed procedures based on sound theoretical approach. The first step in

psychological tests development is the identification and definition of the behaviour to be measured. The test developer must determine the specific construct or trait of interest and operationally define it along the measurement continuum. This gives a clear indication of the aspect of behaviour to be measured and is mostly used to form the name of the test.

One of the greatest challenges of test developers is the creation of effective test items. The writing and selection of test items is the next step after the definition of the behaviour or construct of interest. According to Haladyna, Downing, & Rodriguez (2002), creating effective test items may be more of an art than science, although there is a solid scientific basis for many of the well established principles of item writing. In determining the kinds of items that will be good enough to discriminate among testees on the desired behaviour, there is a need for item generation; this can be achieved through the review of literature and consultation with subject matter experts. The synthesis of all the information gathered assists in getting a concise idea for item generation. There may be need to carry out initial job analysis before test item generation for tests being developed for selection and placement purposes (Ehigie 1999). A large pool of items is generated from this approach which will be subsequently screened to arrive at desired number of items.

Training of item writers is an important validity issue associated with test development. (Downing & Haladyna, 1997). Without proper training most novice item writers tend to create poor quality, flawed, low cognitive level test items that do not discriminate in any way. Being a subject matter expert cannot be substituted for effective item writing expertise, effective item writing skills must be learned and practiced. Haladyna (2004) posited that for new item writers, it often helpful and important to provide specific instruction using an item writer's guide, paired with hands-on training workshop. According to Jozefowicz, Koeppen, Case, Galbraith, Swanson, & Glew (2002), the instruction-practice-feedback-reinforcement loop is important for the effective development and maintenance of solid item writing skills. Therefore, the feedback from expert and peers is required in the training of new item writers.

On the generation of test items, it is pertinent to conduct test item selection. The large pool of items generated is usually indistinct, and so a pilot study is required to help revise these items. The data collected from the pilot study or pretesting are scored and analysed. Item or factorial analysis is made to enhance the identification of the items that are not suitable for the test. According to Anastasi (1968), the pool of items initially generated can be reduced or shortened through factor analysis. Factor analysis assists the test developer to identify the items that are good enough to be retained in the final format. The substandard items are either reworded or completely rejected. The pretesting process continues until satisfactory result is achieved. The selection



of items could be made by the test developer as well as in the course of establishing reliability and validity of the test. The process of test construction is completed when the final format of the test is prepared. At this stage all standard items have been revised and the test is taken through the standardization process to establish its psychometric properties in terms of reliability, validity and the establishment of test norms. The nature of the test whether speed, power, or personality test determines the scoring pattern.

Since the inception of test development, the test developer decides the nature of the item format for proposed examination as in the case of achievement tests. There is a need for an objectively scorable item format for most large scale and cognitive achievement testing. According to Haladyna (2004), the multiple choice format (and its variants), with some ninety years of effective use and an extensive research basis, is the item format of choice for most testing programs. The principles of writing effective, objectively scored multiple-choice items are well established and many of these principles have a solid basis in the research literature. (Downing, 2002b, 2004; Haladyna, 2004; Haladyna and Downing, 1989a, b, Haladyna, Downing, Rodriguez, 2002). There is strong research evidence demonstrating the high positive correlation between constructed response and selected response item scores for measuring knowledge and many cognitive skills Rodriguez (2003). This makes the multiple choice item format acceptable for achievement tests. Downing (2002a) argued that the multiple choice item is the workhouse of the testing enterprise. It is an extremely versatile test item form; and can be used to test all levels of the cognitive taxonomy including very high level cognitive processes.

The writing of test manual is the final step in the development of psychological tests. Cronbach (1971) posited that the manual is the principal source of information about the quality of a published and standardised test. The manual provides detailed information on the test which includes administrative and scoring procedure, test objective, psychometric properties, and the norms of the test. Any standardized test that is published for public use must have the manual which must be written explicitly by the test developer. The standards for test development emphasized the documentation of the methods used to establish test specifications and blueprints, their rationale and the evidence the test developers present to support the argument that the particular test specification fairly represents the content domain of interest (AERA, APA, NCME, 1999).

