Chike Okoli Center for Entrepreneurial Studies, UNIZIK - Deputy Director (2010 - 2012).

From the SelectedWorks of Prof. Uche J. Obidiegwu

Fall 2005

CONSTRUCTION OF STANDARDIZED ACHIEVEMENT TEST FOR

Dr. Uche J. Obidiegwu

Available at: https://works.bepress.com/druche_obidiegwu/5/
CONSTRUCTION OF STANDARDIZED ACHIEVEMENT TEST FOR ADULT LEARNERS IN FORMAL EDUCATION: A NECESSITY FOR ADULT LEARNING

Uche J. Obidiegwu, Ph.D
Department of Adult Education
Nnamdi Azikiwe University, Awka

Abstract
This study focused on the necessity of constructing and using standardized achievement instrument for assessing adult learners. Tests and other procedures for measuring learners’ progress serve as basis for instructional decisions taken by educators on learners. For this reasons, educators are supposed to be versed on the production of good quality tests in order to obtain true estimates of learners’ achievement. This study discussed the qualities which a good test should have namely; validity, reliability and usability. It explained the importance of ensuring that items which have desired psychometric characteristics (difficulty, discrimination and distractor indices) are included in an instrument. The study discussed the need, for constructing standardized achievement test for adult learners. It finally recommended that regular in-service training, seminars and workshops should be organized for educators on test construction and item writing rules.

Introduction
In any educational programme for adult learners, the aim is to bring about desirable changes in their behaviour. In order to determine whether a positive change is achieved the objective has to be stated in such a way that its attainment or lack of it can be observed and measured (Okoye, 1996). After objectives have been stated and instructions given, there is need to determine whether the stated objectives have been attained. The interest of an educator and all his activities in a learning situation is based essentially on the possible ways of improving learners’ achievement (Okeke, 2004). An educator does this by facilitating, learning and administering test on the learners. Test and test results are very vital in education. An educator may want to find out how effective his teaching was; which learners have poor understanding, which learners requires counseling, special or remedial classes; the type of learning difficulties learners are experiencing; to what extent learners are ready for the next learning experiences; how learners should be grouped for more effective learning. These are some of the questions which guide educators in taking instructional decision on learners. The above questions can be answered and decisions taken by an educator through administering different types of tests which may include achievement tests, diagnostic tests, self
ratings, periodic quizzes, pretest, scholastic aptitude tests, mastery test etc. These tests are important for diagnostic and prescriptive teaching of adults who may be slow learners, under achievers, gifted and bright learners. Test provides a comprehensive, systematic and objective evidence on which decisions are taken about learners. For this reason, every person who is interested in what happens in the education system should express some concern about production of good quality tests (Okoye, 1996).

In order to realize the goals and philosophies of education, the National Policy on Education, 2004 emphasized continuous assessment for all categories of learners. Against the background of importance of tests for instructional decisions, educators need to have knowledge of the attributes of standardized achievement test and the need to use such instrument in assessing their learners.

**Achievement Test**

Ali, Ezeadi, and Ogbazi (1988) defined a test as an instrument administered to someone or something to determine the presence and absence of a phenomenon being measured. They stated that a test must, satisfy two important conditions namely standardization and numerical indication. Standardization means that all procedures are the same for all candidates and stated before a test is taken. This ensures that all candidates are tested in the same way and with the same questions or problems. Numerical indication means that all scripts should be assigned with scores in the form of numbers such as 0,12,20,etc. Okoye (1996) defined a test as a set of questions, tasks, or statements that can be presented to an individual, response to which would enable the tester establish how much of desired characteristics are possessed by the testee. Baker (2003) defined achievement test as a test used to measure proficiency, mastery and understanding of general and specific areas of knowledge. Achievement tests are used to measure the extent of academic knowledge, skill and competencies which learners must have acquired as a result of exposure, to certain leaning experiences and contents (Onyekuba, 2004). If an adult learner is given a test on mathematics, it is aimed at determining how much of mathematics he knows. If he is given a test on economics, the intention is to find out also how much economics he knows. In the words of Nwana (1981), achievement test is any test given to determine how much learners have learned. To this extent, achievement tests dwell on well specified content areas (Onweh, 1993). Achievement test is designed to indicate degree of success in some past learning activity (Gronlund and Linn 1990). This means that achievement test when effectively constructed assists a teacher to ascertain and diagnose learners achievement level and performance in a course of study.

Achievement tests constructed without following laid down procedures adopted by measurement experts are known as teacher-made tests, while those constructed following standard procedure by measurement experts are named standardized achievement tests. However, Dosumu (2002), Alele-William (2002), Baker (2003) and Obioma (1985), respectively noted that most teacher-made tests lack validity and reliability. In their various explanations they indicated that most teachers seem to lack tests construction skills. In constructing achievement tests, however, there is need for educators to follow laid down procedures adopted by measurement experts by carrying out pre-test of the instrument, subjecting the items to item analysis i.e. determining the difficulty, discrimination and distractor indices of the items and giving specific
instruction on, administration and scoring of the test (Gronlund and Linn, 1990). Essentially, the primary characteristics which any psychological or educational measuring instrument should possess are validity, reliability and usability (Gronlund, 1990). Standardized tests, if well constructed, provides such attributes of an instrument.

