COMPARATIVE ITEM ANALYSIS OF STUDENTS' WAEC AND NECO MATHEMATICS OBJECTIVES TEST ITEM SCORES .

***Thomas, Edet S., **Uchegbue, Henrietta & ***Ugbe, Emmanuel B.**

***edethomas16@gmail.com ** huchegbue@yahoo.com**
***** beshiunim@gmail.com**

*** 08063762108 ** 07033067374 *** 08030407059**

ABSTRACT

The purpose of this study was to compare psychometric properties of students' WAEC and NECO Mathematics objectives test item scores. Two null hypotheses were formulated to direct the study and literature was reviewed on the variable under study. The research design used was the Ex-post facto (causal comparative) design. A total sample of 250 SS 3 students was selected using stratified random sampling procedure from 10 senior secondary schools for the study. Data collected for the study was done using WAEC and NECO SSCE Mathematics multiple choice papers of 2011. The collected data was analysed using the independent t-test statistics. The result of the data analysed showed that item difficulty index and discrimination index of WAEC significantly differ from that of NECO in terms of items. Based on these findings, it is recommended that proper investigation should be conducted by experts to determine the causes of variations between WAEC and NECO examination from the development stage, administration, scoring of scripts and to the release of results.

Keywords: Item difficulty index, discrimination index, performance in mathematics

INTRODUCTION

Examinations are integral and important part of any educational process. It is a tool employed to assess learner's level of achievement. Most times, it is used in selection of students into secondary or tertiary institutions. In order to achieve the conduct of valid and credible examinations, independent examination bodies were established. These include: The West African Examinations Council (WAEC), The Joint Admissions and Matriculation Board (JAMB), The National Business and Technical Examinations Board (NABTEB), and The National Examinations Council (NECO). The duties of these bodies are drafting of questions, time tabling, administration of examinations, marking, scoring, grading, releasing of results of candidates seeking admission into secondary or tertiary institutions and given admission to students.

However, the examination bodies have had their own share of criticism by most school administrators, Ministry of Education, teachers, students, parents and the general public in the way they conduct their examinations. These criticisms cut across setting of such examination, taking of the examination, marking, grading, release of results and the issuance of certificate. Prominent among these criticisms levelled against these public examination bodies according to Temitope (1999) and Kolawole (2001) were mass leakage of examination papers at times traceable to the officials of the council, unnecessary delay in releasing results, uncontrollable population explosion of the candidates and over load of work as a result of too many

examinations conducted by the council buttressing unreliability of the examination all the school subjects especially Mathematics.

Mathematics cuts across all aspects of human endeavours. This is because man's social, economic, political, geographical and technological life is centred on numbers. In education, Mathematics is the bedrock of all sciences and technologically based subjects. As a core subject, it is offered by all students in schools whether they are science or art inclined and as such it is a " Must-pass" subject for any student seeking admission into any field of study either in universities or other allied tertiary institutions. In recent times, the performance of students over the years in Mathematics has remained an issue of great concern to stakeholders and educators. The poor performance got worse in 2010 when there was a public outcry against the decline in the performance of candidates in Mathematics in both WAEC and NECO examinations. A close look at the results over the years showed a steady deterioration in students' performance as there has been a decline in the number of those who obtained five credits and above including Mathematics which is a basic requirement for transition to higher education. The results revealed that students' performance from 2006 to 2010 were below 50%. It was only in 2008 that students' performance was slightly above average. The above fact implies that less than 50% are qualified to seek for admission into the University and other tertiary institutions (Salman, Mohammed, Ogunlade & Ayinla, 2012).

The importance of Mathematics in admission process has made most students who as a result of fear of failure, see the subject as being difficult and do not have flair for it, to be desperate and as such resort to other forms of examination malpractices. Many people believed that NECO examinations are the most difficult of the examinations being conducted by these examination bodies. Adeniran (2000) claimed that NECO is inferior to WAEC in all standard while Kolawole (2002) concluded that a given 'X' grade in NECO Mathematics test is equal to 'X + 1' grade in WAEC meaning that WAEC is more difficult than NECO. Based on these, this study therefore sought to compare the difficulty and discrimination indices in students' WAEC and NECO Mathematics objective test item scores.

Examination is judged worthwhile when it possesses difficulty and discrimination indices.

Difficulty index is simply the percentage of students taking the test who answered the item correctly. The larger the percentage getting an item right, the easier the item while the smaller the percentage getting an item right, the difficult the item. This process involves counting all the number of test takers who answer each item correctly and converted into proportion or percentages. This can be computed by dividing the number of students who score that item right over the total number of students who attempted the item (Joshua, 2005). The proportion for the item is usually denoted as p-value and is called item difficulty. However, in interpreting difficulty index of an item, many factors need to be considered. Some of these factors as pointed out by Joshua (2005) are the wording of the items, students' learning experiences, and the structure of the subject matter.

Difficulty index, sometimes called percentage passing, according to Anastasi and Urbina (2010) is defined in terms of the percentage (or proportion) of person who answer item correctly. This implies that the easier the items, the higher the difficulty index. They further say that difficulty index is useful in item arrangement i.e. from the simple to complex. This arrangement

gives the test takers confidence in approaching the test and also reduces the likelihood of their wasting much time on items beyond their ability to the neglect of easier item they can correctly complete. Thonstone (as cited in Anagbogu, 2009) stated that indices of item difficulty expressed as percentages or normal curve units are limited to the ability range covered by the sample from which they were obtained.

Bandele and Adewale (2013) carried out a study on comparative analysis of the item difficulty levels of WAEC, NECO and NABTEB Mathematics Achievement Examination in Nigeria. A descriptive research of the survey type that involved the use of a correlation design was used in this study with a sample of 600 final year students randomly drawn from selected Government Technical Colleges and Senior Secondary School from Ondo, Ekiti and Osun States of the South-West Geo-political zone of Nigeria. The sample was then grouped into three homogenous groups of 200 students each. i.e. 200 students from three selected Government Technical Colleges and the remaining 400 students from the selected senior Secondary Schools. The instruments used consisted of adopted WASSCE, NECO and NABTEB Mathematics Examinations. Analysis of variance (ANOVA) was employed to analyse the data. It was revealed that there was no significant difference in the item difficulty levels of WAEC, NECO and NABTEB Mathematics achievement Examinations.

In the same vein, Ojerinde and Alonge cited in Alonge (2003) carried out a research work on the qualities of mathematics classroom Achievement Test and its relationship with an External standardized achievement test of the fifty (50) test items whose difficulty and discriminating indices were found, twenty-eight (28) of them had satisfactory discriminating power (0.22 and above) and all have reasonable difficulty indices (between 0.42 and 0.76). This clearly showed that the test items were 50 percent acceptable. Thus, the findings of this study clearly revealed that item difficulties (i.e. p-value) and discriminating indices (i.e. d-value) of the Mock Mathematics Classroom Achievement Tests were significantly related to those of WAEC.

Also, WAEC (1995) carried out a study on students teachers' and experts' perception of the difficulty of SSCE essay questions in Mathematics, English Language and Chemistry, through a survey. The study involved a total of 287 respondents made up of 260 students, 15 teachers in some selected secondary schools in Lagos Metropolis and 15 experts who are University of Lagos Lecturers in the subject areas. The teachers and experts were requested to rate the questions as "Hard", Medium", or "Easy" while the students in addition to rating, answered the easy questions frequency count, percentages, arithmetic mean, standard deviation and Kendall tau's rank correlation constituted the statistic used for analysis. It was found that student perception of difficulty of examination questions depended on the level of preparedness they have had. The consequences of these findings were centred on improved teaching and learning activities to enable students have confidence in examination with a corresponding better performance. In a detailed result in chemistry, students' perceived examination in Chemistry as moderately difficult based on the rating: hard 44 percent, medium 47 percent and easy as 8 percent.

However, on the same note, experts viewed that the questions are moderately difficult within the scope of average SSCE students. For the teachers, they viewed questions as moderately difficult. For mathematics, students rated mathematics to be moderately difficult. The teachers rating showed that the questions were moderately difficult and the experts equally assessed the questions as moderately difficult. The findings showed that the students' rating

agreed with that of experts, although difficulty is not restricted to objective test. It is easier to access or to estimate in objective test than subjective, because subjective test presents many avenue of expressing same facts thus obstructing the comparison of performance and even the facts. Hence making the determination of the ratio the average of score by all examinees that form part of the measurement easily and objective accessible.

Sotaridona, Pornel and Valleiyo (2003) conducted study on some applications of item response theory to testing and found that the estimates of classical test theory item difficulty showed that the data estimates consistently differs between the two ability groups indicating that an item looks easy when administered to low ability examinees. This shows clearly that the estimates are dependent on the group of examinees who took the test.

Beck (1978) carried out a study to determine the influence of item difficulty on other test item parameters among low achievers. He used a total of 165 third graders as subjects; the instrument used for the study was made up of 4-options 65 multiple choice test items. The test was administered in a normal classroom setting assisted by class teachers in supervision of the conduct of the test. The choice of the item difficulty was made based on National p-value metropolitan standardization. The National p-value was defined as those items having difficulty index value of .40 or lower. The correlation analysis of the data collected showed that there was a significant difference in variability of item difficulties. Further probe showed that higher achievers could answer test items with lower p-value than lower achievers. It could therefore mean that item difficulty can significantly influence other item parameters.

Item discrimination index indicates whether an item differentiates between test takers having varying degrees of knowledge or abilities. Items on scholastic ability test should differentiate between students with higher grade average who should answer an item correctly and more frequently than students with lower grades average. Discrimination may be used in the correlation between scores on the item and scores on the criterion as grades would be the basis for computation. When applied to the teacher made test, such external criterion are not made available. Thus, the total score of the test is used as a criterion. The basic assumption of the discrimination indices therefore is that the test as a whole is an adequate measure of the domain.

To compute the discrimination index, the scores of an individual item (total) scores on the test may result if a student scored high on the test tends to answer the items correctly and those who score low, answer incorrectly, the item test correlation would be positive, but if there is no relation in answering them and the test scores, then the discrimination index would be zero. A rule of thumb will be to at least, require (r) to be .20 or higher (Brown, 1983).

When a group of examinees scores are divided into two or more sub-groups on the basis of the test scores, then the possible discrimination index (D) would be;

$$D = \frac{U-L}{\frac{1}{2}N}$$

Where U = Number of candidates in the upper group who answered the itemcorrectly?

L = Number of candidates in the lower group

who answer the item correctly?

N = Total number of candidates in both the Upper and lower groups

D = Range from .00 to 1.00

Negative values of D indicate that low ability candidates performed in the item than high ability candidates. If the result of the discrimination index is – 1.00, this implies that all candidates in the low ability group scored the item right, while all candidates in the high ability got the item wrong. This means that there is a perfect discrimination, but negatively.

Positive value of (D) indicates that high ability candidates perform better than their low ability counterparts. In this case, if the discrimination is + 1.00, this means that all candidates in the high ability group scored the item right while none in the low ability group scored it right. In this case, it is a perfect discrimination, as it is naturally expected that students with high ability perform better than students with low ability.

When discrimination index indicates 0, this implies that both high ability and low ability groups scored on the item equally. Such an item cannot discriminate, and, as such, is referred to as "dead wood". High discrimination indices are desirable in test development. Ebel and Frisbie (1991) provided a guideline for evaluating item discrimination. Generally, item discrimination indices from + 0.25 to + 1.00 may be considered adequate for use in norm reference test items.

Discrimination index or sometimes called choice of criterion, according to Anastasi (1999) refers to the degree to which an item differentiate correctly among test takers in the behavior that the test is designed to measure. In another approach, Denga (2003) stated that the purpose of discrimination test is to distinguish as much as possible among students (examinees) at all levels of achievement. It indicates the effectiveness or power of an item in discriminating between bright and dull students. Similarly, Kelly and Linacre (2002) described discrimination as an indication of the extent of which success on an item corresponds to the success on the whole test. They point out that since all the items in a test are geared towards jointly generating an overall score, any item with a negative or zero discrimination undermine the test, in other words, any item that falls short of appropriate discrimination requirement is not suitable for inclusion in a test battery. Positive item discrimination is considered productive unless it is so high that the item is merely repeating the information provided by other items.

They hold the view that if the discrimination index (D) is computed from equal size high and equal size low being groups in the test ending up with an index range of +1 and -1 this could ultimately result in a feedback loop.

Kline (2000) in his work on item analysis within the classical test theory (CTT), approaches, statistical analysis and interpretation, found that the higher the discrimination index, the more the item discriminates. To determine the discrimination, he grouped those who have the highest and lowest overall test scores. The upper group was made up of 25-33 per cent who are the best performers, i.e. those with the highest overall test scores; and the lower group weremade up of the bottom 25 per cent to 30 per cent who are the poorest performers (have the lowest overall

test scores). In doing this, he used 27 per cent of the distribution, as the crucial ratio that separates the mean of the standard normal distribution of responses error.

On the other side, Oosterhot (1976) has indicated clearly that differences may exist in procedure and assumptions, most item discrimination index provide closely similar results. The numerical values of the indices might differ in the items that have been retained and those that are rejected on the basic of the different discrimination indices are largely the same. The author further analyzed that variations in item discrimination data from sample to sample is generally greater than that among different methods. In line with this view, Denga (2003) stated that discrimination indices are influenced by some students and examination factors which include:

i The learning experience of the examinee

ii The appropriateness of the stem to structure the question for the examinee

iii The extent of ambiguity in the item

iv The attractiveness of foils (attractiveness) to fools those who do not know the correct answer.

v. The difficulty of the item

vi. The presentation of the best foil which will appeal to the upper group

Despite the influences of these students and examination factors on discrimination, Ebel and Frisbre (1991) noted, a test with a high average discrimination index is always better indexed. They added that, despite the index, the former test will always produce more reliable scores than the later one.

In another empirical finding by WAEC (1995), the option format in terms of better discrimination between high and low achievers, revealed that between subject officers, item writers and teachers, the three categories of respondents had a greater percentage of teachers representing 60.50 percent and item writers 66.7 percent asserted that the 5-option format discriminates relatively better than 3- option or 4 - option formats. The result appears to suggest that the 5-option format is preferred followed by the 4-option format and 3-option in terms of better discrimination between high and low achievers.

Considering the reduced sample size, fewer examination bodies, different study area, different statistical techniques and different research design used in this work when compared with the findings of Bandele and Adewale (2013), there is a great significance difference in the item difficulty of the two examination bodies.

Hypotheses

1. WAEC Mathematics multiple choice items does not significantly differ from those of NECO for 2011 SSCE examinations in terms of their difficulty indices.

2. There is no significant difference between WAEC Mathematics multiple choice items and those of NECO for 2011 SSCE examinations in terms of their discriminating power.

METHODS

The Ex-post facto (causal comparative) design was adopted for this study and the population comprised all Senior Secondary three (SS III) students in the 86 public secondary schools in Uyo Education zone of Akwa Ibom State totalling 12,499. The stratified random sampling technique was employed and a sample size of 250 SS III students drawn from ten (10) selected secondary schools on a sampling fraction of 0.02 indicating that each person in the study sample represented by 2% of the students in the sample frame. Stratification was done based on Local Government Area within the zone.

WAEC and NECO multiple choice test items (objective) in Mathematics were fully adopted as the instruments for data collection for this study. Data collected were analysed using Independent T-test statistics. From the item analysis done, the difficulty index (p-value) range of $0.40 \leq p \geq 0.60$ were preferred while p-values of less than 0.40 and greater than 0.60 needed review. Items with discrimination index (d-value) of 0.30 and above were preferred while items with very low, zero or negative discrimination indices needed careful examination and review.

RESULTS

The following are the results of the data analysis:

Hypothesis One: there is no significant difference between WAEC Mathematics multiple choice items and those of NECO for 2011 SSCE examination in terms of difficulty indices. In testing this hypothesis, the scores generated from the test were transformed, and Item analysis was done as well as Independent t-test and the summary of the result presented in Table 1.

Table 1: Summary of item analysis of WAEC and NECO Mathematics multiple choice items in terms of their difficulty level: Number of items with appropriate difficulty level

Examination	**Item with appropriate difficulty index**	**%**	**Item with inappropriate difficulty index**	**%**	**Total**
WAEC	37	74%	13	26%	50
NECO	34	56.7%	26	43.3%	60

$0.40 \leq p \leq 0.60$ = appropriate; $p < 0.40$ and $p > 0.60$ = inappropriate

The acceptance range of difficulty was fixed at 0.040 – 0.60. All items having difficulty index lower or higher than this acceptance range were considered to be inappropriate. As shown in Table 1, WAEC has lower percentage of items with inappropriate difficult index as 26% of the items in 2011 in multiple choice items in Mathematics. The higher number of items with inappropriate difficulty index was recorded by NECO which had the percentage of 43.3 in the same year in Mathematics.

Table 2: Independent t-test analysis of the difference between WAEC and NECO Mathematics multiple choice items in terms of their responses.

EXAMINATION	N	Mean	SD	t-value	p-value
NECO	250	0.59	0.15	5.762*	0.00
WAEC	250	0.52	0.12		

*Significant at 0.05; $p<0.05$; df = 498.

The information in Table 2 shows that there is a significance difference between WAEC and NECO Mathematics multiple choice items in terms of their difficulty level since the calculated t-value of 5.762 tested at 0.05 levels of significance was found to be greater than the critical t-value of 1.96. Therefore, this result implies that the null hypothesis which states that the difficult level of WAEC Mathematics multiple choice item does not significantly differ from those of NECO for 2011 SSCE examination was rejected while the alternate was not rejected.

Hypothesis Two: there is no significant difference between WAEC Mathematics multiple choice items and those of NECO for 2011 SSCE examinations in terms of their discrimination indices. In testing this hypothesis, the scores generated from the test were transformed, and Independent t-test was done as well as Item analysis and the summary of the result presented in Table 3.

Table 3: Summary of items analysis of the Discrimination Power of test items by WAEC and NECO in Mathematic Examination Instrument: Number of items with appropriate Discrimination Index

Examination	Item with Good Discrimination Power	%	Item with Poor Discrimination Power	%	Total
WAEC	50	100%	0	0	50
NECO	31	51.7%	29	48.3%	60

$D \geq 0.30$= good< 0.30= poor

The acceptance range of discrimination power was placed at 0.30 and above. In this study, it was observed in Table 6 that all WAEC items had better discrimination power than NECO. This implies that WAEC items discriminate well between the upper scorers and lower scorers in Mathematics than NECO.

Table 4: Independent t-test of the discrimination index of WAEC and NECO Examination instruments

EXAMINATION	N	Mean	SD	t-value	p-value
WAEC	250	0.71	0.32	7.01*	0.00

NECO	250	0.53	0.25

Significant at 0.05; $p<0.05$; df = 498.

The information in Table 4 shows that the calculated t-value of 7.01 is greater than the critical t-value of 1.96 at 0.05 level of significance with 478 degrees of freedom. This means that the null hypothesis is rejected while the alternate hypothesis which stated that there is significant difference between the discrimination index of WAEC Mathematics multiple choice items and those of NECO was retained. This implies that the WAEC instrument discriminates more than the NECO instrument.

DISCUSSION

The result of this study reveals that there is a significant difference between the difficult level index of WAEC Mathematics multiple-choice items and those of NECO in SSCE examinations. This implies that NECO instrument possesses more difficulty indices than WAEC examination instrument. This finding disagrees with the study of Bandele and Adewale (2013) that compared the item difficulty level of WAEC, NECO, and NABTEB Mathematics achievement examination in Nigeria. A descriptive research of the survey type that involved the use of a correlation design was used in this study with a sample of 600 final year students randomly drawn from selected Government Technical Colleges and Senior Secondary School from Ondo, Ekiti, and Osun states of the South – West geopolitical zone. The instruments used consisted of adopted WASSCE, NECO, and NABTEB Mathematics examinations. Analysis of Variance (ANOVA) was employed to analyze the data. It was revealed that there was no significant difference in the item difficulty levels of WAEC, NECO, and NABTEB Mathematics achievement examination. On the other hand, the result of this finding agrees with the findings of Sotaridona, Pornel, and Valleyo (2003) who on their various ways have established that the influence of difficulty item in response variability affect item difficulty index and changes in items response pattern. However, the variance of the result of Bandele and Adewale (2013) may be connected to the type of research design that was used in the study and the number of instruments that were compared.

The findings of this study reveal that there is a significant difference between the discrimination indices of WAEC and NECO examination instruments in Mathematic. This means that there is a significant difference in the two examination instruments in pulling apart the bright and dull students setting in an examination. The result shows that WAEC instruments pull apart the dull students from the bright ones more than the NECO instrument. The findings agrees with earlier findings like Abel and Frisbre (1991) who conten d that a test with a high average discrimination index is always better index and that such a test would produce a more reliable zone than the other.

The findings are also in agreement with Hotni (2006) as it is seen that discrimination index is a useful measure of item quality whenever the purpose of the test is to produce a spread of scores reflecting differences in students' achievement. This enables distinctions to be made among the performance of respondents, particularly as discrimination measures the extent to which item responses discriminate between individual examinees who have higher overall scores on test and those that get a lower overall score as in the case of the two examination instruments.