#### *Typology of psychological tests*

Several categories of psychological tests been developed to measure different types of behaviours, some of the common psychological tests are personality tests, achievement tests, intelligence tests etc.

#### *Personality inventories*

Personality inventories may be described as questionnaires on which individuals report reactions or feelings in certain situations. The basic tenet of this method of personality assessment is individuals' self-observation and self-report. Personality inventories are not often called tests as there are no right or wrong answers. The term, inventory, is considered more adequate as they are paper-and-pencil questionnaires. A personality inventory may be designed to measure a single dimension of personality such as self-esteem, or several personality attributes simultaneously (Ehigie 1999).

Popular personality inventories include Minnesota Multiphasic Personality Inventory (MMPI), California Psychological Inventory (CPI), sixteen Personality Questionnaire (16PF) etc. The Minnesota Multiphasic Personality Inventory (MMPI), which is well described in the classic volumes by Dahlstorm, Welsh, and Dahlstorm (1972, 1975). The MMPI was developed by means of criterion-keying approach. The criterion groups for the development of most of the scales of the MMPI consisted of patients in the neuropsychiatric sections of Minnesota hospitals in the United States. All of these patients were classified in one of the following eight diagnostic categories of hypochondriasis, depression, hysteria, psychopathic deviate, paranoia, Psychasthenia, schizophrenia, and hypomania. Items which distinguished between members of these psychodiagnostic groups and normal control groups were included in the questionnaire (Huysamen 1982).

The MMPI represents an important milestone in the development of objective personality measurement and served as the basis for the development of other questionnaires such as the California Psychological Inventory (CPI) developed and revised by Harrison Gough (1987). What the MMPI means for the assessment of pathological behaviour, the CPI is said to do for normal behaviour. The MMPI was originally designed to identify people with serious personality disorders but it has presently been widely used in studying normal population. It is therefore used in determining the appropriate psychiatric label for people whose behaviours are not perfectly disordered. While drawing about half of its items from the MMPI, the CPI was developed specifically for use with normal populations from 13 up. On the whole however, the CPI is one of the best personality inventories currently available.

The Sixteen Personality Factor Questionnaire (16PF) is a multiple-choice personality questionnaire which was developed by Cattell (1946). In developing the Sixteen Personality Factor Questionnaire, Cattell used factor-analytic methods to obtain to obtain collections of homogeneous items. First he obtained a large collection of items by listing the names of all the personality traits which could be found in dictionaries, or in the psychological and psychiatric literature. The item scores were correlated and subjected to factor analysis, the result was the Sixteen Personality Factor Questionnaire (16PF).



***Achievement tests***

Achievement tests are designed to assess what a person has learned after a specific course of instruction. They try to measure performance of an individual after an exposure to a prescribed content. Standardised achievement tests were designed to evaluate a student's knowledge after a standard course of training. They represent terminal evaluation of performance and thus reveal how much a student has learnt. Course contents and syllabuses serve as a standardization measure for achievement tests. Downing (2002a) advocated that the multiple choice format is appropriate in achievement testing; he posited that the multiple test format can be used to test all levels of the cognitive taxonomy.

***Intelligence tests***

According to Encyclopaedia Britannica (2010), intelligence tests are series of tasks designed to measure the capacity to make abstractions, to learn, and to deal with novel situations. The most widely used intelligence tests include the Stanford-Binet Intelligence Scale and the Wechsler scales. The Stanford-Binet is the American adaptation of the original French Binet-Simon intelligence test; it was first introduced in 1916 by Lewis Terman, a psychologist at Stanford University. The individually administered test, revised in 1937, 1960, and 1972, evaluates persons two years of age and older and is designed for use primarily with children. It consists of an age-graded series of problems whose solution involves arithmetical, memory, and vocabulary skills.

Intelligence tests have provoked a great deal of controversy about what kinds of mental abilities constitute intelligence and whether the IQ adequately represents these abilities, with debate focusing on cultural bias in test construction and standardization procedures. Critics have charged that intelligence tests favour groups from more affluent backgrounds and discriminate against less privileged racial, ethnic, or social groups. Consequently, psychologists have attempted to develop culture-free tests that would more accurately reflect an individual's native ability.

