Content Validity and Reliability Indices of National Examination Council Mathematics for Basic Examination in Adamawa State, Nigeria.

Amuche, C.I. Ph.D, Obiji Donatus Ph.D, Winstone Julius & Samson Edison

Department of Educational Foundations, Taraba State University, Jalingo

Abstract

This study examined the content validity and reliability indexes of national examination council certificate mathematics for basic certificate examination in Adamawa State, Nigeria. To achieve the purpose of the study, two null hypotheses were formulated to direct the study. Literature review was done according to the variables under study. Evaluation research design was adopted for the study. A sample of 556 JSS III students was randomly selected from 12 Local Government Areas of the state. Two research instruments were used for the data collection. The instruments were Basic Education Mathematics Achievement Tests (BEMAT) for the year 2018, 2019 and 2020 and a content validity checklist. Descriptive statistics was used to present the research questions. The null hypotheses were tested using analysis of co-variance (ANCOVA) and Kruskal-Wallis H test. The result of the analysis revealed that, NECO Mathematics BECE question has moderate content validity and reliability. Based on the findings of the study, it was recommended that examination bodies such as NECO should not include items with low validity and reliability indexes. Experts in educational measurement and evaluation should carry out the rigorous process of item analysis by determining item validity and item reliability indexes and merely not by hand, picking of items from past questions.

Keywords: Basic certificate, content validity, mathematics and reliability indexes

Introduction

Examination is an instrument to determine examinees extent of competence in a particular educational programme. Examinations, in their different varieties such as written or oral, have been largely used by educational institutions, examination bodies and other establishments providing training. According to Obemeate (2000), test items used for this purpose, usually take time to organize since items should possess acceptable level of validity and reliability. In addition, items of a test were expected to possess high degree of usability. For suitability of test items, items must be developed by experts in the field from the syllabus.

Examination bodies such as Joint Admissions and Matriculation Board (JAMB) and West African Examinations Council (WAEC) may have much in common with National Examination Council (NECO) in terms of item development even though the purposes are different. WAEC and NECO develop items that are used for certification of learners, while JAMB test is used to obtain admission into tertiary education such as universities, polytechnics, and College of Education.

For a test to be effective it must possess the necessary psychometric properties. Igwe (2016) defined psychometric properties of an examination as "the validity, reliability and item characteristics of the examination.'' Some important psychometric properties of test include difficulty, discrimination indices and the distracters effectiveness. According to Moyinoluwa (2015) psychometric characteristics of examinations refer to certain attributes inherent in tests upon which an assessment of candidates is based. These characteristics include validity, reliability, the difficulty indices, the discrimination indices and the power of distracters.

Validity, as applied to a test, is a judgment or estimate of how well a test measures what it purports to measure in a particular content (Cohen, 2009). Validity of an instrument is defined by Ajai and Amuche (2015) as the accuracy with which a test measures what it attempts to measure. A valid test must measure accurately and consistently what it is designed to measure. A test can be judged valid if it measures what it is intended to measure (Hathcoat, 2013). Similarly, Messick cited in Domino (2012) defines validity as an integrated evaluation judgment of the adequacy and appropriateness of interpretations and based on the assessment measure. More specifically, validity is a judgment based on evidence about the appropriateness of inferences drawn from test scores. An inference is a logical result or deduction. Characterizations of the validity of tests and test scores are frequently phrased in terms such as "acceptance" or "weak". Validity reflects a judgment about how adequately items of the test measure what they purport to measure. There are different types of validity. They are (i) face validity (ii) content validity (iii) construct validity (iv) Criterion-Related validity.

Content validity test describes a judgement of how adequately a test covered selected content taken as a representative of the universe of topics that the test was designed to sample. With respect to educational achievement tests, it is customary to consider a test content-validity measure as when the proportion of material covered by the test approximates the proportion of material covered in the course. Content validity refers to the comprehensiveness of the instrument in covering the content areas that have been treated during instruction (Anikeze, 2012). It implies the adequacy with which the test items appropriately represent the sample of the content area to be measured. Content validity is often built in by a concentrated effort of the examiner to make sure that the sample of behaviour, that is the test, is truly representative of the domain being assessed. Such an effort requires first of all, a thought knowledge of the domain.

