

Construction of Valid and Reliable Test for Assessment of Students

DR. P.U. OSADEBE

Department of Guidance and Counselling, Delta State University, Abraka., P.O. Box 88 Obiaruku, Delta State
Nigeria

E-mail: drosadebeuzo@gmail.com

Abstract

The study was carried out to construct a valid and reliable test in Economics for secondary school students. Two research questions were drawn to guide the establishment of validity and reliability for the Economics Achievement Test (EAT). It is a multiple choice objective test of five options with 100 items. A sample of 1000 students was randomly drawn to determine the validity and reliability of the test. After item analysis, the result showed that the Economics achievement test has a high face and content validity. The test item validity was determined through Difficulty and discrimination indices. A difficult index or p-value of 0.5 for each item was considered after applying the formula of correction for guessing. The discrimination index was established through point biserial statistics for each item with a correction coefficient of at least 0.3. The test has a reliability coefficient of 0.95 established through the use of Kuder-Richardson formula 20. The test is valid and reliable for assessing students internally and prepare them for external examinations.

Keywords: Construction, Economics, Achievement Test and Education.

Introduction

Construction of valid and reliable test in various subject areas has not been given enough attention in the senior secondary schools in Nigeria. It is expected that the schools should have enough valid and reliable tests for assessing their students when they have covered the curriculum content areas as well as to prepare them for external examinations. It has been observed that most teachers are not good in constructing test in their various subject areas. (Osadebe, 2001 & 2012). As a result of this problem, the researcher constructed a valid and reliable multiple choice Economics Achievement Test for use in schools as well as to prepare the students for external examinations such as West African Examination Council (WACE) and National Examination Council (NECO). The constructed test will serve as a major contribution to the need of valid and reliable Economics Achievement Test in senior secondary schools.

One of the greatest problems in the senior secondary schools in Nigeria is that teachers after assuming to have covered the Senior Secondary Certificate Examination (SSCE) Economics curriculum would resort to assessing senior secondary three (SSS III) students with unreliable Economics Achievement Tests on SSCE curriculum often administered during school's mock examination. This is usually the last test for SS III students before they sit for the Senior Secondary Certificate Examination. Most teachers hurriedly copy questions from any past question paper to compose their summative achievement tests. As a result, teachers do not establish validity and reliability for such tests.

The unreliable Economics Achievement Test administered during the school's mock-SSCE, is often used to assess and prepare the students before their SSCE is conducted by West African Examination Council (WAEC) or National Examination Council (NECO). The use of poorly designed Economics Achievement Test is a major problem as it affects students' interest and achievement in Economics. It has already been pointed out that poorly designed tests could make the students lose interest in a particular subject (Osadebe, 2001; Ohuche & Akeju, 1988). Similarly, Onunkwo's (1998) observation that most examiners find it easier to construct test items in the lower cognitive levels (knowledge and comprehension) than the higher cognitive levels (application, analysis, synthesis and evaluation) points to the fact that most of these examiners who are teachers in junior and senior secondary schools have been constructing tests that are not highly valid. This is a problem. However, the construction of test items is an art that only few people seem to master (Nunnally, 1981; Osadebe, 2001). Therefore, there is the need for experts to construct enough valid and reliable tests for use in senior secondary schools.

It is not uncommon that senior secondary school students score poorly in SSCE Economics. They are often assessed with poorly prepared Achievement Tests. The content areas of their tests in Economics are not spread out to cover the content of Economics curriculum. Consequently, such tests limit students' scope of reading and this leads to poor achievement in Economics.

Students' good achievement in Economics at SSCE will depend to a large extent, coverage of the Economics curriculum as well as the use of valid and reliable Economics Achievement Tests to assess them.

In Nigeria, students are expected to study Economics as a subject in senior secondary school for three years or have covered the SSCE Economics curriculum before they could be assessed with the Economics

Achievement Test. The Federal Government of Nigeria (1981 & 2004) recommends three years for senior secondary school during which Economics is taught as a subject to students. Generally, Economics is defined as the science which studies human behaviour as a relationship between ends and scarce means, which have alternative uses. In their own view, it is the study of how society decides what, how and for whom, to produce (Begg, Fisher & Dornbusch, 1994). A good knowledge in Economics at secondary school will later help students in tertiary schools to study Economics, Banking, Finance, Accounting and other related courses. Most students choose Economics in the senior secondary school because of their interest, ability and its relevance to their future careers.