**Validity of a Test**

According to Okoye, (1996) the primary characteristics which any psychological or educational measuring instrument should possess are validity and reliability. Validity is the degree to which a test measures what it is supposed to measure and consequently permits appropriate interpretation of scores (Gay, 1996). Validity as defined by Kerlinger, (1979) and Ali (1996) is the ability of a test to measure what it is supposed to measure. The validity of a test is the expression of the degree to which the test measures the qualities, abilities, skills and information which it is designed to measure (Green and Jorgenson, 1988). These definitions essentially reveal that the validity of a test shows the extent to which test items actually cover what they are designed to measure. Ali (1996) remarked that a test which is not valid is worthless. The validity of a test however, depends on the purpose of the test. This implies that a test which is valid for one purpose may not be valid for another, for instance a test which is valid for assessing achievement in physics will not necessarily be valid for assessing achievement in chemistry. Gronlund and Linn (1990) noted that validity pertains to the result of a test, or evaluation instrument and not to the instrument itself. There are many methods in which validity of a test can be obtained. They are face validity, criterion-related validity and predictive validity (Okoye, 1996). For the purpose of this work, the content validity which is the most appropriate form of validity for achievement test will be discussed.

**Content Validity**

The content validity of any given test refers to the extent to which a test measures both the subject matter content and the instructional objectives designed for a given course (Gronlund and Linn 1985). It is the most appropriate form of validity for achievement test (Tyler and Walsh, 1979; Lindvall and Nitko, 1975; Gay, 1996, Okoye 1996 and Ugwuije, 1996). Content validity involves adequate coverage of test items, its content and objectives as specified in the syllabus. It concerns a systematic examination of the test content to determine if the test items represent the behaviour base and subject matter (Okoye, 1996; Tyler and Walsh, 1979). According to Nwana (1979) to ensure sound construction of a valid achievement test, questions should be set from all parts of the syllabus, also, the number of questions set in each section must reflect the relative importance of each section. For these to be achieved, a test developer needs to prepare and be guided by a test blue-print or table of specifications. A test blue-print is a two-way grid table which specifies the level of objectives as they relate to the content of a subject.

**Reliability of a Test**

Next to validity, reliability is the most important characteristics of evaluation result. Reliability provides the consistency that makes validity possible and indicates how much confidence we can place on results (Gronlund and Linn 1990). Reliability is
the consistency of scores obtained by the same individuals on different occasions or with
Reliability refers to the consistency of measurement, that is, how consistent test scores or
other evaluation results are from one measurement to another (Anastasi, 1961). When a
test is reliable, it is expected that such test should give consistent information about the
person whose behaviour is being measured. It is expected that using a reliable instrument,
if we measure a trait in one person a number of times, we would continue to obtain the
same value.
Different methods of estimating reliability include test-retest, equivalent forms, test-
retest with equivalent forms, split-half method and Kuder-Richardson method. However,
out of the five methods of estimating reliability, Kuder-Richardson methods is mostly
used for determining reliability of an achievement test. Kuder-Richardson provided a
number of formula which can be used to estimate the reliability coefficient of a test when
each of the items is scored dichotomously (Okoye, 1996).

**Usability of a Test**

Usability of a test refers to practicability of an evaluation procedure. In selecting
tests and other evaluation instruments, practical considerations cannot be neglected
(Gronlund and Linn, 1990). The instrument should be short to avoid boredom and
disinterest on the part of the testee (Green, 1975; Mkpa 1992). However, if we attempt to
cut down too much on the time allotted to testing, we may reduce drastically the
reliability of our scores (Grunlund and Linn, 1990). Usability is concerned with a wide
range of factors including: cost of testing, ease of administration, ease of scoring and ease
of interpretation and application.

**Item Analysis**

In addition to determining the validity, reliability and usability of a measuring instrument,
it is necessary in the process of developing such an instrument to ensure that items which
possess the desired psychometric characteristics are selected for inclusion in the final
instrument. The process of working at the various items of the draft instrument with a
view to ascertaining how good they are is known as item analysis (Okoye, 1996). Item
analysis is aimed at improving the validity and reliability of a test. Inomesia (1986)
pointed out that the aim of item analysis is to find out the effectiveness of a single item
within a test. Okoro (1991) stated that, item analysis of a test is concerned with analyzing
of responses of individual test items to ensure that each item is valid and is measuring
what the test as a whole measures. It deals with how hard each item is (Difficulty Index
Analysis), whether each item distinguishes between the poor and brilliant students
(Discrimination Index Analysis) and how all the options attract responses (Distractor
Index Analysis). For the purpose this work, Difficulty, Discrimination and Distraction
indices would be briefly discussed.