A mathematics NECO test would be considered as content-valid, if the proportion and type of mathematics problems of the instrument approximates the proportion and type of mathematics problems presented in the course syllabus. Content validity is built into a test from the choice of appropriate items. Anastasi and Urbina (2008) had reported that for educational tests, the preparation of the items is preceded by a through and systematic examination of relevant course syllabic and textbooks as well as by consultation with subject-matter experts. On the basis of the information thus collected, test specifications are summed up for the item writers. Content validity evidence is established by comparing test items with instructional objectives develop with. The aid of a test blueprint as to determine whether the items match or measure the objectives. After such an examination had taken place, a test is judged either to have or not to have validity evidence for a particular subject. No correlation coefficient is computed. Instead, human judgment is relied upon.

In validating a test, several methods of quantifying content validity have been developed. One method for determining agreement among raters or judges was developed by Lawshe in 1975. According to Lawshe cited in Cohen (2009), if more than half of the panelists indicate that an item is essential, that item has at least some content validity.

Reliability is a very vital quality that scores on a test must possess for it to be accepted as a good one. According to Anastasi as cited in Ajai and Amuche (2015) reliability means consistency of scores obtained by same individual when re-examined with different sets of equivalent items or under other different examining conditions. Anikweze (2016), opined that reliability of test indicates the degree of accuracy with which a test measure what it is designed to measure. It

implies consistency of test results over time and item. However, a reliable instrument can measure accurately and consistently but not necessarily what it set out to measure. Hence, a reliability test may not be valid although every valid test must have the properties of reliability. Ajai and Amuche (2015) explained that the reliability test is expressed numerically, usually a coefficient ($r_{xx)}$. The subscripts in the reliability of test is a correlation between two forms of the same test. The coefficient varies between two values 0.00 (no reliability) and 1.00 (perfect reliability). Snowman and Biehler (2003) stated that there are varieties of methods to occasions and measure the extent to which the ranking change over time. This approach results in test-retest reliability.

A reliable and valid instrument could be unusable because of other factors that affect the quality of a good instrument. Utility in the context of testing and assessment refers to the usefulness or practical value of testing to improve efficiency (Cohen, 2009). A number of considerations are involved in making a judgement about the utility of a test such as; psychometrically soundness, costs, and benefits. A test is said to be psychometrically sound for particular purpose if the reliability and validity coefficients are acceptably high (Cohen, 2009). An index of reliability can tell us something about how consistently a test measure what it measures; and an index of validity can tell us something about whether a test measures what it purports to measure. Thus, it is expected that Mathematics NECO questions used in Nigeria and Adamawa State in particular should be valid and reliable.

Given the importance attached to Mathematics, the prospect of students has become a matter of concern because of the possibility that students will no longer be self-employed and self-reliant as students leave school. Students will also be unable to study Mathematics-related subjects like Physics, Chemistry, Economics, Accounting and Marketing at the tertiary level of education (Ugodulunwa & Barko, 2015). The Federal Ministry of Education guidelines on construction of Basic Education examination specify that NECO question papers are to be constructed using questions developed by teachers in workshop under the management of the State examination office. These questions are edited and some are selected to create the final paper by office staff assisted by technical experts. This does not appear to be any pre-testing and their quality varies considerably in both content and presentation (Wilmut & Yakasai, 2016). It is the nature of the results that comes out of BECE Mathematic question in recent years that prompted this study to undertake the item analysis of Mathematics objective items. Thus, the main purpose of this study is to determine the nature of the item and their adequacy for use as NECO Mathematics.

Statement of the Problem

The recent decline in quality of education in Nigeria calls for public concern. This situation becomes crucial point of reference because candidates sitting for examinations are expected to have been taught and prepared adequately by their teachers through several methodological activities in the classroom. One of the myriads of these activities is achievement tests, which are to be used to obtain response information that would enable the school to place the candidate to particular achievement level after results are released. This then serves as feedback for the candidates' parents and government, both of which are stakeholders in the day-to-day running of the school system.