During the three years in the senior secondary school, the students are taught by Economics teachers to cover the SSCE curriculum. Students are examined at every level of the senior secondary school with teachers made achievement tests until the end of the third year when they are finally assessed by WACE or NECO.

The Economics Achievement Tests constructed in this study is a multiple choice Economics Achievement Test. The multiple choice item can measure at both the knowledge and understanding level (Gronlund, 1976; Osadebe, 2001). Moreover, the multiple choice has the advantage of covering the broad content of Economics curriculum and measure different outcomes of the cognitive (Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation).

The multiple choice Economics Achievement Test for this study should be administered to SS III students when they have covered the WAEC or NECO Economics syllabus. Therefore, it was as a result of the use of unreliable achievement test poorly designed by teachers, and the need to provide a more valid and reliable achievement test in senior secondary school, that the researcher conceived the idea to construct one for assessing students at SS III and prepares them for external examinations. The achievement test is very useful when students cover the SSCE Economics curriculum.

Research Questions

The following research questions guide the study;

1. What is the validity of the Economics Achievement Test?
2. What is the reliability of the Economics Achievement Test?

Literature Review

The construction of valid and reliable test demand a special knowledge. There are literature on the test construction. When a teacher constructs a test, it is said to be a teacher made test that is poorly prepared. Then when an expert constructs a valid and reliable test, it is called a standardized test. Be it as it may, a teacher can construct a test if well guided. The problem of teachers for constructing poor test is a major issue in education that requires special attention. However, teachers should consult an expert before using a test (Osadebe, 2012). It is even advisable for experts to produce tests for teachers to use. Thus, non-standardized and standardized tests could be produced by experts. The construction and standardization of test are special areas for those who have the background (Nunnally, 1981; Osadebe, 2001).

The researcher who is an expert in test construction and standardization decides to construct a valid and reliable Economics Achievement Test for teachers as an area of need. The test will only be given out when needed so as to avoid misuse. It should be used in assessing students' achievement after teaching and learning Economics particularly when the content areas have been covered. Therefore, the main purpose of the study was to construct a valid and reliable achievement test for teachers and others to use particularly when the content areas of senior secondary school ordinary level Economics curriculum have been covered. Specifically, the study focused on how to establish validity and reliability for the Economics Achievement Test.

The procedure used in constructing the test was similar to that of Gronlund (1985), Nunnally (1981), Okonkwo (1995) and Osadebe (2012). The procedures used include: planning, item writing, item analysis or trial testing, composition of items, test theory, reliability, printing and manual preparation. The teachers and other users are expected to administer the test and score, to be guided by the manual as supported by Ukwuije and Opara (2012).

The systematic planning of the test requires identifying the instructional and behavioural objectives to be measured; identifying the content areas for the test, deciding on the test format and table of specifications. The test format is multiple choice objective test items. The table of specifications helped to establish high content validity (Osadebe, 2013; Ukwuije & Opara, 2012). The face validity was established through expert judgment by inspection of the items and approval. This is similar to that of Okonkwo (1997), Osadebe and kpolovie(2008). The instructional objectives were derived from the content areas of ordinary level (O/L) Economics curriculum. The behavioural objectives were drawn from the cognitive domain of Bloom as stated in Gronlund (1985). The behavioural objectives or cognitive levels include knowledge, comprehension, application, analysis, synthesis and evaluation. The table of specifications was drawn vertically and horizontally to determine the content areas and behavioural objectives for item writings. 200 items were generated to select the 100

required for the study. The items that survived the item analysis were selected. It is advisable to double sample of items because some may not survive item analysis. It was better to have enough items after item analysis to select the most valid items. This supported Aiken (1979) and Osadebe (2012). The test theory used was three parameters item response theory. This help to determine the difficulty level, discriminatory index and check guessing.

The composition of test is in line with Osadebe (2012), recommended that an Achievement Test may be composed in ascending order (preparing items or questions according to how the topics were taught), descending (setting questions starting from the last topic), and randomization (randomly selecting questions from the pool of items). The method used was ascending order because the researcher considered achievement in order of teaching and learning.

The reliability of the test was considered. There are different methods of estimating reliability. These include: test-retest, parallel form, split-half, Kuder-Richardson 20/21 and Cronbach Alpha. The method appropriate for the study is Kuder-Richardson 20. This is because the Achievement Test is a multiple choice objective with dichotomized response (either pass = 1 or fail = 0). This is related to the approach on reliability by Uzunoz (2011), and Osadebe (2001).