These tests are no doubt useful and to very large extent objective measures of proposed constructs but the generalisability are questionable. Some of the items of these tests are designed for American and western societies. Ethnic and cultural differences were not considered and this makes generalisability somewhat difficult. Some items on the tests may not make sense to Nigerians or other non western citizens because of cultural differences, thereby making it difficult to respond to such items and ultimately influencing the test results. Consequent upon cultural implication of the tests, certain societies have adapted some of these western tests to suit their culture. An example is the PHSF (Personal, Home, Social, and Formal relations) Questionnaire published in 1970 by HSRC was designed to measure the adjustment of white high

school pupils, university students, and adults. This questionnaire was adapted and normed for high school pupils between the ages of 14 and 19 inclusive (Huysamen 1982).

#### Challenges of developing psychological tests in Nigeria

Psychologists in Nigeria have started constructing tests to assess personality in various settings. This became necessary to ensure cultural relevance of personality tests in Nigeria (Ekore 2001). Some procedures used by the western societies are being used to develop local tests in Nigeria. The strong point of these tests is the cultural consideration for test content and context. The deficiency in the foreign tests is necessitating psychologists in Nigeria to delve into test construction in order to meet local demand. However there exists still a need for local tests in Nigeria (Ekore 2001). The bulk of the tests available to psychologists are foreign and not suitable for African cultural background. The scarcity of local tests in Nigeria may be as a result of the challenges psychologists in Nigeria face in the process of test construction. Some of the challenges may include:

##### *a. Shortage of manpower in the area of psychometrics in Nigeria*

Most psychological tests in Nigeria were developed by non psychometricians due to the scarcity of psychometricians in the country. Downing & Haladyna (1997) argued that knowing the principles of effective item writing is no guarantee of an item writer's ability to actually produce effective test questions. Although item writers must be experts in their own disciplines, there is no reason to believe that their subject matter expertise generalizes to effective item writing. This explains the need for psychologists to be trained as psychometricians in order to ensure effective test construction. It is interesting to know that the study of psychometrics at the postgraduate level in Nigerian universities is almost nonexistent.

##### *b. Financial and time constraints*

Test development involves a lot of financial resources which are not readily available to most psychologists in Nigeria. As a result of financial constraint in test development, many of the psychological tests found in Nigeria are adapted from western countries. Time is a major constraint to test development in Nigeria. The basic principles and steps of test construction should be strictly adhered to in order to develop effective tests. These steps are time consuming and require a lot of efforts and commitment. This explains some of the challenges psychologists in Nigeria encounter in the process of test construction.

##### *c. Work overload*

Characteristic of the job of an average psychologist in the academic milieu is work overload. Lecturers often teach more courses than necessary and are also saddled with administrative responsibilities. It is also important that these lecturers write and publish papers and articles to remain relevant in the profession. There are multiple deadlines to be met as a lecturer in the



discharge of duties and the ability to multitask is crucial in this setting. The challenge of the psychologists here will be to meet deadlines and there is hardly room for extra engagements. Test development is not given any consideration as the psychologists already have all of their work hours occupied.

*d. Lack of mentoring and training programs for new test developers in Nigeria*

As a result of scarcity of psychometricians, young psychologists lack mentors to coach them in the art and science of test construction. According to Haladyna (2004) for new item writers is often helpful and important to provide specific instruction using an item writer's guide, paired with hands on training workshop. Effective item writing skills must be learned and practiced.

*e. Lack of attraction to psychometrics as a field in psychology*

Lack of attraction to psychometrics as a field in psychology is another reason for the challenges psychologists in Nigeria encounter in the development of local tests. As a result of 'psychometrics' involvement in the theory and technique of educational and psychological measurement including the measurement of knowledge, abilities, attitudes, and personality traits, many young psychologists avoid it in an attempt to escape figures. A good number of students in Nigeria have phobia for mathematics and therefore avoid it every way possible.