Since one of the examinations used in Nigeria to assess students' academic achievement at the end of 9-year Basic Education is set by NECO in collaboration with the Federal Ministry of Education, the study observed that this examination is fraught with reliability and validity challenges as record from schools that registered for the examination show a very high percentage of pass returned. The

result does not discriminate well between students in high ability group and low ability group. All these lead to significant measurement error into the measurement process. When measurement is poor, then there will be an inaccurate data-base inference, which in turns leads to wrong decision-making. The use of a poorly designed instrument is a major problem as it affects students' interest, career choice and academia achievement. When an instrument does not possess the necessary characteristics, it ought to, this means that the examination is not valid and reliable, and the effort to achieve the educational objectives will be in futility. Could this be the case of Mathematics NECO BECE administered in Adamawa State?

Research Question

The following research questions were formulated to guide the study;

i. What is the Content Validity indices of objective NECO BECE Mathematics items for 2018, 2019 and 2020 in Adamawa State, Nigeria.?
ii. What is the coefficients of internal consistency of objective of NECO BECE Mathematics questions (items) for 2018, 2019 and 2020 in Adamawa State, Nigeria?

Research Hypotheses

The following null hypotheses were postulated to guide the study. All null hypotheses were tested at 0.05 level of significance.

i. Content Validity indices of objective NECO BECE Mathematics items for 2018, 2019 and 2020 in Adamawa State, Nigeria do not differ significantly.
ii. Coefficients of internal consistency of objective of NECO BECE Mathematics questions (items) for 2018, 2019 and 2020 in Adamawa State, Nigeria do not differ significantly.

Empirical Review

Nwaogazie (2014) in a study investigated the influence of content validity on secondary school students' academic achievement in Imo State, Nigeria. A sample of four hundred and twenty students was randomly selected from fifteen thousand, five hundred senior secondary school two (15,500) students of 2009/2010 academic session in Imo State. The data used were collected mathematics achievement students' scores using the researchers' constructed instrument called mathematics achievement and psychometric indices instrument (MAAII). The reliability index of the instrument was 0.9. The data were analyzed using mean and standard deviation for the research question while the hypothesis was tested using Z-ratio at 0.05 level of significance. The result also showed that the calculated Z-ratio between the two variables is 6.04 while the critical value of Z is 1.96 at 0.05 alpha level under 418 degrees of freedom. Since the calculated Z-ratio (6.04) is greater than the critical value of Z (1.96) at 0.05 level of significance at 418 degrees of freedom, the null hypothesis was rejected. This indicates that there was a significant difference between the mean academic achievement of students tested on item with high content validity indices and their mean academic achievement on items with low content validity indices. This study investigated the influence of content validity on students' academic achievement while current study determined content validity of Mathematics NECO questions.

Adebule (2009) examined the reliability and difficulty indices of Multiple Choice (MC) and True or False (TF) types of objective test items in a Mathematics Achievement Test (MAT). The study

used survey descriptive design. The instrument used were variants – 50 - items Mathematics achievement test based on the multiple choice and true or false test formats. The population for the study consisted of all senior secondary school three students who were preparing for the Senior Secondary School Certificate Examination (SSCE) in Mathematics. A total of five hundred (500) students randomly selected from ten senior secondary schools in five Local Government Area of Akure, Ondo State served as sample for the study. The result revealed that reliability coefficients for MC and TF were 0.35 and 0.25, respectively. The result of the null hypotheses tested at p=0.05 showed that there was no significant difference between the reliability coefficients of MC and TF test items. Also, there was no significant difference between the difficult indices of MC and TF test items. Hence, it was recommended that a combination of both MC and TF type of tests should be used to evaluate students' achievement in schools. This study is similar to the present study in determining reliability and item difficulty but differ in research design, sample and study area.

Nwaogazie (2014) in a study investigated the influence of content validity on secondary school students' academic achievement in Imo State, Nigeria. A sample of four hundred senior and twenty students was randomly selected from fifteen thousand, five hundred senior secondary two students of 2009/2010 academic session in Imo state. The data used were collected from the students' scores using the researcher's constructed instrument called mathematics achievement and psychometric indices instrument (MAAII). The reliability index of the instrument was 0.9. The data were analyzed using mean and standard deviation for the research questions while the hypothesis was tested using Z-ratio at 0.05 level of significance. The result also showed that the calculated Z-ratio between the two variables is 6.04 while the critical value of Z is 1.96 at 0.5 alpha level under 418 degrees of freedom. Since the calculated Z-ratio (6.04) is greater than the critical value of Z (1.96) at 0.05 level of significance at 418 degrees of freedom, the null hypothesis was rejected. This indicates that there was a significant difference between the mean academic achievement on items with low content validity indices. This study investigated the influence of content validity on students' academic achievement while current study determined content validity of mathematics NECO questions.

Research Design

This study employed an evaluative research design, since the data involved in the study was collected from the source without any manipulation. According to Trochim (2006), an evaluation design is the systematic acquisition and assessment of information to provide useful feedback about some objects. This design is mostly used to provide feedback on an event, organization, programme, policy, technology, person, activity among others. Hence, evaluation research design is considered most appropriate for this study since the study sought to assess the quality of NECO as an instrument to obtain feedback on item characteristics of the examination.

The population of the study consisted of 6,633 JSS III students from the Adamawa (NECO office, Adamawa state 2020). The population comprises of both male and female students that took NECO BECE. JSSIII is chosen as population of the study because it was the class that took the NECO BECE examination at the end of the 9-year Basic Education.

A sample of 556 JSS III students was randomly selected from 12 Local Government Areas (LGAs). The sample size was obtained using Krejci and Morgan (1970) probability sampling table. Hat-and draw with replacement technique was employed in the selection of the respondents from 12 LGAs. Probability proportional to sample (PPS) method was used in selecting respondents from the selected LGAs.

Two research instrument was used for the data collection. The instrument was Basic Education Mathematics Achievement Tests (BEMAT) for the years 2018, 2019 and 2020 and a content validity checklist. Basic Education Mathematics Achievement Tests (BEMAT) were past objectives Mathematics NECO question for 2018, 2019, each having 60 multiple-choice test items. The past question papers for 2018, 2019 and 2020 were tagged "Type `A`, `B` and `C" respectively. The instrument was administered by the researcher and three trained research assistants under the supervision of the researcher to JSS III students (the group that will take 2020 examination). The instruments consisted of 60 multiple-choice items each with 5 options (letter A-E). Each sampled student answered either type `A`, `B` or `C` of the BEMAT. The other instrument that was used in data collection is content validity checklist; which was used in obtaining data that was used in determining content validity.

To determine the validity of the BEMAT, copies of the instrument were given to three experts from Faculty of Education, Taraba State University Jalingo. The experts were requested to check for both the face and content validity of the instrument and to rate the instrument. The comments from the raters enabled the researcher to review the instrument and incorporate their observation in the final version of the instrument based on the comments from the three validators. Validator I commented that the instrument should be administered during third term and to JSS3 students, validator II commented that the items are Ok while validator III commented that the instrument is good enough to help answer questions raised and the hypothesis stated.

A pilot study was carried out: the purpose of the study was to determine the reliability of BEMAT 60 JSS 3 students were randomly selected from three schools in Jalingo for the test. 20 copies each of BEMAT type A, B and C were distributed to the sampled students. The schools use was outside the sample of the study, but have some degree of similarities with sampled schools. Their response was scored and analyzed using split-half formula to determine the internal consistency of the instrument. The pilot study was meant to serve as a model and insight to the main study.

Reliability of an instrument was indicative of its consistency or the relatedness within a measure. For the purpose of this study, the reliability of BEMAT type `A`, `B` and `C` were analyzed using split-half method at 0.05 significant level, the result of the reliability analyses revealed that the spearman-Brown coefficient (Equal Length) was .997, .887 and .993 respectively, which were above the acceptance level of .80 for achievement test as asserted by Ajai and Amuche (2015).

The researcher with the aid of an introductory letter from the department of educational foundation, Taraba State University visited the sampled schools. On arrival at each school, the introductory letter was presented to the school principal or any delegated authority seeking for permission to conduct the study.

Three assistants were trained by the researcher and were used in the invigilation of the examination, as they are instructed to sit the students as in the examination situation, asked the students to read the instruction very careful while waiting to start and attend to the students in times of difficulties to the questions for clarity in each of the sample schools, the researcher with the help of a trained research assistants administered the BEMAT to the sampled JSS III students.

