Innovations

Construction and validation of physics achievement test for senior secondary schools in Nigeria

[1]Nwabeze, Prince. Chukuka, [2]Odili, J.N., and [3]Ossai, P.A.U.
Department of Guidance and Counselling, Delta State University, Abraka, Nigeria

Corresponding author: Nwabeze, Prince. Chukuka

Abstract
The purpose of this study was to develop and validate an achievement test for senior secondary schools in physics using three parameter logistic Item Response Theory. Three research questions were raised. This study used the instrumentation research design. The population for the study comprised of 37,441 in Delta and Bayelsa States. The sample of the study consisted of 1,200 senior secondary school three students using stratified random sampling technique. The instrument used for the study was Physics achievement test made up of fifty items. The content and face validity of the instrument were established by the use of table of specification and experts judgement while, the reliability of the instrument was established using Kuder-Richardson formula 20 and a reliability coefficient of 0.92, was obtained. The IRT Bayesian Modal Estimator (BME) software and the statistical packages for social sciences (SPSS) were used for the analysis. The fit statistics was used to identify the items that fitted the one parameter (1PLM), two parameter (2PLM) and the three parameter (3PLM) logistic model at 0.05 alpha level. The findings of the study showed that the Physics achievement test had good psychometric properties. The results showed that the overall reliability co-efficient (KR-20) of the PAT was 0.93, the fit to the model was good. The physics achievement test measured one single trait for physics ability.Based on the summary of the results, it was determined that the Physics achievement test constructed by the researcher has good and relevant psychometric properties of a test. As such it would serve as an effective formative evaluating instrument for Physics students. It was suggested, among others that test experts and developers should explore the use of IRT model in the construction and validation of Physics achievement test for senior secondary school students.

Keywords: 1.*Construction, 2. Validation, 3.Physics and Achievement test*

Introduction

In Nigeria, education was informal prior to the introduction of Western schooling. There was no official assessment or testing mechanism to determine the degree of behavior change brought about by the teaching process. The development of Western education gave rise to a planned and regulated method of examination. Internal and external examinations are the two categories into which these tests fall. Exams created and conducted within the context of the institution, such as promotion exams and terminal exams, are considered internal examinations. Omorogiuwa and Iri–Aghedo (2016), On the other hand, external exams are tests that are created, administered, and evaluated by public testing organizations. Examples of public examination bodies include the State Ministry of Education, West African Examination Council (WAEC), National Examination Council of Nigeria (NECO), National Business and Technical Examinations Board (NABTEB), and Joint Admission and Matriculation Board

(JAMB), among others. These exams are external to the students and are used for placement, certification, and job opportunities, to name just a few.

Omorogiuwa and Iri–Aghedo (2016), opined that test is a set of item(s) or series of tasks or set of questions presented to a learner in which they are to respond orally or simply in writing that enables it to evaluate differences between learners. However, a test is seen as any form of device or procedure that is used for measuring achievement, ability, aptitude, interest and any other personal attributes or traits. As observed by Thomas (2003) and Spike (2005), testing serves the following purposes, such as selection of students into educational programmes, prediction of students' future performance or potential, classification of students, provision of information and significant data for counselling and diagnosis, evaluation and verification of academic programmes as well as certification after graduation from a given educational programme. As is done in any other subject areas, testing is done in physics for the purposes mentioned.

Physics is a branch of physical science that focuses primarily on the interaction between matter and energy. Physics is a fundamental subject whose advancements continue to fuel the rapid advancement of our contemporary technological society, as mentioned by Usman and Abubakar (2019). The basis of contemporary science and technology is physics. It is the cornerstone for fundamental understanding, comprehension, and knowledge of the fundamental concepts on which its application considerably improves the quality of life in the technologically advanced society in which we live.

An achievement test is a test created to determine how much information a person has acquired via teaching in a particular field or collection of areas. According to Ani (2014), an achievement test is a tool used to apply stimuli to an individual or group in order to elicit specific desired or expected responses that represent that person's ability. Every measuring device, such as a test, is supposed to have specific features so that the results it yields may be trusted (Ani, 2014). According to Obilor (2019), an achievement exam is an aptitude test created to assess the level of intent that a person has mastered as a result of deliberate prior experience or training. It is used to determine the degree to which a person has succeeded or acquired specific knowledge or skills as a result of exposure to training. The teacher can evaluate student development in a particular subject and determine whether or not the specified goals have been met by using achievement tests.