As soon as a test is valid, reliable and composed into the required number of questions, the next stage is printing. The complete test is printed for use. This will be accompanied by the manual. The manual is necessary to guide the users of the test. At this stage the test is keep in a safe place for use. The aim of the research was to produce the Economics Achievement Test as an instrument for teachers and other users in assessing the achievement of students and prepare them for external examinations.

Method

The study is instrumentation in nature. It is instrumentation because it is a study aimed at the construction, validation and production of valid and reliable test for teachers and others to use in assessing students achievement in Economics. A sample of 1000 testees was used for validation. This applies to the validity, item analysis and reliability of the test. The Economics Achievement Test is a multiple choice objective test of 100 items..

The content validity was established through expert agreement determined through Kendall Coefficient of concordance. The test items were drawn in a blue print or table of specifications based on the content of senior secondary school certificate Economics curriculum and the educational objectives of Bloom taxonomy of knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. Preparing a table of specifications helped the test to obtain high content validity (Ukwuije, 1996). Again, ensuring the content validity of an Achievement Test by explicit plan is more appropriate (Nunnally, 1981; Osadebe, 2013). The test was composed from the table of specifications after item analysis. The test made up of 100 items is for 2 hours. Testees have spaces to indicate their names, sex, location, school-type and class on the answer sheet. There are five options (A, B, C, D, E) for each item: made up of one correct answer (key) and four wrong answers (distracters). The distracters are plausible and each was randomly distributed. Generally, the stems or questions were well constructed. A p value or difficulty index of 0.5 for each item was considered after correction for guessing formula was applied. A point biserial correlation of at least 0.3 was considered for each item to determine the discrimination level. This helped to determine how the test discriminates between those who did well and those who did poorly on each item. During the item analysis, items with inappropriate distracters were retouched or discarded. 100 items were eventually selected.

The estimate of reliability of the Economics Achievement Test was determined through Kuder-Richardson formula 20(K-R₂₀). Copies of the final test were administered to a randomly drawn sample of 1000 testees. Each either passes an item with (1) or fails with (0). The K-R₂₀ helped to establish the internal consistency of the Economics Achievement Test.

Results

Research Question One: What is the validity of the Economics Achievement Test?

Table I: Computation of the Economics Achievement Test blue print or table of specifications for validity.

Content Area	Knowledge 30%	Comprehension 20%	Application 20%	Analysis 10%	Synthesis 10%	Evaluation 10%	Total 100%
Definition of Economics to Demand and Supply 30%	9	6	6	3	3	3	30
Theory of price determination to money and inflation 40%	12	8	8	4	4	4	40
Financial institutions to major natural resources 30%	9	6	6	3	3	3	30
Total	30	20	20	10	10	10	100

The table I shows a very good distribution of the Economics Achievement Test items. This was as result of the high degrees of agreement among the five judges on the percentage weight assigned to the objectives in the cognitive domain and content areas of senior secondary school Economics curriculum respectively. The degrees of agreement were high and significant with the use of Kendall co-efficient of concordance. An index of 0.86 for cognitive levels or objectives and 0.76 for content areas were obtained.

The result indicates that all the objectives and content areas were well covered. Thus, the agreement among the five judges and the distribution items to the objectives and content areas are enough evidence that the Economics Achievement Test has a high content validity. The face validity was also considered appropriate during the experts judgment.

Research Question Two: What is the reliability of the Economics Achievement Test?

Table II: Economics Achievement Test estimate of reliability using Kuder-Richardson formula 20.

$$(KR-20): rf = \frac{n}{n-1} \left[1 - \frac{\sum Pq}{SD^2} \right]$$

No. of students	No. of items (n)	$\sum pq$	\bar{X}	SD	SD ²	rf	SEM
1000	100	23.1	62.7	20.4	417.1	.95	4.6

The table II shows an estimate of the Economics Achievement Test (EAT) reliability using Kuder-Richardson formula 20. This approach became necessary because the EAT is a single multiple choice objective test with expected response of either pass (1) or fail (0). Thus, a reliability estimate of .95 was obtained. This helped to established the internal consistency of the EAT. The result reveals that the EAT has a high reliability and should be used. Since the EAT has 100 items, the high reliability seems to explain the fact that the longer a test, the higher the reliability coefficient. A standard error of measurement (SEM) of 4.6 was obtained. The SEM is low and this confirmed that the EAT is highly reliable.

