DEVELOPMENT AND VALIDATION OF ECONOMICS TEACHER-MADE TEST FOR ASSESSING STUDENTS' ACHIEVEMENT IN THE NORTH CENTRAL STATES OF NIGERIA

Allahnana Kwanza Maikudi, Akande Martina Taiwo, Uwelo Danladi & I. J. Kukwi

Department of Educational Foundations, Nasarawa State University, Keffi, Nigeria

maikudiallahnana@gmail.com

+234-(0)8061257907

Abstract

This study developed and validated an Economics teacher-made test for assessing students' achievement in the North Central States of Nigeria. In terms of geographical scope, the study was restricted to Nasarawa, Benue and Plateau States of the North Central States of Nigeria. Instrumentation design type with cross-sectional survey research design was used and focused on senior secondary II curriculum. The population of this study comprises 200,530 students. Multi-stage random sampling technique was used to select a sample- of 400 respondents for the study. Economics Teacher-Made Test (ETEMAT) developed by the researchers was the instrument for the study. Kudder-Richardson (KR-20) statistic was used to establish the reliability of 0.89 (Internal consistency) of the instrument with validity index of 0.80. t-test for independentsample was used for the analysis Findings from the study showed that, the content validity index was found goodat 0.80 validity index and the ETEMAT was found to be highly reliable at 0.89 reliability index. It concluded thatthe EconomicsTeacher-Made Test is a valid evaluation instrument. The study recommends that,Economics teachers and teachers generally should develop valid and reliable Teacher-Made Test for assessing students' achievement.

Keywords: Development, validation, teacher-made test, achievement.

INTRODUCTION

In Nigeria, particularly in North Central States, Economics is considered as an important subject and is taught at the senior secondary school level. Economics occupies a very important position in the life of man and society. It is a subject concerned with the efficient utilization or management of limited productive resources for the purpose of attaining the maximum satisfaction of human wants (Ochuba, 2011). For Anyaele (2009), Economics is a social science which studies human behaviour as a relationship between ends and scarce means which have alternative uses. Given the foregoing definitions, there is need to add that Economics as a social science helps man to understand and manage his scarce resources in order to meet his numerous wants. Toachieve the goals or objectives of Economics at Secondary school level, the teaching and learning of Economics have to be properly done, especially in assessment practices of teachers, not just for those intending to pursue a career in Economics, but also, more generally, as a part of educational -foundation which every student should have before leaving school. The framework for assessment begins the same way curriculum design begins.

Teacher-Made tests measure knowledge of facts, concepts, principles, skills, interest and attitude. Teacher-Made tests are primarily used in making classroom-level decisions and are designed with particular reference to the course objectives/learning goals of a specific course, study program or class (Mahajan, 2015). Teacher-Made tests indicate present, not future, proficiency. Such tests evaluate students' understanding of a particular instructional domain in order to make decisions regarding the advancement or capability of the students.However, assessmentin education is the form of achievement test, serves the purposes of identifying the learners' extent of mastery of knowledge and skills, as well as contribute to effective teaching of a subject(Chime,2012). This implies that achievement test should be used for guiding the teaching process for enhanced learning to occur. Iwuji (1990) cited inChime(2012), defined an achievement test as an instrument given at the end of teaching-learning programme. Achievement test is used to find out how much a student is able to achieve in a course he/she has been taught. Nwagu (1992)seen achievement test as a systematic and purposeful quantification of learning outcomes. It involves the determination of the degree of attainment of individuals on tasks, courses or programmes to which the individuals were exposed. Often time, teachers ask questions before, during or after their lessons to ascertain how much information, issues and skills concerning the instructional theme the students have learnt. Teachers also organize tests weekly, termly or yearly to assess the students in terms of achievement

in the various content areas of instruction. The result of testing provides the teacher and the students with some feedback on the students' progress in the subject. This would enable a teacher to decide whether to carry on with lessons as planned certain areas for specific groups of students(Chime, 2012).