Descriptive statistics of mean and variance were used to answer questions. Research question one was answered using mean by comparing the items in the instrument with the content in the national curriculum. Research question two was answered using Kuder-Richardson 20 formula. Research question three and four was answered using IRTPro (a specialized software for determining item

characteristics based on the assumptions of IRT). All research hypotheses are tested using Analysis of Variance (ANCOVA). ANCOVA was an inferential statistic used in measuring significant among means of three groups.

Research question one

What are the validity indices of objective test items of Mathematics BECE question for 2018. 2019 in Adamawa state?

Table 1: Summary of objectives items of mathematics BECE question for 2018, 2019 and 2020 in Adamawa state, however these themes were collapse into five themes as some of the themes overlap to next level.

S/N	Themes	Weights Based On number of topics in a theme	Weights Based On number of items from each theme 2018 BECE	Weights Based On number of items from each theme 2019 BECE	Weights Based On number of items from each theme 2020 BECE
THEME A	Everyday Statistic	9 (16.67%)	12 (20%)	20 (33.33%)	17 (28%)
THEME B	Number and Numeration	15 (27.78%)	29(48.33%)	27 (47%)	27 (47%)
THEME C	Basic Operation	11(20.37%)	18(30%)	12 (20%)	16 (27%)
THEME D	Algebraic	8 (14.81%)	0 (0%)	0(0%)	0(0%)
THEME E	Geometry and mensuration	11 (20.37%)	1 (1.67%)	1(1.67%)	0(0%)
	Total	54 topics	60 items	60 items	60 items
Content validity index			**(36/54)=.67**	**(36/54)=.67**	**(35/54)=.65**

Table 1 provides content coverage of mathematics BECE for 2018, 2019 and 2020, with breakdown of weighting based on number of themes in each theme and weighting based on number of items from each theme. The table shows that in 2018 and 2019 only one topic was set from THEME B. The remaining 59 items measured topics from THEMES A, B and C. In 2020 mathematic BECE, all the 60 items in the instrument were measuring themes from number and numeration, basic operation, algebraic processes, geometry and mensuration, with no single item measuring everyday statistics. Based on this information, it can be seen that the instruments had moderate content validity.

Research question two

What are the coefficients of internal consistency of objective test items of mathematics BECE question for 2018, 2019 and 2020 in Adamawa state?

Table 2: Kuder-Richardson KR-20 reliability of BECE mathematics objectives Test Items for 2018, 2019 and 2020.

BECE Internal Consistency					C0-efficeint of
2018	1.017	12.579	50.777	0.248	0.752
0.765					
2019	1.017	13.095	60.013	0.218	0.782
0.795					
2020	1.017	13.151	64.598	0.204	0.796
0.810					

In Table 2, it was observed that Kuder-Richardson 20 analysis of internal consistency of BECE 2018 indicated that Mathematics objectives test items had a variance of 50.777, standard deviation of 7.126 and a coefficient of internal consistency of 0.765; BECE 2019 Mathematics objective test items had a variance of 60.013, standard deviation of 7.747 and a co-efficient of internal consistency of 0.795; while BECE 2020 Mathematics objective tests items had a variance of 64.598, standard deviation of 8.037 and co-efficient of internal consistency of 0.810. This shows that tests had high co-efficient of internal consistency. The table further shows that 2020 Mathematics had the highest co-efficient of 0.81; while BECE Mathematics 2018 had the least co-efficient of 0.765.

Hypothesis one:

Content validity indices of objective Mathematic BECE question for 2019, 2019, and 2020 in Adamawa state did not differ significantly.

Table 3: One way analysis of variance (ANCOVA) of the content validity indices of mathematics BECE objectives Test for questions for 2018, 2019 and 2020 in Adamawa State.

	Sum of Squares	Df	Mean square	F	Sig
Between Groups	.000	2	.000	.000	1.000
Within Groups	.466	12	.039		
Total	.466	14			

Analysis of variance in table 9 found that there was no statistically significant difference among the mean content validity indices of 2018, 2019 and 2020 objectives test items of mathematics BECE question in Adamawa State (F= .000, p=1.000). on this premise null hypothesis one which says that there is no significant difference among the content validity indices of objectives test items of mathematics BECE question for 2018, 2019 and 2020 in Adamawa State retained.

Hypothesis two

Coefficient of internal consistency of objective test items of mathematics BECE question for 2018, 2019 and 2020 in Adamawa State did not differ significantly.

Table 4: Kruskal-wallis independent samples Test of difference among the coefficient of internal consistency of objectives test items of mathematics BECE question for 2018, 2019 and 2020 in Adamawa State.

	Content Validity Indices	Decision
Chi	2.000	Retained Null Hypothesis
Df	2	
Asymp. Sig.	3.368	

Kruskal-wallis H test in Table 4 shows that there was no statistically significant difference among the coefficients of internal consistency of objective test items of Mathematics BECE questions for 2018, 2019 and 2020 in Adamawa State, $x^2(2)=2.000$, p=0.368, with a mean rank coefficient of internal consistency of 0.765 for 2018, 0.769 for 2019 and 0.810 for 2020. Based on this results, null hypothesis two which says that there was no significant difference among the coefficients of internal consistency of objective test items of mathematics BECE questions for 2018, 2019 and 2020 in Adamawa State is retained. This implies that there was no significant difference among the coefficients of internal consistency of objective test items of mathematics BECE question for 2018, 2019 and 2020.

Discussion findings

The research question one on the extent of content coverage of the examination showed that the instruments had moderate content validity but items were not fairly spread across in the theme. This might be as a result of non-adherence to the NECO BECE in the generation of test items by test developers. This finding is similar to that of Faleye and Ayeloja (2014) but contradicts the finding of Ugodulunwa and Barko (2015) who reported that mathematics JSSCE in Plateau State had low content validity which they attributed to improper item generation and arrangement. This finding also contradicts the findings of Oribhabor and Emafo (2016) who found low content validity in mathematics Test constructed by senior secondary school teachers in Edo State, and Joe-Kinnance and Opara (2017) who found that English language JSSCE question from 2014 to 2016 lacked content validity.

The findings of hypothesis one showed that there was no significant difference among the content validity indices of objective test items of Mathematics NECO BECE question for 2018, 2019 and 2020 in Adamawa State, with F= .00, p=1.000. Based on this result the null hypothesis which says that there was no significant difference among the content validity indices of objective test items (Adebule, 2009). By implication this result indicates that Mathematics BECE question have moderate content validity.

The findings research question two on the reliability of objectives test items in 2018, 2019 and 2020 mathematics BECE revealed coefficient of internal consistency of .765, .795, and .810 respectively, obtained using KR-20 formula which were reliable. This findings is in agreement with Ugodulunwa and Barko (2015), Oribhabor and Emafo (2016) and Joe-kinance and Opara (2017) who found reliability coefficient above o.70. this result implies that BECE result can be relied upon in making educational decisions such as placement of students into secondary school and guiding students in making carrier choice.

The findings of hypothesis two showed that there was no significant difference among the coefficient of internal consistency of objectives test items of Mathematics BECE question from 2016 to 2018 in Adamawa State, with $x^2(2)= 2.000$, p=0.368. based on this result the hypothesis which says that there was no significant difference among the coefficient of internal consistency of objectives test items of Mathematics NECO BECE question from 2018 to 2020 n Adamawa State is retained. This implies that there was no significant difference among the coefficient of internal consistency of objective test items of Mathematics NECO BECE question from 2018 to 2020. This result is similar to Adebule (2009) who found that there was no significant difference between the reliability coefficient of MC and TF test items.

Conclusion

Based on the findings, the following conclusions were made:

NECO Mathematics Basic Education Certificate Examination (BECE) administered in Adamawa state for 2018, 2019 and 2020 had coefficient of internal consistency of 0.765, 0.795 and 0.810 respectively and the examination was found to be reliable. The instrument of NECO BECE administered in Adamawa state had moderate content validity but there was no fairly equitable distribution of items among themes. It is also clear that, the NECE BECE examinations administered in Adamawa state was able to discriminate well between the upper and lower ability students.

Recommendations

Based on the findings of the study, the following recommendations were made:

Items that failed to meet the validity; discrimination and reliability indices should be eliminated or revised by the examination body.

There should be equitable distribution of items among the topics across all the five themes of Mathematics as contained in the national curriculum; test development and content experts should be involved in developing and validating the test items in order to obtain valid and reliable result which lead to valid inferences by stakeholders.

Experts in Educational Measurement and Evaluation should carry out the rigorous process of item analyses by determining item validity and items reliability indices, not by hand-picking from past questions.

References

Adebule, O. S. (2009). Reliability and levels of difficulty of objective test of ten senior secondary schools in five local government areas of Akure, Ondo state. Educational Research and Review, 1), 585-587.

Ajai, J. T. & Amuche, C. I. (2015). Educational research methods and statistics. Academica House Publishers Nig. Ltd Jos, Plateau State.

Amuche, C. I. &lgba, G. J. (2016). Influences of content validity of Teacher-made-Tests on Physics students' academic achievement in Taraba state. Journal of Science and Technology, Mathematics and Entrepreneurial Education. (JSTMEE). 1(1): 1 11-121.

Anastasi, A. & Urbina, S. (2008). Psychological testing. 8th Edition. Upper Saddle River, NJ: Prentice-Hall.

Anikweze, C. M. (2009). Simplified approach to educational research. Enugu: SNAAP Press (Nig.) Ltd

Anikweze, C. M. (2010). Measurement and evaluation: For teacher education. (2nded.). Enugu: SNAAP Press Ltd.

Cohen, R. J. (2009). Psychological testing and assessment: An introduction to tests and measurement. (7thed.). New York: McGraw-Hill.

Cohen, L, Manion, L. and Morrison, K. (2013) research methods in education (6th ed.) London: Routledge.

Domino, G. , & Domino, M. L. (2012). Psychological testing. An introduction, UK. Cambridge press.

Faleye, B. A. , &Ayeloja, O. R. (2014). Taxonomical analysis of selected Teacher-made multiple choice Tests in Obafemi Owolowo University Nigeria. Journal ofEducational and Social Research. 4(3), ISSN 2240-0524

Hathcoat, J. D. (2013). Validity semantics in educational and psychological assessment. Practical Assessment, Research and Evaluation, 18(9):1-14.

Ifeakor, A. C. (2011). Psychological measurement & evaluation in education: Issues 75and application. Onitsha: Folmech Printing and Publishing Co.Ltd.

Joe-Kinanee, J. N. B. &Opara, I. M. (2017). Psychometric properties of English language in junior secondary school certificate examination questions in rivers state, International Journal of Scientific Research in Education, 577-589.

Moyinoluwa, T. D. (2015). Analysing the psychometric properties of Mathematics in public examinations in Nigeria. Research on Humanities and Social Sciences www.iiste.orgISSN (Paper)2224-5766 ISSN (Online)2225-0484 (Online)Vol.(5)7.

Nwaogazie, J. I. (2014). Influence of content validity of secondary school students' academic achievement in Imo State, Nigeria. JORIND, 12 (l), 182-190.

Obemeata, J.O. (2000): principle of essay and multiple choice test construction. Ibdan:Stiling-Horden Publishers (Nig) Ltd.

Oribhabor, C. B. &Emafo, O. D. (2016). Determining the reliability and content validity of the Mathematics tests constructed by senior secondary school Mathematics teachers in Edo State, Nigeria. African Journal of Education, Science and Technology, 3(2).

Trochim, W. M. (2006). Research methods knowledge base. Drake University. Retrieved from: http://www.socialresearchmethods.net/kb/intreval.htmon 12th April, 2018.

Ugodulunwa, A. C., &Barko, L. (2015). Analysis of psychometric properties of Business studies junior secondary certificate examination in Plateau state, Nigeria. Online Interdisciplinary Research Journal, 5(2), 1-7. Retrieved December 20, 2015, from http://www.oiirj.org/oiij/mar-apr2015/21 .pdf

Wilmut, L, &Yakasai, I. M. (2016). A brief review of the assessment of student achievement in Kaduna, Kano and Kwara states of Nigeria. DFID/World Bank SESP preparation mission.