Discussion

The table I result shows a very good distribution of the Economics Achievement Test items. This was as a result of the high degrees of agreement among the five judges on the percentage weight assigned to the objectives in the cognitive domain and on content areas of senior secondary school Economics respectively. The result indicates that all the objectives and content areas were well covered in the test blue print or table of specifications. This is an evidence that the Economics Achievement Test has a high content validity. The judges helped to established face validity. Therefore, the various analyses that helped to establish high content and face validity for EAT, are in line with current development that validity should be determined by a quantitative approach for more objectivity. This has been pointed out by Okonkwo (1997) and Osadebe (2013), that opinion of experts on content validation should be expressed on a clearly defined scale. This approach also supported the suggestion of Nunnally (1981), that validity should be expressed in index form. However, this quantitative approach is contrary to the age long qualitative approach where content validity only determined by experts' judgment or a rational affair which end in an expression of opinion of experts as supported by some authors (Guilford 1971, Anastasi, 1976; Thorndike & Hagen, 1977; Nunnally, 1981). Face validity is concerned with

judgment about the test after constructions (Nunnally, 1981). In their own view, the concept of face validity implies that a test should appear valid to a variety of judges in addition to being valid from a content point of view (Nevo, 1985; Zeidner, 1987). It is believed that when views are expressed by those recognized as experts on the issue of interest, they are more likely to acquire greater weights and enjoy more people's confidence than otherwise. It should be pointed out that experts sometimes do not agree on issues. Then if a scale is defined for the measurement of validity, it would ensure more objectivity. But some experts including Ebel (1979) have opposed attempts to determine validity indices of test. It should be noted that objectivity is a major concern in the measurement of test quality (Okonkwo, 1997). Hence, valid items after item analysis were selected for the Economics Achievement Test. This implies that the EAT has a high content validity. Thus, the use of difficulty index of at least 0.5 as recommended by Guilford (1954) and Joe (1995) as well as a point biserial correlation of at least 0.30 as recommended by Nunnally (1981) and Osadebe (2001) for each item, were all appropriate. Therefore, the quantitative approach used to determine how valid is the Economics Achievement Test helped to ensure high content and face validity for the test.

The reliability result was presented in table II, through Kuder-Richardson formula 20 analysis. The result shows that a reliability estimate of 0.95 was obtained. This help to establish the internal consistency of the Economics Achievement Test. The result reveals that the EAT is highly reliable and could be used for senior secondary three (SS III) students when they cover the SSCE Economics syllabus. The use of KR20 in this study was appropriate. Since this study is an individual research in the behavioural science and items are scored dichotomously then there is no excuse for not computing KR20 (Nunnally, 1918; Osadebe, 2013). The reliability estimate of this study (.95) is above the .87 of Rubenstein (1985), .77 of Obioma (1985), and .88 of Okonkwo (1995). However, the EAT reliability estimate measured up to the standard recommended by some experts. That is .90 and above by Ebel (1979), .70 to .98 of Guilford (1965) as well as .80 and above of Nunnally (1981). Further evidence of high reliability of the EAT is provided by a low value of the standard error of measurement (SEM = 4.6). This is in line with Thorndike (1971) who pointed out that a measure is very reliable as represented by a low standard error of measurement.

Conclusion and Recommendations

The study was carried out to construct and produce a valid and reliable Economics Achievement Test mainly for teachers and other users. The test should be used to assess senior secondary school III students' achievement in Economics when they have covered the content curriculum. Two research questions which centered on validity and reliability were drawn to guide the study. Theoretical and empirical literature related to the study were reviewed.

Various samples were used to ensure that the test is valid and reliable. A sample of five judges was used to determine the face and content validity. Then, a sample of 1000 testees was used to determine the reliability of the instrument. The validity of the instrument was determined through the use of test blue print or table of specifications and expert judgment. This helped to establish high face and content validity.

The reliability was estimated through the use of Kuder-Richardson formula 20. An index of 0.95 was obtained as a measure of internal consistency.

Hence, the constructed Economics Achievement Test has a high validity and reliability. The test should be used by teachers to assess students achievement when they have covered the content areas of senior secondary three Economics curriculum.