Inadequate valid Economics teacher-made tests according to Allen (2005) are a reason many teachers continue to assign invalid grades to students. If the grades are not accurate measures of the student's achievement, then they do not communicate the truth about the level of the student's academic achievement. Since important decisions are often based on a student's grade, invalid Economics teacher-made tests-might produce grades that may result in dire consequences for the student. If students receive grades lower than ones that accurately depict their true level of Economics academic achievement, it may lead them to believe that they lack the ability to succeed academically in Economics and lower their sense of self-efficacy as well as their motivation to do well in WAEC and NECO Economics examinations (Osadebe, 2010). So, valid items for evaluating students' achievement in Economics secondary school are rare and the possibility of developing such items by the classroom teacher is limited because it is an art that only experts in test development does. It involves a couple of steps scrupulous analysis, and substantial time (EsomonuandAgbonkpolo, 2010). Ali (2012) stated that the validity of a test is the degree of accuracy with which the test measures what it is intended to measure. Similarly, Onunkwo (2002) explained that validity of an instrument means the degree of qualities, abilities, skills, traits, information it was designed to measure. A valid test ensures that questions are set from all parts of the syllabus. This emphasizes the need to ensure adequate coverage of both subject matter area and the instructional objectives which the students' learning centred on. Eze (2011) explained that a table of specification is used to ensure a systematic coverage of the entire course content and instructional objectives.

Nwagu in Chime (2012), test reliability indicates the extent to which individual differences in scores are attributed to chance errors of measurement, and the extent to which they are attributable to true differences in the characteristics under consideration. However, there is non-existence of valid and reliable instrument which would yield dependable and authentic results of Economics. This calls for the need to develop and validate an instrument for authentic assessment in Economics.Development and validation of test items, refers to construction of a test items, and ensuring that the instrument (test) used measured what it was designed to measure (Nwana, 2008). The general trend in the development of teacher-made test has been-; the definition of the constructs and content to be measured, identification of the target population, item collection and preparation, pilot study, item review, main study, and data analysis with regard to test characteristics (Ritter, Boone &Rubba, 2001).

A valid and reliable test should have test characteristics that fall within the accepted range of values, for each characteristic, such as-; validity, reliability, discrimination index, difficulty index, and readability, and it should not be biased against any designated sub-group of test takers, such as gender and school location. These two factors, gender and schoollocation are differed on student achievement. Okereke (2011) found out that the development of the instrument yielded a logical validity index at 0.8, 0.71, 0.78, and 0.63. The finding also supports the work of Opara (2013) who found that the Mathematics Achievement Test is a valid and reliable instrument for measuring achievements in mathematics tests. The content validity index was found perfect. Based on the numbers obtained from 2 raters, the value of content validity was 0.80, this was carried out on a 4-point rating scale of very relevant, quite relevant, somehow relevant, and not relevant.The Achievement Test (MAT) was found to be highly reliable with three reliability indices of 0.73, 0.52, and 0.44. The Mathematics Achievement Test (MAT) was found to be of appropriate difficulty index and distracted positively.

Adonu (2009) found out that there was a significant gender related difference in the performance of male and female students in psychomotor tasks. This difference is in favour of the males.In another research development, Onah (2009) in his finding revealed that there was a significant difference in the mean achievement scores of urban school students and rural school students.(b) there was a rejection of null hypothesis on the influence of sex on student's achievement in ASAT in favour of males.Osadebe (2010) carried out his study on construct valid and reliable test in Economics for

secondary school students. The finding test has a reliability coefficient of 0.95 index. Moneth (2012) from the results of the analysis, it was found out that: the developed Economics achievement test instrument for Senior Secondary schools has high psychometric properties in terms of facility and discrimination index; the instrument has high reliability index; there was significant difference between the achievement of male and that of female students in Economics at the senior secondary school level in favour of the male students; there was a significant difference in mean achievement between students in schools located in urban areas and those in rural areas in favour of urban students. All the above studies were different from the present study. Some of the study based on gender issue with different instrument for assessing students' achievement. But the present study based on the development and validation of teacher made test for assessing students' achievement in North Central States of Nigeria. This formed the knowledge gap of the study.

Most Economics teachers have continued to develop instruments for measuring students' authentic assessment according to their varied abilities in test construction. Observations show that those teacher-developed testing instruments are generally of doubtful psychometric features since no serious attention might have been paid to their development and validation. For such instruments, either face validation or possibly content validation was employed. In other words, most of the Economics teachers in secondary schools do not seem to possess the competencies required in instrument development and validation. This means that for Economics teachers to use valid and reliable tests experts in test development have to develop them, otherwise the objectives of educational system may not be achieved. Consequently, there is need to develop and validate an instrument for measuring students' attainment through Teacher Made Test.On teacher-made test development practices generally, teachers have problems with-: proofreading the test, using a sufficient number of items, and examining student achievement on the items. It is against this background that this study stems up to develop and validate Economics teacher-made test for students' achievement in North Central States of Nigeria.