Test construction refers to the methodical process of assembling test items or the creation of a test by creating and compiling a set of questions that serve as the test taker's task (s). As a result, test construction is a specialty area for differential psychologists and test and measurement experts, according to Osadebe (2011). Consequently, creating a successful test takes expertise and skills.He suggested that teachers should be guided in construction of items/test following the steps stated below;

1) Planning of Items Generation/ Test
2) Item writing
3) Item analysis – trial testing
4) Composition of item
5) Reliability
6) Printing and administration
7) Marking and scoring (measurement)
8) Manual

Hence, the physics teacher, having undergone a course of training in measurement and evaluation should be able to construct, validate, select, administer and score a test as well as interpret the scores generated from tests. There is a notable short fall of test material standardized or normed in Nigerian culture for use in educational services. This situation creates a problem in educational development of the students, for example in Nigeria; it is unfortunate that many of the inappropriate vocation choices are as a result of ignorance and illiteracy on the parts of parents. They have fixed ideas of the course of study, which they would want their children to study, and also the interest of their children in the said course of study (Osadebe&Nwabeze, 2018).These led students into choosing courses that are either too low or above their mental capacity.

It should be noted that students' low intellectual ability, poor aptitude disposition coupled with low level of aspiration for particular courses like physics have a devastating implication on the students. These problems are as a result of absence of requisite test materials in physics and also in educational and vocational counselling. It should be noted clearly that many failures recorded in the sciences today especially in physics is as a result of wrong selection of courses to be studied.

Adegoke (2013) investigated the comparability of item statistics produced using the classical test theory (CTT) and two-parameter item response theory frameworks (IRT). Four research questions guided the study. Instrumentation research design was adopted for the study. The population of the study consisted of seven hundred and twenty four senior secondary school two (SSS II) students Age (16-18 years) participated in the study. The students were selected from 16 randomly selected senior secondary schools in Ibadan Educational Zone I, Oyo State, Nigeria. All the students were offering all science subjects including physics. There were 451 boys (62.3%) and 273 girls (37.7%). A 60-item Physics Achievement Test was developed and administered to 724 senior secondary school two students (Age 16-18 years), who were randomly selected from 16 senior secondary schools in Ibadan Educational Zone I, Oyo State, Nigeria. Results showed that item statistics obtained from both frameworks were quite comparable. However, item statistics obtained from IRT 2-parameter model appeared more stable than those from CTT. Moreover for item selection process, IRT 2-parameter model led to deletion of fewer items than CTT model. Orangi and Dorani (2010) conducted a research to develop a social studies achievement test for the high school students based on item-response theory (IRT). The purpose of the study was to develop a social study achievement test for high school student's first grade based on item response theory. The results showed that the constructed forms were of high reliability, they were at the same time acknowledgeable through the analysis based on the classical method and they were also in accordance with the three – factors of the Item Response Theory. Taking into account the item characteristics curve, both of the forms produced the knowledge for the students with average ability. In this analysis a kind of rank-percentile norm for the both sex was formulated. This study has one feature with the current study, for example: the design was the same but differ with the sampling technique and sampling size.

Ani (2014) conducted a study on application of item response theory in the development and validation of multiple choice tests in economics. Six research questions and two hypotheses guided the study. Instrumentation research design was adopted for the study. Data was gathered using the 50 multiple-choice questions about economics that the researcher had constructed.It was possible to get a reliability index of 0.89. The data from the study were examined using the likelihood estimate technique of the computer application BILOG-maximum MG. The data analysis showed that 50 test items for economics survived, hence the final instrument created to evaluate students' economics proficiency included 50 items with the proper indices.Based on the three parameter model (3pl) models, the study's findings indicated that 49 items of the multiple choice questions in economics were reliable. The results also revealed that 31 out of the 100 questions on the economic multiple-choice test were challenging. The results also showed that things function differently for male and female students in economics.