The item difficulty index is defined as the percentage or proportion of persons answering
each item correctly (Anastasi, 1976). It assists in ascertaining that suitable items are
included in the final version of a test. It also helps in arranging items in a test in an
approximate order of decreasing facility (Nworgu, 1992). Difficulty index is sometimes
referred to as easiness percentage. It is the percentage of the students tested who
responded correctly to an item for example, a difficulty index of 75% indicates that 75% of the students responded correctly to an item. The higher the difficulty index, the easier the question item is. Item difficulty index ‘D’ is computed with the formula
\[ D = \frac{U+L}{N} \]
Where U, is the number of candidates in the upper ability group who answered the item correctly.
‘L’ is the number of candidates in the lower ability group who answered the item correctly.
‘N’ is the total number of candidates in both the upper and lower ability groups (Gronlund and Linn, 1990).

The value of ‘D’ ranges from 1 to 0. A difficulty index of 1 or 100% indicates an item which every person scored correctly, that is, a perfectly easy item. Conversely, a difficulty index of zero indicates an item which no person scored correctly, i.e. a perfectly difficult item. An item difficulty index of .50 is considered moderate. However, a difficulty index of 30 to 70 was recommended by Anastasi and Ferguson (1981).

The item discrimination index is the difference between the proportion of candidates scoring an item correctly in the upper group and those scoring the item correctly in the lower group. It is the ability of a test item to discriminate between high ability candidates and low ability candidates. Item discrimination index ‘R’ is computed as follows: \[ R = \frac{U-L}{\frac{1}{2}N} \]
where U, L and N still retain their value in the item difficulty index discussed above. Item discrimination index ‘R’ ranges from -1.00 to 1.00. The negative values of ‘R’ indicates that low ability candidates performed better on that item than high ability candidates. An ‘R’ of positive value indicates that high ability candidates performed better than low ability candidates on an item. An ‘R’ of zero indicates that equal number of students in the high ability group and low ability group passed the item. Such an item cannot discriminate. When all candidates in the two groups missed the item, an item discrimination index of zero equally results.

Distractor index involves checking the distractors (options) to see if they are plausible enough to be retained in the item. For a distractor to be plausible, it must appeal to some candidates as the correct answer, and more to the low ability group than to high ability group. The distractor index ‘DI’ for each distractor is computed with the formula ‘DI’= \[ U-L/N \]
where: U, L and N still retain their value in the item difficulty index discussed above. Negative value of ‘DI’ indicates that the distractor is effective. While a Positive value of ‘DI’ indicates that the distractor is ineffective. A ‘DI’ of zero indicates that the distractor is dead, it does not confuse or distract any candidate (Gronlund and Linn, 1990).

**Need for Constructing Standardized Achievement Test in Adult Learning**

A lot of decision are taken about learners based on achievement tests. Standardized achievement testing provides a more sophisticated method of learners assessment because of its peculiar characteristics namely: the test items are of high technical quality, they are developed by educational and test specialist, tried out experimentally (pretested), selected on the basis of difficulty, discriminating power and
subjected to a clearly defined and rigid set of specifications (Gronlund and Linn, 1990). Standardized achievement tests enable adult learners to benefit from writing valid and reliable achievement test. It is useful for both formative and summative evaluation of learners. Using standardized instrument for formative evaluation makes the learners aware of factors that would be considered in testing them in their continuous assessment or examination. Such awareness reduces anxiety and tension of whether they will pass or not in their examination. In addition, it saves educators time which they may use for writing classroom tests (teacher made tests). Educators can use standardized instruments either in whole or in part to assess their learners since it has a wide content coverage.

A test which is well developed, interpreted correctly and applied effectively contributes to a more intelligent educational decision. Standardized achievement tests make it possible to compare learners in terms of their achievement in a subject area. Its content is derived from objectives common to many schools in a state or country. An educator could use standardized test to obtain necessary information of adult learners over the years and also to obtain necessary standing of learners with respect to a norm group. This type of information may serve as a useful feedback to the educators regarding the effectiveness of their instruction. The information enables an educator to identify learners who has problems in specific content areas. Remedial instruction could be planned in order to help them out of their problems. Standardized achievement test also enables curriculum developers to know how adequately the curriculum of adult learners is covered and the particular area that needs revision.

**Recommendation and Conclusion**

In order to be able to achieve the National Philosophy on Education, acquisition of test construction skills by educators becomes very necessary. Since assessment is done on continuous basis in Nigerian education system and assessment results used to take decisions on learners, it has to be taken seriously by all those concerned-teachers, learners, parents and curriculum planners. Based on the foregoing, it is recommended that regular in-service training, seminars and workshops be organized for educators on test construction and item writing rules so that they can be able to develop comprehensive, systematic and objective instruments which will enable them to obtain true estimates of students achievement.

**References**


Okoro, M.O. Programme evaluation in education. Obosi: Pacific