Since the Economics Achievement Test is highly valid and reliable, it was recommended that it should be used by teachers and other personnel to assess senior secondary three students' achievement when they have covered the content areas of Economics curriculum. The test should be kept in a safe place so that it will remain valid and reliable when used from time to time. The test should be used to prepare senior secondary school students for internal and external examinations. The test manual should always be consulted when using the test.

References

- Anastasi, A. (1976). *Psychological Testing*. New York: Macmillan Publishers
- Begg, D., Fisher, S., & Dornbusch, R. (1994). *Economics*. (4th Ed.) London: McGraw-Hill Publishing Company.
- Ebel, R.L. (1979). *Essentials of Educational Measurement* (3rd Ed.). New Jersey: Prentice – Hall Inc.
- Federal Government of Nigeria. (1981). *National Policy on Education*. Lagos. Federal Ministry of Information Printing Division.
- Federal Government of Nigeria (2004). *National Policy on Education*. Lagos: Nigeria Educational Research and Development Council Press.
- Gronlund, N.E. (1976). *Measurement and Evaluation in Teaching*. New York: Macmillan Publishing Co. Inc.
- Guilford, J.P. (1954). *Psychometric Methods*. New York: McGraw-Hill Book Company Ltd.
- Guilford, J.P. (1965). *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill Book Company.

- Guilford, J.P. (1971). *Psychometric Methods*. New Delhi: Tata McGraw-Hill Publishing Company Ltd.
- Joe, A.L. (1995). *Basic Concepts of Educational Measurement and Evaluation*. Port Harcourt: Paragraphics.
- Nevo, B. (1985). Face Validation Revisited. *Journal of Educational Measurement*, 22, 289-393.
- Nunally, J.C. (1981). *Psychometric Theory*. New Delhi: Tata McGraw-Hill Publishing Company Ltd.
- Obioma, G.O. (1985) *Development and Validation of a Diagnostic Mathematics Test for Junior Secondary Schools*. Unpublished Doctoral Dissertation, University Of Nigeria, Nsukka.
- Ohuche, R.O., & Akeju, S.A. (1988). *Measurement and Evaluation in Education*. Onitsha: African –Fep Publishers Ltd.
- Okomkwo, S.C. (1995). *Development and Standardization of a Readiness Test for Junior Secondary School Mathematics*. Unpublished Doctoral Dissertation University of Nigeria.
- Okonkwo, S.C (1997). Formative Evaluation Approach to Content Validation. *Journal of Technical and Science Education*, 6(1&2), 62-66.
- Onunkwo, G.I.N. (1998). Cognitive Hierarchies of the Junior Secondary Certificate Examination Questions in Integrated Science. *Review of Education*, Xv(1), 200-208.
- Osadebe, P.U. (2001). *Construction and Validation of Test*. A Seminar Paper Presented at the University Of Port Harcourt.
- Osadebe, P.U., & Kpolovie, P.J. (2008). Validation of Educational Research and Evaluation in Nigeria. *Journal of Vocational Science And Educational Development*. 8(1), 103-108.
- Osadebe, P.U (2012) *Procedures for Construction, Validation and Standardization of Test*. A Seminar Paper Presented at Delta State University, Abraka.
- Osadebe, P.U. (2013). Evaluation Techniques, DELSU, *Journal of Educational Research and Development*, 12(1), 56-63.
- Rubenstein, R. N. (1985). Computational Estimation and Related Mathematical Skills. *Journal of Research in Mathematics Education*, 16(2) 106-119.
- Thorndike, R.L. (1971). *Education Measurement Washington: American Council on Education*.
- Thorndike, R.L., & Hagen, E.P. (1977). *Measurement and Evaluation in Psychology and Education*. New York: John Wiley and Sons.
- Ukwuije, R.P.I & Opera, M.I. (2012). *Test and Measurement for Teachers*. Port Harcourt: Chadik.
- Uzunoz, A. (2011). Reliability and Validity Towards Environment Attitude Scale For Secondary School Students. *Journal of Educational Research and Review*, 6(16), 898-901.
- Zeidner, M. (1989). Essay Versus Multiple Choice Type Classroom Examinations: The Student's Perspective. *Journal of Educational Research*, 80, 352-358.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:

<http://www.iiste.org>

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: <http://www.iiste.org/journals/> All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: <http://www.iiste.org/book/>

Academic conference: <http://www.iiste.org/conference/upcoming-conferences-call-for-paper/>

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar

