World Journal of Research and Review (WJRR)
ISSN:2455-3956, Volume-7, Issue-5, November 2018 Pages 16-22

Assessing Perceived Efficiency of Tailored Testing and Its Useability over Uniform Testing in Measuring Achievement of Secondary School Geography Students in Benue North-East Education Zone

Yeke C. M., Ogboji J. M., Akaakase T.

***Abstract*— This study assessed perceived efficiency of tailored testing and its usability over uniform testing in measuring achievement of Secondary School Geography Students in Benue North – East Education Zone. The study adopted experimental research design. The sample size for the study was 266 secondary school Geography students representing 34.6% of the population drawn from 877 Geography students using Taro Yamen's formula drawn from 238 schools that offer Geography out of 1046 secondary school in the Zone. A 50 item National Examination Council Geography Past Question Paper was adopted. Also 50 item standardized achievement test, a 50 item teacher-made achievement test and tailored tests were developed and used for data collection. The results revealed that there was a significant difference in the mean achievement scores of students using uniform test and tailored test items. Based on the findings it was concluded that when instruments of items tailored to individual ability of the testees are used, there is the tendency that the students' true ability can be revealed. It was recommended that the Government should organize workshops to train teachers on item construction with particular reference to tailored testing to instill in the examiners the culture of testing students based on their individual ability, School owners should make introduction of item banking necessary in all post primary schools in order to enhance validity of testing in the school system, Relevant examination bodies like NECO, WAEC, and NABTEB should ensure that only experts are involved in item development to make the test items valid and reliable. Conclusions were drawn and suggestions for further studies were made.**

***Index Terms*— Perceived efficiency, tailored testing, uniform testing,school Geography students, Benue North – East.**

I. INTRODUCTION

The knowledge of capability of a person on a job or the ability of a student to cope with a specified task at any given level of study is of great importance to any institution that is progress - driven. Therefore, there is need for the right instrument to be used in measuring the ability of persons before they are assigned a task or promoted to the next level. When a wrong instrument is used in measuring the ability of someone, it is obvious that the result will be invalid and unreliable (Useni, 2012). There are therefore, a number of instruments that can be used in measuring one's achievement. These are standardized test, teacher – made test as well as tailored test.

Historically, testing as we know it today, had its roots in the earliest attempts of man to measure certain obvious human characteristics and behaviours. The history of testing movement is the record of the extension and gradual refinement of the measures used. As man later turned his attention to the more intangible element of human behaviour (trait), the problems of testing became more complex, (Gerberich, 1963).

Test could mean different things to different people depending on the area of interest. For instance, in Medicine, test means medical examination conducted in order to check or discover somebody's health condition (Timothy & Gray, 2009). It could also be seen as an activity or instrument used or intended to find out whether somebody or something has the required qualities (Rouse, 2007). However, Onuka (2013) sees test as a systematic procedure for observing a person's behaviour and describing it with the aid of numerical scale or category. In a different view, Odinko (2014) sees test as an experiment intended to show that something works or works well. A test therefore, is a measurement instrument or device administered to someone to determine the relative value of the trait or skills which include cognitive, affective and psychomotor skills. For instance, a spelling test measures how well someone spells or the extent to which someone has learned how to spell a specific list of words (cognitive skills) (Alonge, 2002). A Geography interest test measures the extent to which someone has interest, likeness towards Geography (affective skills), while a test on filling the burette with acid measures someone's dexterity in filling the burette with acid (psychomotor skills).

Test has several forms or types as viewed by several authors. In a classification made by Emaikwu (2011), Adejoh and Obinne (2013), test could be classified based on the number of items to be assessed, the mode of response as well as the purpose. According to the authors, classification based on both the purpose of test and the method of construction involved achievement tests such as standardized achievement test, teacher made achievement test among others.

Yeke C. M., Department of Educational Foundations, Federal University of Agriculture, P.M.B. 2373,Makurdi, Nigeria
Ogboji J. M.,Department of Educational Foundations, Federal University of Agriculture, P.M.B. 2373,Makurdi, Nigeria
Akaakase T.,Department of Educational Foundations, Federal University of Agriculture, P.M.B. 2373,Makurdi, Nigeria

Achievement test according to Emaikwu (2011) is an instrument designed to measure the relative accomplishment in a specified area of work. While Ogomaka, Onah and Amadi (2017) regard achievement test as one which provides a satisfactory psychometric properties of the concept it was designed to measure. According to the authors, the widely used instrument by psychometricians and psychologists to provide specific objectivity in measuring ones achievement is standardized achievement test.

Standardized achievement test according to Kolawole (2001) is the test constructed, administered and marked by experts. It however, has universal applicability with defined objectives syllabus which is common to all its users. In a related view, Anikweze (2012) sees it as test in which the procedure, apparatus and scoring have been fixed so that precisely the same teaching procedures can be followed at different times in different locations. In a similar view, Adikwu, Aduloju and Agi (2016) consider standardized achievement test as one designed to obtain and interpret scores in some objective form which can be used to evaluate individual testee's performance based on an acceptable standard. It is therefore, a test carefully constructed by expert test developers. Thus, standardized achievement test significantly differs from teacher-made achievement test.

Teacher-made achievement tests according to Akem and Aduloju (2003) are those test usually constructed, administered and scored by the school teachers with little or no outside assistance while Emaikwu (2011) regards teacher-made achievement test as one constructed by the classroom teacher for usage in a particular class under conditions of choice by the teacher. According to the author, this kind of test tries to measure students' achievement over a period of time for the purpose of achieving some stated behavioural objectives. In a related view, Aminu (2014) considers it as the classroom test designed, administered and scored by those subject teachers teaching in the various classes. The author added, constructing valid and reliable test items for the classroom is a time consuming process. It requires teachers to thoroughly consider the content, goal and outcomes of the assessment results which always vary in-line with the variation that exists among test takers. Therefore, tailored tests have been advocated to accommodate the various ability levels among learners or testees in every given classroom.

When test items that tally with the individual ability of the testees are constructed tailored test is inferred. Meanwhile, Tailored Testing is a process that systematically matches the difficulty of the test items with the abilities of the examinee being tested(Stage, 2003). It is also a process in which different sets of test questions (items) are administered to different individuals depending on each individual's status on the trait being measured. Tailored testing has also been variously referred to as adaptive, programmed, response contingent, computerized, automated, individualized, branch and sequential testing (Emaikwu, 2011). The author added that, Tailored testing can also result in efficient and more accurate mastery classifications and provides an efficient and practical approach to the measurement of individual change. Tailored testing solves the problem by selecting from an item bank for each individual a test designed of items which are appropriate in difficulty level for each examinee. This is in line with the idea of Item Response Theory (IRT).

According to Emaikwu (2011), Ukozor and Ukwuoma (2017)Item Response Theory deals with the characters of an item with respect to item difficulty, item discrimination as well as item response pattern of examinee. It is interested in determining what a particular examinee might do when confronted with test item. Such information is necessary if the test designer desires to predict test scores characteristics in one or more population of examinee. An appealing feature of item response theory is that, with its application, once an examinee ability level has been established, it is possible to determine the probability of a correct response to an item the examinee has never taken assuming that certain item parameters have already been determined. According to the authors , in tailored testing, one or more items are administered to an examinee, and scored correct or incorrect. Based on the response of the examinee, additional items are selected from an item bank with items of known difficulties and discrimination. Therefore, a routine testing is done to establish the individual abilities of the testees. So, items selected for administration to the examinees during the process of testing are selected to be those in the items bank, which are most appropriate for measuring those individuals, primarily in terms of their difficulties. In this way, the items are adapted to the characteristics of the examinees during this process of testing. The items administered to each individual are those that are neither too easy nor too difficult for the individual bearing in mind the possibility of having both the high and low ability students in the same class.

High ability learners are easily identified in most cases. In his view, Denga (2003) sees the high ability as those students that have high intelligent quotient. According to the author, those with mid-way intelligent quotient are best described as average ability students, while the low ability students are those that naturally possess low intelligent quotient. This idea is in line with the Bloom's taxonomy where knowledge and comprehension fall in the low level, application and analysis are in average ability while synthesis and evaluation fall in the high ability cognitive levels. Therefore, measuring the achievement of students that vary in all they do using tailored tests are expected to be advocated bearing in mind its relevance.

In a related development, relative efficiency has been viewed by Asemah (2010) as the capability of a special application of effort to produce a specific outcome with a minimum amount or quantity of waste expenses or unnecessary effort in relation or proportion to something else. Thus, tailored test will be explored in relation to other traditional tests such as standardized achievement and teacher-made tests among others, with the aim of establishing its efficiency and effectiveness in measuring students' achievement.

A comparison of standardized and teacher-made achievement tests was made by Emaikwu (2011), according to the author, both are achievement tests, they are made of

uniform test items and are also classified based on purpose and method of construction. In the same view, a tailored test is also an achievement test and it can be classified based on purpose and method of construction in which case individual ability of the testees is taken into account. Therefore, there exist a difference between the traditional forms of test and tailored test which has the tendency to introduce new idea in testing or assessment in our schools system.

Traditional methods of testing are those types and forms of testing such as standardized achievement, teacher-made achievement tests which make use of uniform items that are commonly used by most institutions of learning and other examination organizations to elicit information about the testees.

Several traditional testing techniques such as aptitude testing, intelligence testing, ability testing among others have been identified and put to use all in an attempt to ensure that correct measures are taken of testees. However, it has been observed that, all these traditional methods of testing are carried out using uniform or parallel test items. This perhaps is the reason why they do not satisfactorily measure the true ability of the individual. It is important to note that students or individuals vary in all that they do.

Uniform test items are items of equal strength and fixed for all testees regardless of their level of ability. Standardized achievement tests and teacher–made achievement test also use uniform test items. Examples of teacher-made test are test taken by students during Semester Examination (SE), terminal examinations in both primary and secondary schools while Standardized Achievement test are tests conducted by Joint Admission and Matriculation Board (JAMB), National Examination Council (NECO), West African Examination Senior School Certificate Examination (WASSCE) as well as National Board for Technical Examination Board (NABTEB).

As highlighted by Orluwene (2012), test plays very many crucial roles not just in educational development but also other facets of the nation. These include:

Prediction or forecasting how well an individual will perform in future on a job or task, selection of people into a programme, certification of people after completing an educational programme, classification or assigning people or students into different class of abilities, placement or assigning individuals to the appropriate levels in a programme or discipline based on their performances, ability, interests and aptitudes, evaluation or determine the worth of a programme, source of motivation, improve programmes and curricula, provide feedback or knowledge of results, diagnostic and remedial decisions, guidance among others. These roles facilitate effective measurement.

Measurement can also be conducted by different disciplines in several ways but with hope of obtaining a numeric value . For instance, in the physical sciences, it involves the use of precise units or measures which are used for measurement. Such units and measures include centimeter or metres for distance, kilogrammes for mass weights and decilitres or litre for capacity. In education, Harbor-Peters (1999) sees measurement as the quantitative description of pupils' change in behaviour. It does not imply judgments concerning the worth or value of the behaviour measured. Harbor-Peters listed measuring instruments in education as test, class work, assignments, and projects. In a different view, Anikweze (2012) defines measurement as a way of establishing standard. The author however, distinguishes several aspects of measurement which result from encounters between teachers and learners in schools system. They include: Physical measurement, Psychological measurements, Personality measurement as well as Educational measurement. All the above measurements can therefore, be conducted using tailored test to ascertain the testees' achievement.

Achievement on other hand is the measurable impact made by a programme of study on a student (Atser&Kyum, 2000). The authors added, if the instruments are of uniform strength, there will be the tendency of having wrong information about the testees' achievement. This is because of the variation that exists among students in their intellectual ability. Supporting the view Igeh, (2004) sees achievement as a relative accomplishment in a specified area of work. Students' relative accomplishment on a course of study can therefore be measured using teacher-made achievement test, standardized achievement test as well as tailored test. The author further submitted that, using uniform test items cannot provide adequate and reliable information about the ability of the testees. According to the author, any judgment passed on the students' achievement using parallel or uniform test items remains bias because the test does not take cognizance of individual ability of the testees. The author however, advocated use of tailored test based on research findings. The researcher therefore, sees the need to explore this testing technique in a greater detail using Geography as a school subject. The choice of geography as a subject is informed by its relevance in almost all disciplines. For instance, Agriculture, Biology, Physics, Chemistry, Economics, Statistics among other subjects are studied in Geography. Therefore, using Geography is as good as considering almost all the subjects at post primary school level. Also Physics, Chemistry and Biology form a group of core subjects in the National Policy on Education, (National Policy on Education, 2004). These subjects are all studied in Geography. This explains how indispensible Geography is in post primary school level.

A survey of students' performance for the past ten years shows average achievements except recently. For instance, Statistics shows that 45.7% of students passed Geography during the 2013 examination year, 23.2% in 2014, 17.1% in 2015, 28.4% in 2016 while 20.3% passed geography in 2017 (NECO Result Analysis Report of 2013, 2014, 2015, 2016 and 2017). The question is; could the use of inappropriate testing techniques or test items for the testees be responsible for the poor performances? Also there is a general cry among measurement experts as to how one's achievement can be measured using the right instrument. They have decried the challenge of getting the right instrument that can ascertain the knowledge, skills and ability of people in a specified area.

The curiosity of the researcher revealed that, most teachers in our schools have little or no knowledge of tailored testing and therefore consider its usability unrealistic. Thus, teacher

competence in using suitable instruments on the testee seems to significantly contributes to challenges bedeviling measurement system various institutions in the country.

This challenge, no doubt can be responsible for the corruption we see today at all levels of governance and poor performance of students in Nigerian institutions of learning. Infrastructure in the country decay very fast, partly because they were made either with substandard materials or unqualified hands (Asemah, 2010). Also, when people's competencies are not properly tested before being assigned jobs, there is the tendency of recording awful cases. For instance, the cases of plane crashes and other mishaps as well as examination malpractice which has become rampant in not just post primary schools but also in all institutions of higher learning, are some of the attendant consequences. For instance, 37% of results were withheld by WAEC and 20% by NECO conducted in 2017 due to examination malpractice (Benue State Examination Board Result Analysis Report, 2017). The UTME conducted by JAMB in 2017 recorded 61% failure in Mathematics, English Language and Geography (JAMB Result Analysis Report, 2014). Generally, there has been poor performance in the national examinations for the past five (5) years, (NECO Result Analysis Report 2013, 2014 , 2015, 2016 and 2017). This could partly be traceable to inappropriate testing or measurement of the students' achievement which leaves them with the option that makes them feel the only way to pass examination is through cheating. There is therefore the need to device a more appropriate way of testing achievement of not just Geography students but students from other disciplines also. The uniform test conducted for recruitment and promotion in our schools and organizations do not measure the true ability of the individuals, hence the need to test people's ability using tailored testing.

In study conducted by Igeh (2004) Tailored testing proved reliable in measuring the true ability of Mathematics students than other methods such as, aptitude testing, intelligent testing, personality testing, objective testing, multiple choice testing, true or false testing as well as paired testing probably because, they all use uniform test items. With the revelation by the Statistics, the poor performance of Students in Secondary School in Zone A in Geography is alarming. This calls for immediate attention (Benue State Examination Board 2013, 2014, 2015, 2016 and 2017) since other researchers have studied probable causes of this poor performance in the area of teachers motivation, use of different methodologies among others. It is against this background that this work will be based, giving that much research work is needed to establish empirically the posibilty of using tailored test items over uniform test items in measuring testees ability in Geography. This study is set to explore the possibility of using tailored test because this method is a new innovation and has not attracted much attention in the field of measurement particularly in Geography as revealed from studies reviewed by the researcher.

II. Statement of the Problem

Every student who really studied for six years in post primary school should not have problem passing his/her final examination such as WAEC, NABTEB, NECO and JAMB, because it is expected that a student in Senior Secondary 3 (SS 3) should be conversant with the content of the subject which should enhance his/her performance at the national examinations. Furthermore, to ensure that the real performance of the students is revealed, different modes of testing and assessments have been advocated by (Useni, 2012).

However, despite the attempts made by the use of other forms of tests such as standardized achievement test, teacher-made achievement test among others which do not take cognizance of the testees' individual ability,but use uniform test items. This probably might be responsible for the recent failure or poor performances recorded by students in Geography subject so that 45.7% of students passed Geography during the 2013 examination year, 23.2% in 2014, and 17.1% passed in 2015. 28.4% in 2016 while 20.3% passed geography in 2017. The problem of the study is; could the use of items tailored to the individual ability of the testees enhance their individual achievements in Geography? On this premise therefore, this study is set to investigate relative efficiency of tailored, standardized, and teacher-made test in measuring achievement of secondary school Geography students in Zone 'A' Senatorial district of Benue State.

III. Purpose of the Study

This study is aim at assessing perceived efficiency of tailored testing and its usability over uniform testing in measuring achievement of secondary school Geography students. Specifically, the study seeks to find out:

1. The difference in mean achievement scores of students using standardized test items and tailored test items
2. The difference in mean achievement scores of students using teacher – made test items and tailored test items

Research Questions

The following research questions were raised to guide the study.

1. What is the difference in mean achievement scores of students using standardized test items and tailored test items?
2. What is the difference in mean achievement scores of students usingteacher – made test items and tailored test items?

Research Hypotheses.

The following null hypotheses were formulated to guide the study and would testedat 0.05 levelof significance.

1. There is no significant difference in mean achievement scores of students using standardized test items and tailored test items
2. There is no significant difference in mean achievement scores of students using teacher – made test items and tailored test items

IV. METHODOLOGY

The study adopted experimental research design. There are eight hundred and seventy seven (877) geography students in the 323 post primary schools in the zone during the study. These 877 geography students constitute the sample frame and population for the study respectively. However, using Taro – Yamen formula, a total number of 399 students formed the sample size. Therefore, using purposive sampling, 57 students were sampled from each of the seven (7) LGAs of the zone. these were first administered the uniform from which ability classes were determined with each class having 133 students. So, a total of 266 geography students were eventually used for the study. The instruments for data collection were standardized achievement test, teacher-made achievement test and tailored test. Descriptive statistics of mean and standard deviation were used to answer the research questions and t-test was used to test the hypothesis at 0.05 alpha levels.

V. RESULTS

Research Question 1

What is the difference in mean achievement scores of students using standardized test and tailored test items

The answer to this research question is presented in Table 1.

Table 1: Mean Achievement Scores of High and Low Ability Students using Uniform Test Items

Group	N	Minimum Score	Maximum Score	Mean Achievement Scores	Standard Deviation
Standardized test	133	32.00	42.00	25.41	9.57
Tailored test	133	60.00	78.00	65.42	6.29
Mean difference				40.01	
Total	266				

Table 1 shows that the students measured with standardized test had a mean achievement score of 25.41 with a standard deviation of 9.57. Those measured with tailored test had a mean achievement score of 65.42 with a standard deviation of 6.29. The difference in the mean achievement scores was 40.01.

Research Hypothesis 1

There is no significant difference in mean achievement scores of students using standardized test and tailored test items.

The t-test result of this research hypothesis is shown in Table 2

Table 2: t-test Result of mean achievement scores of students using standardized and tailored test items

Group	T	Df	Sig
Standardized test	44.2	132	.00
Tailored test	54.2	132	.00

The result on Table 2 revealed that t value = 44.2 and 54.2 at df = 132 while p(sig) = .00. Therefore, since p< .05 the null hypothesis which states that there is no significant difference in mean achievement scores of students using standardized achievement test and tailored test items is not accepted.

Research Question 2

What is the difference in mean achievement scores of students using teacher – made test and tailored test items?

The answer to this research question is presented in Table 3.

Table 3: Mean Achievement Scores of Students using teacher-made test and Tailored Test items.

Group	N	Minimum Score	Maximum Score	Mean Achievement Scores	Standard Deviation
Teacher-made test	133	46.00	38.00	42.87	8.94
Tailored test	133	74.00	86.00	68.59	5.74
Mean difference				25.72	
Total	266				

Table 3 shows that, the mean achievement scores of 42.87 and a standard deviation of 68.59 were recorded for the students measured with teacher-made test. The Table also shows that, the students measured with tailored test items had a mean achievement score of 68.59 with a standard deviation of 8.74. The difference in mean achievement scores was 4.28.

Research Hypothesis 2

There is no significant difference in mean achievement scores of students using teacher-made test and tailored test items.

The t-test result of this research hypothesis is shown in Table 4

Table 4: t-test Result of Students Measured Using teacher-made and Tailored Test items

Group	T	Df	sig
Teacher-made test	22.5	132	.00
Tailored test	30.1	132	.00

Table 4 shows that the efficiency of measuring the achievement of students using tailored test items over uniform test is P = .00 and .00 at df=132 and t value = 22.5 and 30.1 Therefore, since P >.05 the null hypothesis which states that there is no significant difference in mean achievement scores of students using teacher-made test and Tailored Test is not accepted.

VI. Discussion

The result in Table 1 showed that, the mean achievement scores were 25.41 and 65.42 with corresponding standard deviations of 9.57 and 6.29. The difference in the mean deviation scores was 40.01. This is an indication that the performance of the students tested with a tailored test was higher than students tested with standardized test as the test items were uniform and could not their individual ability. The test of hypothesis in Table 2 revealed that, there was significant difference in mean achievement scores of students using standardized test and tailored test items. This implies that, use of uniform test items put several students especially the low ability ones on a disadvantaged part in any given examination. The finding is in line with findings of Adoga (2001) and Azu (2012) who found that use of uniform test items that do not take cognizance of the individual ability of the testee is not in the best interest of the education industry. In a study conducted by Stage (2003), it was discovered that human beings are different in every ramification therefore, testing someone with an instrument that match his/her individual ability was better, noting that it reduces the menace of examination malpractice. On the contrary, the findings disagree with that of Tim and Duke (2002) who revealed that, commitment to studies enables students perform well in any given task irrespective of their individual ability. This result therefore implies that, the uniform test items were too difficult for the low ability students which suggest reason for the poor performance; hence the hypothesis was not accepted.

The result in Table 3 showed that, the mean achievement scores was 42.87 and 68.59 with the standard deviation of 8.94.and 5.74. The mean difference was 25.27. Thus, the test of hypothesis in Table 4 revealed that there was significant difference in the mean achievement scores of the students. This is an indication that students improved in their performance as the items were tailored to their individual ability. While the teacher-made test which is also of parallel items recorded poor performance. This finding is in line with Igeh (2004) who found that, testing students with test items that tally with their individual ability enhances their achievement. However in a contrary view, Ter and Uma (2001) observed that tailoring items to the individual ability of the testees might only encourage laziness among students. From the result it is clear that, the students' performance significantly improved using tailored testing, which implies that if schools can discourage the use of traditional way of testing students with uniform test items, there would be an improvement in the performance of students in every given examination. Therefore, the null hypothesis was not accepted.

VII. Conclusion

It can therefore be concluded that, if students are tested using items that tally with their individual abilities, the best of their skills can be discovered especially in their cognitive domain. This will be of value not just to the educational industry but also the society at large.

VIII. Recommendations

Based on the findings, the following recommendations were made:

1. Government should organized work-shops to train teachers on item construction with particular reference to tailored testing to instill in the examiners the culture of testing students based on their individual ability.
2. School owners should make introduction of item banking necessary in all post primary schools in order to enhance validity of testing in the school system.
3. Relevant examination bodies like NECO, WAEC and NABTEB should ensure that only experts are involved in item development to make the test items valid and reliable.

References

[1] Adejoh, M.J., &Obinne, A.D.E. (2013).*Basic Issues in Test, Measurement and Evaluation.*Makurdi: Asofad Printing Ltd.

[2] Adikwu, O., Aduloju, M.O., &Agi, C.I. (2016).*Measurement and Evaluation in Education.*Makurdi: Shibboleth Prints.

[3] Adoga, S.O. (2001). Effectiveness of Tailored Testing in Measuring Achievement of Junior Secondary 3 in Mathematics Examination in Makurdi Local Government Area.*Abacus Journal of Mathematical Association of Nigeria. 48(2)*

[4] Akem, J.A., &Aduloju, M.O. (2003).*Principles of Measurement and Evaluation Continuous Assessment and Psychological Testing in Education.*Makurdi: Confidence Books Ltd.

[5] Alonge, M.F. (2002).*Measurement and Evaluation in Education and Psychology.*Ado-Ekiti: Adedayo printing Nig. Ltd.

[6] Anikweze, C.M. (2012). *Measurement and Evaluation: For Teacher Education.* Ibadan: Malijoe Soft Print.

[7] Asemah, J.I. (2010).*Perspectives in Educational Management and Administration.*Makurdi: Destiny Ventures.

[8] Atser, V.M., &Kyum, K. (2000).Validity of Tailored Testing in Post UTME in BSU.*Journal of Education and Technology. 3(2)*

[9] Azu, C.Y. (2012).*Principles of Testing in Schools and Collages.*Awka: Onitsha Press.

[10] Benue State Examination Board (2017). *Result Analysis Report.* Makurdi: Unpublished.

[11] Benue State Examination Board (2017).*Annual Examination Registration Records.*Makurdi: Unpublished.

[12] Benue State Ministry of Education (2017). *Annual School Supervision Report: Makurdi*: Unpublished.

[13] Denga, D.I. (2003).*Educational Measurement, Continuous Assessment and Psychological testing.*Calabar: Glad Tidings Press Ltd.

[14] Emaikwu, S.O (2011).*Fundamentals of Test, Measurement and Evaluation with Psychometric Theories*, Makurdi: SAP Ltd.

[15] Ezeachi, C (2004). *National Policy on Education4th Edition,* Nigeria: GNO Press.

[16] Gerberich, J.R. (1963).*The Development of Educational Testing.Retrieved from http:// www.Jstor.org/stable /1475207 on 24/09/15*.
[17] Harbor-Peters, V.F.A. (1999).*Note Worthy Points on Measurement and Evaluation*. Enugu: Snaap Press Ltd.
[18] Igeh, C.I. (2004). The use of Achievement Test in measuring achievement of low and high ability mathematics students in Wukari LGA of Taraba State. *Journal of Education and Technology 3(2)*
[19] Kolawole, E.B. (2001).*Test and Measurement.*Ado-Ekiti: Yemi Prints and Publishing Services.
[20] National Examination Council (2013, 2014 & 2015). *Result Analysis Report.* Makurdi Zonal Office: Unpublished.
[21] Odinko, M.N. (2014).*Evaluation Research Theory and Practice. Oyo*: Giraffe Books.
[22] Onuka, A.O.U. (2013).*Issues in Contemporary Evaluation.*Ibadan: Nigeria PG Press.
[23] Orluwene, G.W. (2012).*Fundamentals of testing and non-testing tools in Educational psychology*. Ibadan: Harey publications coy.
[24] Rouse, P. (2007).Performance testing.*Retrieved from http://www.searchoftwarequality.techtarget.com/de on 22/09/15*.
[25] Stage, C. (2003). Classical Test Theory or item Response Theory: The Swedish Experience. *Retrieved from www.cepchile.cl 24/06/2015.*
[26] Ter, V.M., & Uma, C.M. (2001). Relative Efficiency of Teacher – Made Achievement Test in Assessing Performance of Senior Secondary 2 Students in English Language in Otukpo LGA.*Abacus Journal of Mathematical Association of Nigeria. 42(1)*
[27] Tim, O.V., & Duke, C.S. (2000).Relative Efficiency of Tailored Testing in Measuring Performance of Mathematics Students in Senior Secondary 2 in Makurdi Local Government Area of Benue State.*Journal of Education and Technology. 11(2)*
[28] Timothy, K., & Gary, C. (2009).Intelligence and aptitude testing.*Retrieved from http://www.education.com/reference /article /apon27/09/15*.
[29] Useni, S.O (2012) *Test, Measurement and Evaluation in education*, Ibadan: Source Publishers.