Objectives of the Study
The main objective of the study was to develop and validate an Economics Teacher-made test for assessing students' achievement in North Central States of Nigeria. Specifically, the study intends to:
1. determine the content validity index of Economics Teacher-Made test forassessing students' achievement.
2. establish the predictive validity of the developed Economics Teacher-Made test forassessing students' achievement.

Research Questions
The following research questions were raised to facilitate the investigation:
1. What is the content validity index of Teacher-Made test for assessing students' achievement in Economics?
2. What is the predictive validity index of the developed Economics Teacher-Madetest for assessingstudents'achievement?

Hypotheses
To facilitate the investigation, the following hypotheses were raised and tested at the 0.05 level of significance to ensure the development and validity of Economics Teacher-Made test forassessing students' achievement:
1. The content validity of the developed Economics Teacher-Made test will have logical validity index not less than 0.75 as appraised by experts.
2. The predictive validity index above 0.72 will predict Economics Teacher-Made test for assessing students' achievement.

Theoretical Framework of the Study
This study anchored on ItemResponse Theory (IRT):
Item Response Theory by Lord and Novick in Emaikwu (2005)
Item response theory (IRT) was first propounded by Lord and Novick in Emaikwu (2005) in the field of psychometrics for the purpose of ability assessment. It is widely used in education to calibrate and evaluate items in tests, questionnaires, and other instruments and to score subjects on their abilities, attitudes, or other latent traits. During the last several decades, educational assessment has used

more and more IRT-based techniques to develop tests. Today, all major educational tests, such as the Scholastic Aptitude Test (SAT) and Graduate Record Examination (GRE), are developed by using item response theory, because the methodology can significantly improve measurement accuracy and reliability while providing potentially significant reductions in assessment time and effort, especially via computerized adaptive testing. In recent years, IRT-based models have also become increasingly popular in health outcomes, quality-of-life research, and clinical research. For simplicity, models that are developed based on item response theory are referred to simply as IRT models.

Item response theory (IRT) is a collection of measurement models that attempt to explain the connection between observed item responses on a scale and an underlying construct. Specifically, IRT models are mathematical equations describing the association between subjects' levels on a latent variable and the probability of a particular response to an item, using a non-linear monotonic function. As in classical test theory, IRT requires that each item should be distinct from the others yet should be similar and consistent with them in reflecting all important respects of the underlying attribute or construct. Item parameters in IRT are estimated directly using logistic models instead of proportions (difficulty or threshold) and item-scale correlations (discrimination). There are a number of IRT models varying in the number of parameters (one, two and three-parameter models) and whether they handle dichotomous only or polychromous items more generally.

According to Demars (2008), IRT is also sometimes called latent trait theory. This is a modern test theory (as opposed to classical test theory). It is not the only modern test theory, but it is the most popular one and is currently an area of active research. IRT requires stronger assumptions than classical test theory. IRT is much intuitive approach to measurement once one gets used to it. In IRT, the true score is defined on the latent trait of interest rather than on the test, as is the case in classical test theory. IRT is popular because it provides a theoretical justification for doing lots of things that classical test theory does not. Dodeen and Darabi (2009) assert that some applications of IRT include:
Item bias analysis-IRT provides a test of item equivalence across groups. The theory in relevant to this study because it expressed the psychometric properties of teacher made test for assessing students' achievement where teacher should be considered when developing an instrument.

Methods
Instrumentation design type with cross-sectional survey research design were used and focused on senior secondary II Economics curriculum. The population of this study comprises 200,530 students. Multi-stage random sampling technique was used to select a sample of 400 respondents for the study. Economics Teacher-Made Test (ETEMAT) developed by the researchers was the instrument for the study. To ensure that the instrument measures what it is supposed to measure, content, face, convergent and divergent validity was ensured for the instrument by developing a test blueprint or table of specification based on the Benjamin Bloom's taxonomy of educational objectives in the cognitive domains (as cited in Anikweze, 2013) constituting of: knowledge, comprehension, application, analysis, synthesis and evaluation as presented in Table 1.These experts were requested to scrutinize the items (stems, options, keys and distracters) of the ETEMAT in terms of clarity, relevance, adequacy and comprehensiveness of the items. To guide the experts in the validation exercise, the topic of this study and table of specifications together with the draft test were given to the experts. After examining the test, they made some corrections on some of the items. Their expert observations, comments and suggestions were used in the modifications of the ETEMAT.