Akuche and Aliyu, (2019)conducted research on the interaction between the 1-PL of Rasch and the 3-PL Models of IRT in the item selection for a designed test for assessment. To direct the investigation, three research questions were developed. An instrumentation research design was used for the investigation. All pupils in Oyo State's senior secondary school two (SSII) make up the study's population. A sample of ten (10) senior secondary schools was taken. The ten (10) senior secondary schools were chosen using simple random sampling procedures of balloting. The study's sample size was 755 respondents, of whom 75 were tested from each of the nine schools using a non-proportionate stratified random sampling technique, and 80 were drawn from one of the ten secondary schools that were chosen. The Physics Aptitude Test (PAT), which the researcher developed, was the instrument utilised for this investigation. It had 100 items. Three parts make up the test content. The test's questions were based on a carefully thought-out test blueprint that brought together the six levels of the cognitive learning domain. It consists of three aptitude test components: the Verbal Aptitude Test, which has the most items with fifty (50); the Abstract Aptitude Test, which has forty-three (43); and the Numerical/Quantitative Aptitude Test, which has fifty-seven (57) items (57 items). This displays the distribution of the 100 PAT test items among the subject areas and learning objectives.After the test items had undergone experimental try-out and revision, a total of 50 items that made up the

PAT were selected, utilising the Classical Test Theory (CTT) approach. The difficulty and discrimination indexes were utilised to choose fifty test items in total.

Aliyu (2015) conducted a study on the use of the rasch model in the development and validation of mathematics achievement tests. Nine research questions and four hypotheses guided the study. Instrumentation research design was adopted. All senior secondary class III students in Oyo and Delta States made up the study's population. A multistage sampling procedure was used in order to choose one thousand five hundred (1500) pupils from the population. The tool included 25 social economic status questions and 150 items that were self-developed (SES). The table of specifications and factor analysis were used, respectively, to analyse the content and construct validities, while some experts' opinions on the item development were used to determine the face validity. Using rotational component matrix and principal component analysis (PCA), the experts' chosen items were reduced to 100 (RCM).

The reliability value of the items using KR20 was 0.85.The degree to which the mathematics fit the Rasch model of IRT was investigated using fit analysis of Winsteps 3.75 and a t-test. In order to assess test takers' accomplishments, the Rasch model addressed a number of measuring concepts that were essential to construct validity. The findings revealed that 65 items had high psychometric qualities in addition to satisfying the Rasch model's requirement that a measurement construct be both fitting and invariant.

Research Questions

The study was guided by the following research questions:

i. What is the Validity of Physics Achievement Test in Delta and Bayelsa States?
ii. What is the Reliability of Physics Achievement Test in Delta and Bayelsa States?
iii. What is the Difficulty level of the Physics Achievement Test in Delta and Bayelsa States?

Method

This study adopted the instrumentation research design to construct and validate the Physics Achievement Test for Senior Secondary School Three (SS3) students in Delta and Bayelsa.

The population of this study comprised of all the public senior secondary schools in Delta and Bayelsa States. The total population of secondary schools in Delta State is 450 with SS3 students population of 25,000 in the 25 Local Government Area (LGA) while Bayelsa State had 205 schools with SS3 students population of 12,441 in the 8 Local Government Area.

The sample of this study consisted of one thousand two hundred (1,200) from the population of twelve thousand (12,000) senior secondary school three students in the six senatorial districts of Delta North, Delta Central and Delta South; Bayelsa Central, Bayelsa East and Bayelsa West of Delta and Bayelsa States. The multi-stage sampling technique was adopted in this study. In the first stage, stratified random sampling technique was used to select two (2) Local Government Areas in each of the six senatorial districts which culminate to twelve (12) LGAs. From the twelve LGAs, four (4) schools was selected in each of the LGA, making a total of forty-eight (48) senior secondary schools selected for the study in the six (6) senatorial districts of Delta and Bayelsa States.

In the last stage, the researcher used simple random sampling technique to select twenty five (25) senior secondary school three students from the forty-eight (48) senior secondary schools. Hence, from the 12,000 students 10% was used to select 1,200 students. This involves writing random numbers in a piece of paper. The piece of papers were carefully folded and placed in the basket. The piece of papers in the basket were picked with replacement to get 600 students to represent SS3 students in each of the sampled states drawn from Delta North, Delta Central, Delta South, Bayelsa Central, Bayelsa East and Bayelsa West Senatorial District of Delta and Bayelsa States, making a total sample of 1,200 students for the study.

The research instrument for this study is Physics Achievement Test (PAT) generated by the researcher. The instrument contained 100 items on Physics Achievement Test. Finally Factor Analysis of the SPSS was used to select 50 items for the study. The items in the instrument were drawn from senior secondary school syllabus. The items cover the following areas in Physics curriculum namely: Interaction of matter, space and time, Conservation

principles, Waves: Motion without material transfer, Fields at rest and in motion, Energy Quantization and Duality of matter and Physics in Technology. Each item is made up of stem and four options from which the student is expected to select the correct option.

Validity is one of the most essential psychometric properties of a test. The validity of a test is the extent to which a test measures what it is supposed to measure. In order to establish the content validity of the test items, a table of specification was drawn to ensure adequate coverage of the curriculum in line with the behavioural objectives for the senior secondary school students.

In order to ensure the appropriateness of the instrument, the constructed Physics achievement test was given to three experience physics teachers and two physics lecturers for item appropriateness based on the new physics curriculum. They made their criticisms and corrections on some of the physics items to suit the physics new curriculum. All their corrections were incorporated into the final instrument to suit the physics new curriculum and the higher cognitive order of the bloom taxonomy for appropriateness.

The Kuder-Richardson formula 20 (KR20) was used in establishing a measure of internal consistency of the test. The test was administered to fifty students that were not part of the study for the purpose of reliability. A reliability index of 0.93 was obtained from the computation. From the index obtained, the PAT has a high reliability and should be used.

The researcher administered the test with the help of the Physics teachers in the sampled schools and two research assistants. After the administration of the Physics Achievement Test (PAT), the answer scripts were collected for marking. Each item was scored either right or wrong by the researcher.

The test items scored were entered into the computer for the purpose of analysis. The chi-square statistics was used to identify the items that fitted the one parameter (1PLM), two parameter (2PLM) and the three parameter (3PLM) logistic model at 0.05 alpha level. The IRT Bayesian Modal Estimator (BME) software and the statistical packages for social sciences (SPSS) were used for the analysis.

The physics achievement test was limited to 50 items because this length makes it suitable for application within a classroom period of one hour thirty minutes in a school setting. Only Physics Achievement Test items with a benchmark fit statistic of .606 were included in the final form of the Physics Achievement Test using the principal component of factor analysis and the item difficulty values were taken into consideration to ensure a reasonable spread of items along the entire test. This is important because the measurement of the varied ability levels in the target population requires a similar variation in item difficulties.

Presentation of Results

Research Question 1

What is the validity of Physics Achievement Test?

Table 1: Computation of the Physics Achievement Test Blue Print for Validity.

Content Area	**Analysis 40%**	**Synthesis 34 %**	**Evaluation 26%**	**Total 100%**
Interaction of matter, Space and time 30%	6	5	4	15
Conservation principles and wave 25%	5	4	3	12
Energy Quantization and Duality of matter 20%	4	3	3	10
Fields 15%	3	3	2	8
Physics in Technology 10%	2	2	1	5
Total 100%	20	17	13	50

The PAT items were checked with the table of specification to ensure that the test showed the desired proportion of emphasis. The weightings of the content areas were I: = 30%, II = 25%, Ill = 20%, IV = 15% and V = 10%. For the objective domain, the weightings were: analysis = 40%, synthesis = 34% and evaluation = 26%. The rationale for the differential weightings of the content areas and the objective domains was based on the relative emphasis on the content areas and the objective as outlined in the Senior Secondary School Physics curriculum, as well as on interviews the researcher had with a cross-section of physics teachers. Thus, the agreement among the five judges and the distribution of items to the objectives and content areas are enough evidence that the Physics Achievement Test has a high content validity. The face validity was also considered appropriate during the expert judgment.

Research Question 2

What is the reliability of Physics Achievement Test?

Table 2: Physics Achievement Test Reliability Using Kuder-Richardson Formula 20

No. of students	No. of item	KR-20	Mean	SEM	Decision
50	50	0.93	30.56	0.79	High

Table 2 shows the reliability of the Physics Achievement Test (PAT) using Kuder-Richardson Formula 20. This approach became necessary because the PAT is a single multiple choice test with expected response of either pass (1) or fail (0). The result of the analysis reveals that n is 50 and the Kuder-Richardson 20 reliability coefficient of 0.93 was obtained as a measure of internal consistency. The result indicated that the PAT has a high reliability and should be used.

Research Question 3

What is the difficulty level of Physics Achievement Test in Delta and Bayelsa States?

Table 3: Item difficulty level Analysis of Physics achievement test in Delta and Bayelsa States.

Item	Threshold	Remark	Item	Threshold	Remark
ITEM1	0.205	Difficult	ITEM29	0.230	Difficult
ITEM2	0.442	Moderate	ITEM30	0.025	Difficult
ITEM3	0.831	Easy	ITEM31	0.014	Difficult
ITEM4	0.481	Moderate	ITEM32	0.568	Moderate
ITEM5	0.363	Difficult	ITEM33	0.305	Difficult
ITEM6	0.663	Easy	ITEM34	0.254	Difficult
ITEM7	0.230	Difficult	ITEM35	0.090	Difficult
ITEM8	0.413	Moderate	ITEM36	0.832	Easy
ITEM9	0.568	Moderate	ITEM37	0.740	Easy
ITEM10	0.288	Difficult	ITEM38	0.541	Moderate
ITEM11	0.305	Difficult	ITEM39	0.832	Easy
ITEM12	0.554	Moderate	ITEM40	0.460	Moderate
ITEM13	0.813	Easy	ITEM41	0.813	Easy
ITEM14	0.201	Difficult	ITEM42	0.850	Easy
ITEM15	0.813	Easy	ITEM43	0.073	Difficult
ITEM16	0.723	Easy	ITEM44	0.629	Easy
ITEM 17	0.892	Easy	ITEM45	0.140	Difficult
ITEM18	0.813	Easy	ITEM46	0.887	Easy
ITEM19	0.554	Moderate	ITEM47	0.715	Easy

ITEM20	0.461	Moderate	ITEM48	0.209	Difficult
ITEM21	0.844	Easy	ITEM49	0.413	Moderate
ITEM22	0.009	Difficult	ITEM50	0.810	Easy
ITEM23	0.073	Difficult			
ITEM24	0.494	Moderate			
ITEM25	0.175	Difficult			
ITEM26	0.123	Difficult			
ITEM27	0.663	Easy			
ITEM28	0.443	Moderate			

Table 3 reveals the difficulty level of the Physics achievement test based on three parameter logistic (3pl) model. The data showed that items were ranged from 0.009 to 0.892. The items were classified as Difficult, Moderate, or Easy. The criterion for this classification was based on guidelines stipulated by Hambleton and Swaminathan (2013), DeMars (2010), and Santelices and Wilson (2012). Items with difficulty levels from 0.00 to 0.39 were classified as difficult, those with difficulty levels from 0.40 to 0.59 were classified as moderate and those with difficulty levels from 0.60 to 1.00 were classified as easy.

Based on the data in Table 3, nineteen items 38% (item numbers 1, 5, 7, 10, 11, 14, 22, 23, 25, 26, 29, 30, 31, 33, 34, 35, 43, 45 and 48) were difficult, thirteen items 26% (item numbers 2, 4, 8, 9, 12, 19, 20, 24, 28, 32, 38, 40 and 49) were moderate, while the remaining eighteen (18) items (36%) that is items 3, 6, 13, 15, 16, 17, 18, 21, 27, 36, 37, 39, 41, 42, 44, 46, 47 and 50 were easy.

Discussion of Results

The Table 1 result shows a very good distribution of the Physics 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 higher order cognitive domain and on content areas of senior secondary school Physics 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 evidence that the Physics Achievement Test has a high content validity. The judges helped to establish face validity. Face validity is concerned with judgment about the test after constructions (Aliyu, 2015). 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 (Seibert and Brendefur 2018 and Akuche and Aliyu (2019). 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. It should be noted that objectivity is a major concern in the measurement of test quality. This implies that the PAT has a high content and face validity while the findings of this study do not agree with that of Abhuegbeude (2016) and Nworgu (2003)

The reliability result was presented in Table 2, through Kuder-Richardson formula 20 analyses. The result shows that a reliability estimate of 0.93 was obtained. This helped to establish the internal consistency of the Physics Achievement Test. The result reveals that the PAT is highly reliable and could be used for senior secondary three (SS III) students when they cover the SSCE Physics syllabus. The use of KR20 in this study was appropriate, since the items are scored dichotomously; there is need to use KR20 in computation (Osadebe, 2013; Osadebe, 2014). The reliability estimate of this study (.93) is above the .86 of Aliyu and Akinoso (2020); and .85 of Aliyu (2015) and .82 of Akuche and Aliyu (2019). However, the PAT reliability estimate measured up to the standard recommended by some experts. The result of this study is in line with the findings of Aliyu (2015); Akuche and Aliyu (2019) and Aliyu and Akinoso (2020) while the findings of this study do not support that of Abhuegbeude (2016).

Results of Research Question 3 showed that the items ranged in difficulty from 0.009 to 0.892. The result showed that 19 items (38%) were difficult with value ranged from 0.00 to 0.39; 13 items (26%) were moderate with value ranged from 0.40 to 0.59 and 18 items (36%) were easy with value ranged from 0.60 to 1.00. The table

showed that the items had positive difficulty estimates, suggesting that the physics achievement test covers a wide spectrum of ability of senior secondary school physics students. The results support previous findings (Ani, 2014; Nkpone, 2001; Akuche and Aliyu (2019 and Aliyu, 2015). According to Siebert and Brendefur (2018), item difficulties lower than - 2.5, and higher than +2.5 are usually not useful for achievement testing. The difficulty estimates, therefore, fall within the recommended range. The finding also agrees with the result of Chong (2013) that difficulty parameter or the threshold parameter value tells us how easy or how difficult an item is. The finding of this study corresponds with Obinne (2008) that negative difficulty estimates indicate that the items are easy while positive difficulty estimates indicate that the items are hard while the findings of this study disagree with that of Abhuegbeude (2016) and Orluwene and Ukwuije (2009).

Conclusion

Based on the findings of this study, the researcher drew the following conclusion:

The Physics achievement test constructed by the researcher had high and relevant psychometric properties of a good test. As such it would serve as an effective formative evaluation instrument for Physics students. The study also concluded that students in Delta and Bayelsa States need more preparation for the students to score higher. The validated Physics achievement test did not have high discrimination index for senior secondary school three students.

Recommendations

Based on the findings of this study, it is hereby recommended that:

1. The development and validation of the Physics achievement test for senior secondary school students should consider the use of the IRT three-parameter logistic model.
2. Secondary school students should take the study of Physics more serious to obtain a higher score in Physics.
3. Testing instrument should meet or possess the required standard regarding psychometric properties for item construction.

References

1. *Abhuegbeude, A.B.I. (2016) Development and Standardization of Mathematics Achievement Test for Middle Basic Education Pupils in Delta and Edo States. Unpublished Ph.D thesis Delta State University, Abraka.*
2. *Adegoke, B. A. (2013) Comparison of Item Statistics of Physics Achievement Test using Classical Test and Item Response Theory Frameworks. Journal of Education and Practice,4(22); 87-96.*
3. *Akuche, U. E. and Aliyu, R. T. (2019) Examining the Relationship between 1-PL of Rasch and 3-PL Models of IRT in Item Selection in a Constructed Test for Assessment. International Journal of Operational Research in Management, Social Sciences & Education, 5(2); 204-215.*
4. *Aliyu, R. T. &Akinoso, S. O. (2020). Examining the relationship between the Rasch and 2-pl models of IRT in selecting items in a constructed test for curriculum development. Nigerian Online Journal ofEducational Sciences and Technology (NOJEST), 1 (1), 77-98.*
5. *Aliyu, R. T. (2015). Development and validation of Mathematics Achievement Test using the Rasch Model. An Unpublished Ph.d thesis in Delta State University, Abraka.*
6. *Aliyu, R. T., and Akuche, U. E. (2019) Assessment of Winsteps and Bilog MG3 softwares in detecting item parameters in physics achievement test. Abacus (Mathematics Education Series) 44(1); 366-381.*
7. *Ani, E.N. (2014) Application of Item Response Theory in the Development and Validation of Multiple Choice Test in Economics. Unpublished M.Ed. Dissertation, University of Nigeria Nsukka. Nigeria.*

8. *Omorogiuwa, K. O. &Iro-Aghedo, E. P. (2016) Determination of differential item functioning by gender in the national business and technical examinations board 2015 mathematics multiple choice examination. International Journal of Education, Learning and Development. .4.10, 25-35. www.eajournals.org.*
9. *Osadebe, P.U (2012) Procedures for Construction, Validation and Standardization of Test. A Seminar Paper Presented at Delta State University, Abraka.*
10. *Osadebe, P.U. (2013). Evaluation Techniques, DELSU. Journal of Educational Research and Development, 12(1), 56-63.*
11. *Osadebe, P.U. (2014) Construction of Economics Achievement Test for Assessment of Students. World Journal of Education, 4(2), 58-64.*
12. *Osadebe, P. U., &Nwabeze, C. P. (2018). Construction and validation of Physics aptitude test as an assessment tool for senior secondary school students. International Journal of Assessment Tools in Education, 5(3), 461-473.*
13. *Siebert, C.F. and Brendefur, J.L. (2018) Examining Validity and Reliability of a Mathematics Assessment Tool for K-2 Students. Journal of Education and Development, 2(1); 31-44.*
14. *Spike, E.S. (2005). Sex Difference in Intrinsic Aptitude for Mathematics and Science. Havard university press*

Corresponding Email: ***chuks01nwa@gmail.com***