### **Conclusion and Recommendation**

Psychological tests serve a whole lot of purposes in both applied and research settings. Therefore, efforts should be made to ensure the construction of local tests in Nigeria. Universities in Nigeria should offer psychometrics as an area of specialization in psychology, as this will promote psychometrics and increase the number of psychometricians within the country.

Psychometricians in Nigeria should rise to this challenge and encourage young psychologists in the field through mentoring and coaching. Research clusters should be formed and financed towards the development of psychological tests.

The government, educational institutions, and professional bodies should give grants to psychologists in the pursuit of test construction; sabbatical leave should also be given to psychologists for the purpose of test development.

### **References**

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.

Anastasi, A. (1968). *Psychological testing* (3<sup>rd</sup> ed.) New York: Macmillan

- Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7<sup>th</sup> ed.). New Jersey: Prentice-Hall, Inc.
- Cattell, R.B. (1946). *The description and measurement of personality*. New York: World Book.
- Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.
- Dahlstrom, W.G., Welsh, G.S. & Dahlstorm (1972). An MMPI handbook vol.1. *Clinical interpretation*. Minneapolis: Minnesota University Press.
- Dahlstrom, W.G., Welsh, G.S. & Dahlstorm (1975). An MMPI handbook vol.2. *Clinical interpretation*. Minneapolis: Minnesota University Press.
- Downing, S.M. (2002a). *Assessment of knowledge with written test forms*. In Norman, G.R., Vander Vuleuten, C.P.M., Newble, D.I. (Eds). *International handbook for research in medical education* (pp 467-672). Dordrecht. The Netherlands: Kluwer academic publishers.
- Downing, S.M. (2002b). Construct-relevant variance and flawed test questions: Do multiple choice item writing principles make any difference? *Academic Medicine*, 77(10), 103-104
- Downing, S.M. (2004). *The effects of violating standard item writing principles: The impact of flawed test items on classroom achievement tests and students*. Paper presentation at the annual meeting of the American Educational Research Association, San Diego, C.A
- Downing, S.M., & Haladyna, T.M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement Education*, 10(1), 61-82
- Ehigie, B.O. (1999). *Psychological tests and testing. Psychology: Perspective in Human Behaviour*, pp 89-121



- Ekore, J.O. (2001). *Total quality management implementation: employees' needs, organizational type, and design as predictors*. An unpublished Doctorial thesis, University of Ibadan, Nigeria.
- Encyclopædia Britannica (2010). *Intelligence test*. Encyclopædia Britannica Online. Retrieved 23, November, 2010:  
<http://www.britannica.com/EBchecked/topic/289811/intelligence-test>.
- Fernald LD (2008). *Psychology: Six perspectives* (pp. 12–15). Thousand Oaks, CA: Sage Publications.
- Gray, P.O. (2010). Foundations for the study of psychology. *Psychology* (6th ed.). New York, New York: Worth Publishers.
- Gough, H.G. (1987). *California psychological inventory administrator's guide*. Palo Alto, CA: Consulting Psychologists Press, Inc.
- Haladyna, T.M., (2004). *Developing and validating multiple choice test items*. (3<sup>rd</sup> Ed.). Hillsdale, NJ, Lawrence Erlbaum Associates.
- Haladyna, T.M., & Downing, S.M. (1989a). *A taxonomy of multiple-choice item writing rules*. *Applied measurement in Education*, 1, 37-50
- Haladyna, T.M., & Downing, S.M. (1989b). *A validity of a taxonomy of multiple-choice item writing rules*. *Applied measurement in Education*, 1, 51-78
- Haladyna, T.M., Downing, S.M. & Rodriguez, M.C. (2002). *A review of multiple-choice item writing guidelines for classroom assessment*. *Applied measurement in Education*, 15 (3) 309-334
- Huysamen, G.H (1982). *Psychological measurement: An introduction with South African examples*.
- Jozefowicz, R.F., Koeppen, B.M, Case, S., Galbraith, R., Swanson, D., & Glew, H. (2002). *The quality of in house medical school examinations*. *Academic medicine*, 77, 156-161
- Rodriguez, M.C. (2003). *Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations*. *Journal of Educational Measurement*, 40(2), 163-184