Table 1: Table of Specification for 100 Items Test in Economics for SS 2 Students

S/N	CONTENT AREA	Time (Hrs)	Know	Com	App	Ana	Syn	Eva	Total Items	%
1.	Demand; types, law, factors (Equilibrium)	2	**5**(,31, 34, 45, 85)	**5**(19,2 7, 67, 34, 82, 100)	**3(**, 56**)**	**5**(2, 14,, 56, 47, 54,68 , 95)	**5** (12, 15, 22)	**2**(3, 4, 5, 6, 11,7, 8, 9)	25	25
2.	Supply; types, law, factors (Equilibrium)	2	**5**(5, 9, 84, 85, 87)	**3** (8, 18, 15, 53, 10)	**5**(19,1 2,94, 96)	**5**(28, 16, 17,81 ,83)	**5**(3, 72, 73, 77)	**2** (13, 82,52)	25	25
3.	Elasticity of demand and supply with the theory of Cost	2	**5** (22, 24,28, 29, 30)	**5**(23, 16, 17)	**5**(26, 29, 13,)	**3**(18, 20,21 , 27)	**5**(4,6, 91)	**2** (22, 23,24, 25)	25	25
4	Industrialization, money and theory of cost	2	**3(**59, 58,60, 61)	**5**(88, 89, 90)	**5(**70,7 1,72,7 4,75,7 6)	**5(**50, 51)	**5**(53,54 ,55,78,7 9)	**2**(89,9 7,98,9 9)	25	25
	Total of items	**6**	**18**	**18**	**18**	**18**	**20**	**8**	**100**	**100 %**

KEY:

Know =Knowledge = Recall
Com = Comprehension = Understanding = U
App = Application
Ana = Analysis
Syn = Synthesis   Thinking = T
Eva = Evaluation

Kudder-Richardson (KR-20) formula was used to establish the reliability of 0.89 (Internal consistency) of the instrument with validity index of 0.80. For the purpose of this study, one instrument developed by the researchers was used for data collection. The instrument consists of 100 items multiple-choice test drawn from various Economics topics as recommended by the NERDCcurriculum for SS II. The instrument was developed by the researchers. Each test item in the instrument has five response options, namely A, B, C, D and E with only one option as the key while others are distracters. Instructional Assessment Resources (IAR, 2011) asserted that "an item analysis involves many statistics that can provide useful information for improving the quality and accuracy of multiple-choice or true/false (question)". The ETEMATwas administered to 400 SS2 Economics students during second term of 2018/2019 academic session when subject teachers and SS2 students had completed the teaching and learning of the test content. The test try-out was for the purpose of item analysis.
Item analysis: It is done to ensure the quality of the items. It involved the following steps:
Step 1: Identify the higher and lower achievers.
Step 2: Process test responses.
Step 3: Calculate item difficulty index.
Step 4: Calculate item discrimination index.
Step 5: Calculate the distracter indices.
Step 6: Selection of good items.
Eleje, Abanobi and Obasi (2017) viewed that an item was considered good for inclusion in the final output of the test if it had difficulty index of 0.30 to 0.70, discrimination index greater than 0.20 and a positive distracter index. However, items with appropriate difficulty indices but with discrimination indices of less than 0.20 were not accepted as good. Also items with appropriate discrimination index but have difficulty index of less than 0.20 or more than 0.80 were rejected. Factor analysis was use to answer research questions and establish norms for the ETEMAT while t-test was used to test the formulated hypotheses at 0.05 level of significance.

RESULTS

Question 1: What is the content validity index of Teacher-Made test forauthentic assessment of Economics developed by the researcher?

The content validity index (CVI) of Teacher-Made test forassessing students' achievement developed by the researcher was computed based on the joint ratings of relevance of Economics Teacher-Made Test (ETEMAT) items by three content experts.

Table 2: Showing Ratings of the Relevance of ETEMATItems, by ThreeContent Experts Rating on 100 Items

Experts	Content Validity Indices	Average of the index	Guess Indices	No of Items
1	0.80		0.067	
2	0.70	0.80	0.065	100
3	0.64		0.076	

This was carried out using a 5-point scale for the analysis. This implies that 80% of items which is equivalent to 80 items out of 100, as they were rated quite relevant and very relevant to the component objectives. Therefore, the content validity index of EconomicsTeacher-Made Test for authentic assessment of Economics was 0.80. This implies that Economics Teacher-Made Testwas valid during the exercise.The percentage scores from the validating scales were summed up and their means for the respective instruments translated into the logical validity indices otherwise called rational validity.
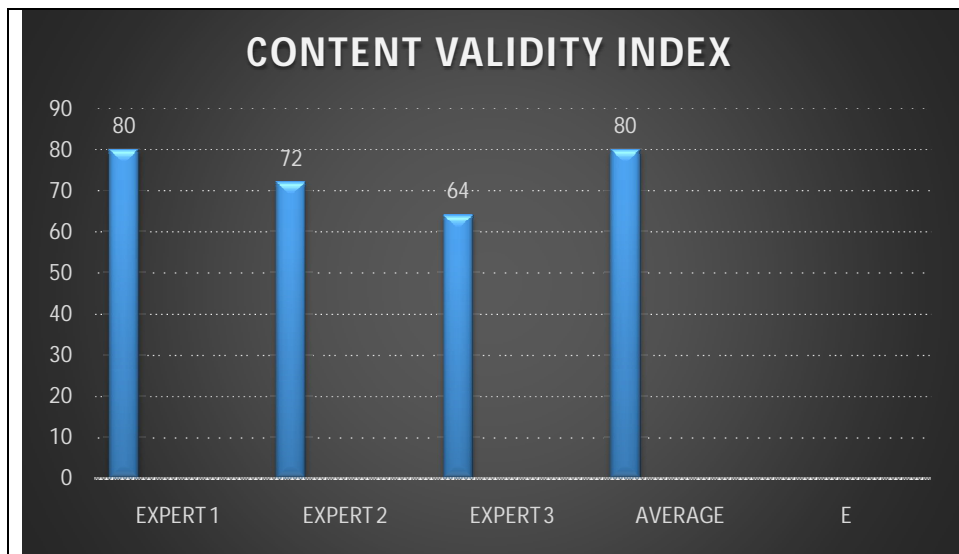

Figure 1: Simple Bar Chart Represent Content Validity Index (CVI) of Teacher-Made Test

Research Question 2: What is the predictive validity index of the developed Economics Teacher-Made test for assessingstudents' achievement?

Table 3: Predictive validity index of the developed EconomicsTeacher Made Test

Variables	Predictive Validity Index Statistics		
		Guess	
Predictive Validity index	Teacher Made Test Standardized	Items Index	N of Items
0.80	0.64	0.048	100

Table 3shows that the predictive validity index was 0.80. This implies that, predictive validity index of ETEMAT predicted the achievement of students' in Economics Teacher Made test.
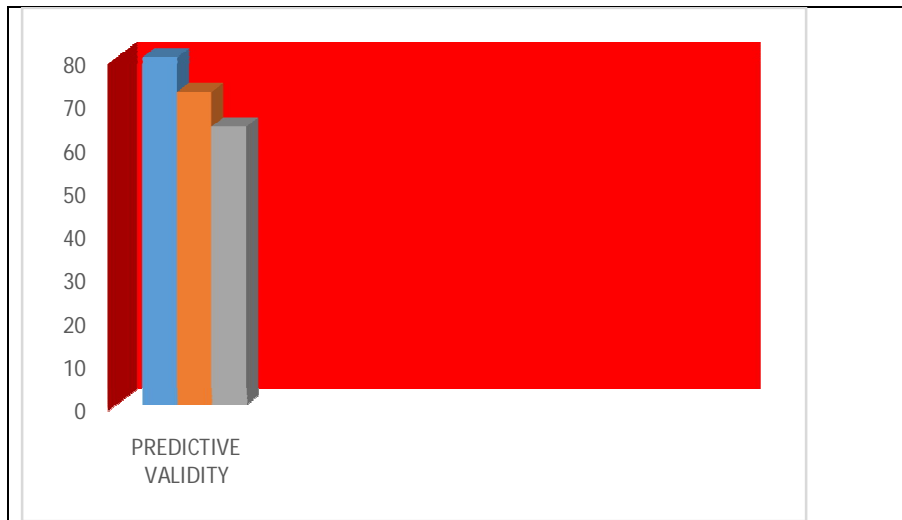
Figure 2: Simple Bar Char of the Predicted Validity index of ETEMAT
Testing of the Hypotheses
Hypothesis 1: The content validity of the developed Economics Teacher-Made test will have logical validity index not less than 0.75 as appraised by experts.
Table 4:Result of t-Test Statistics shows Content Validity of the Developed Economics Teacher-Made Test of Logical Validity Index as Appraised by Experts
t-test for independent samples

Variables	N	Mean	SD	df	Level of Significance	T	Sig. (2-tailed)
Content Validity	241	2.34	1.10				
				398	0.05	0.796	0.784
Achievement	159	2.11	1.08				

Analysis of the data using t-test for independent sample indicated means of (Content validity) and Authentic Assessment) with   398 degree of freedom at 0.05 level of significance. T-calculated was 0.796. This means the null hypothesis was rejected in both content validity and authentic assessment responses and alternative hypothesis was accepted which revealed that content validity of the developed Economics Teacher-Made test has logical validity index not less than 0.75 as appraised by experts. That is, the logical validity index of 0.80 was obtained by experts and as showed in Table 1. This shows that the Teacher-Made test used for this test research was valid for the study.

Hypothesis 2: There is no significance difference between predictive validity index of 0.72 and Economics Teacher-Made Test

Table 5: Result of t-test of Significant for the Predictive Validity of Teacher Made Economics Test

t-test for independent samples

Variables	N	Mean	SD	Level of Sign.	T	df	Sig. (2-tailed)
Predictive validity	238	2.65	1.45				
				0.05	0.593	398	0.567
Achievement	162	2.43	1.32				

Analysis of the data using t-test for independent sample indicated a predicative validity index of 0.82 with a degree of freedom of 398 at the 0.05 level of significance. This means the alternative hypothesis was accepted in both validity and null hypothesis was rejected which revealed that predictive validity index of 0.82 predicted Economics Teacher-Made Test for assessing students' achievement in Economics as shown in Table 2.

Discussion of Findings

The first finding of the study showed that the content validity index of the developed Economics Teacher-Made test has a logical validity index of 0.87. This finding agreed with the work of Okereke (2011) who found out that the development of the instrument yielded a logical validity index at 0.8, 0.71, 0.78, and 0.63. The finding also supports the work of Opara (2013) who found that the Mathematics Achievement Test is a valid and reliable instrument for measuring achievements in mathematics tests. The content validity index was found perfect. Based on the numbers obtained from 2 raters, the value of content validity was 0.80. The Achievement Test (MAT) was found to be highly reliable with three statistics of 0.73, 0.52, and 0.44. The Mathematics Achievement Test (MAT) was found to be of appropriate difficulty index and distracted positively. The Economics teacher made test items are suitable test items. The suitability of the items was a function of the average rating score accorded each item of the appropriateness of each item. Second finding of the study revealed that predictive validity index of 0.82 predicted Economics Teacher-Made test on authentic assessment of students' achievement in Economics. This finding agreed with the work of Osadebe (2010) who found that validity of 0.78 predict students internally and prepares them for external examinations. The finding also supports the work of Adonu (2009) who found that there was no significant difference among theaters in their rating of the student's psychomotor skills on the instrument. This also agreed with Moneth (2012) who found out that there was significant difference between the achievement of male and female students in Economics at the senior secondary school level in favour of the male students; there was a significant difference in mean achievement between students in schools located in urban areas and those in rural areas in favour of urban students.

Conclusion

Most Economics teachers in secondary schools do not seem to possess the competencies required in instrument development and validation. This means that for Economics teachers to use valid and reliable tests experts in test development have to develop them, otherwise the objectives of educational system may not be achieved. In view of this, there is need to develop and validate an instrument for measuring students' attainment through Teacher Made Test. TheEconomics teacher made Achievement Test (ETEMAT) is a valid evaluation instrument. The content validity index was found perfect. Based on the numbers obtained from 3 raters, the value of content validity index was 0.80. The ETEMATwas found to be highly reliable.ETEMAT was found to be of appropriate difficulty index and distracted positively. The ETEMATtest items are suitable test items. The suitability of the items was a function of the average rating score accorded each item of the appropriateness of each item.

Recommendations
Based on the findings of the study, the following recommendations were made;
1. Economics teachers and teachers generally should develop teacher made test based on the content for assessing students' achievement.
2. Economics teachers and teachers generally should use the test to predict students that will do well in Economics in their final class as well as those that will have good performance in external examinations.

REFERENCES
Adonu, C. (2009). *Development and Standardization of a Readiness Test for Junior Secondary School Mathematics.*Unpublished Doctoral Dissertation University of Nigeria.

Ali, A. (2012). *Conducting research in education and social sciences.* Enugu:Tashiwa network limited.

Allen, W. (2005). Introduction to test theory and development process. Port-Harcourt: *Chris-Ron integrated services.*

Anikweze, C.M. (2013). *Measurement and evaluation for teacher education.3^{rd}Edition.*IBADAN: Constellation Publishers (NIG) Limited. ISBN*:978-049-498-7. Pg.105-107.*

Anyaele,B.(2009). *Principles of tests construction and administration.* Lagos: *Publishing Consultant.*

Chime, U. M. (2012). Development and Validation of Economics Achievement test for senior secondary school students. *Master's thesis, Department of Science Education, Univer sity of Nigeria, Nsukka.*

Demars, T. (2008).Online social interchange, discord and knowledge construction.*Canadian Association for Distance Education. Journal of Distance Education,*13(1), 57-74. development.

Dodeem, H. &Darabi, T. (2009). Using Constructivism in Technology-Mediated learning; Constructing Order out of the Chaos in the Literature.*Radical Pedagogy,*1(2) Retrieved on1/9/2018 http://radicalpedagogy.icaap.org/content/issue1_2/02ka nuka1_2.html.*IJEDICT*

Eleje, L.I, Abanobi, C.C& Obasi, E (2017). Development and Validation of Economics Achievement Test for Secondary Schools .*Asian Journal of Education and Training.*

Emaikwu, S.O. (2015). The development of Self- and Peer-assessment Strategies for a Design and Project-
based Curriculum. UltibaseArticles.[Accessed online] http://it.coe.uga.edu/itforum/paper63/pa per63.htm.01/10/2018.

Esomonu, C. &Agbonkpolo, A. (2010). Impact of parents' socio-economic status on university students' academic performance.*Ife Journal of. Educational. Studies.* 7 (1): 31-39.

Eze, L. (2011). Perspectives on the evolution and future of educational measurement. Westport, CT:*American Councilon Education/Praeger.*

Iwuji, V.B.C. (1990), M*easurement and evaluation for effective teaching and learning,* Onitsha: Summer Educational Publishers (Nig.) Limited.

Mahajan, A. (2015). Measurement and Evaluation in Education. Onitsha, Nigeria: *Africana-FEP Publishers Limited.*

Moneth, I. (2012). Procedures for Construction, Validation and Standardization of Test: Motivation based traditional method: its effects on female gender 'schools' performance in chemistry. 42^{nd}annual conference proceedings of science teachers' association of Nigeria.

Nwana, O.C. (2008). Educational measurement for teacher.Lagos: *Thomas Nelson and Sons (Nig.) Limited.*

Nwagu E.K.N. (1992), *Development and standardization of social sciences achievement test (SSATS),* Unpublished Ph.D Thesis, Department of Education, University of Nigeria Nsukka.

Ochuba, J.(2011). Development and Validation of an Achievement Test in Mathematics, *International Journal of Mathematics and Statistics Invention(IJMSI).*

Okeke, E. (2011). Towards gender equity in the 21stcentury. In C.V. Nnaka and Okereke, S. C. (2008). Development and preliminary validation of an instrument for the identification of mathematically gifted pupils in Ebonyi State. (Unpublished Ph.D Thesis), University of Nigeria Nsukka.

Onah, J. (2009). Formative Evaluation Approach to Content Validation. *Journal of Technical and Science Education, 6(1&2).*

Onunkwo, P. S. (2002). *Measurement, Assessment and Evaluation.* Lagos: Concepts Publication Limited.

Opara, M.I. (2013). Computational Estimation and Related Mathematical Skills. Journal of Research in Mathematics Education,16(2) 106-119.

Osadebe, P.U. (2010). Validation of educational research and evaluation in Nigeria. *Journal of Vocational Science and Educational Development. 8(1), 103-108.*

Ritter, T., Boone, U. &Rubba, M. (2001). Effect of Mathematical Games on Academic Achievement of Senior Secondary School Students in Mathematics in Abuja, Nigeria. *ABACUS: The Journal of Mathematical Association of Nigeria.* 42(1), 40-47. Routledge.

Rubrics, H. (2011).Gender differences in Mathematics interest and performance.   University   of Michigan.