Denga (2003) reported that discrimination indices are influenced by some factors such as the previous learning experiences of the examinees, the appropriation of the item to structure the question for the examinees, the difficulty of the item and the presentation of the test foil which will appeal to the upper group. This view supports the position of this study as the two examination bodies' scores were found to be different in their own levels. This is observed in their number of options and items in each instrument. Moreover, it is important to note that this result may be attributed to the examination administration which is not considered in the study.

CONCLUSION

On the basis of these findings, it was concluded that the NECO examinations was more difficult than the WAEC examination items. It was also concluded that the WAEC examination instrument discriminates more than the NECO examination items.

RECOMMENDATIONS

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

1. NECO should review their item options from 5 to 4 by deleting options that are not plausible as it will help to distract more of dull students than bright ones.

2. A body of experts should be commissioned to scrutinize items for national examinations such as NECO and WAEC in order to standardize such examination.

REFERENCES

Adeniran, T. (2000) NECO'S Certificate: Valid for varsity admission. The punch, Lagos, pp: 29.

Akpan, S. M. (1989). Factors that influence students item response pattern in a mathematics test. Unpublished M. Ed. Thesis, faculty of Education, University of Calabar, Nigeria.

Akuezuilo, E. O & Agu, N. (2003).*Research and statistics in education and social sciences, methods and applications*. Awka: Nuelcenti Publishers & Academic Press Ltd.

Alonge, M. F. (2003). Assessment and examination: The pathways to educational development. *9th inaugural lecture presented at university of Adi-Ekiti*.

Anagbogu, G. E. (2009). Analysis of psychometric properties of WAEC and NECO examinations and students' ability parameters in Cross River State, Nigeria Unpublished Doctoral Dissertation. University of Calabar, Calabar, Nigeria.

Anastasi, A & Urbina, S. (2010) *Psychological testing*. New Delhi: PHI Learning Private Limited.

Anastasi, A (1999) *Psychological testing*. New York: Macmillan

Anastasi, A. (1992). Some emerging trends in psychological measurement: A fifty year perspective. *Applied Psychological Measurement*, 9, 121 – 138

Bandele, S. O. & Adewale, A. E. (2013) Comparative analysis of the item difficulty levels of WAEC, NECO and NABTEB Mathematics Achievement examinations. *Mediterranean Journal of Social Science*, 4(2) 761-765.

Brown, F. G. (1983). *Principles of Educational and Psychological testing*. New York: Holt, Reinhart and Winston

Denga, D. J. (2003).*Educational Measurement, continuous assessment and Psychological testing.* Calabar: Rapid Educational Publishers.

Ebel, R. L. & Frisbie, D. A. (1991).*Essentials of Educational Measurement*. Prentice Hall Englewood Cliffs.

Ekpenyong, D. B. (1991).Influence of Psychological variables on classical test item parameters. Unpublished M. Ed. Thesis, faculty of Education, University of Calabar, Nigeria

Enyi, G.S. (2002). *Introduction to measurement and evaluation for universities, polytechnics and colleges of education: A professional approach*. Enugu: Hamson Pal.

Federal Republic of Nigeria (2004).*National Policy on Education* (4thed). Lagos. Federal Government Press.

Hotni, C. E. (2006). Personal elements in teachers' works. *Journal of Educational Research,* 12, 49-55.

Joshua, M. T. (2005) *Fundamentals of text and measurement in education*. Calabar: University of Calabar Press.

Kelley, I. & Linacre, J. M. (2002) Intern Discrimination Indices. *Research Measurement Transition,* 16(3): 883-884.

Kline, P. (2000). *A psychometric primer*. London, UK: Free Association Books.

Kolawole, E. B. (2001) A comparative analysis of psychometric properties of two Nigerian Examining Bodies for Senior Secondary Schools mathematics. *Research Journal of Applied Sciences,* 2 (8) 913-915.

Kolawole, E. B. (2001). *Test and measurement*. Ado-Ekiti: Yemi Prints and Publishing services.

Kolawole, E. B. (2002) Assessment of West African Examination Council (WAEC) and National Examination Council (NECO) results in both Mathematics and English in Eketi State Secondary School Examination. *Junior Mathematics Science Education,* 2, 152-165.

Nwaobia, E. M. (1990). *Monograph on introduction to advanced measurement and evaluation.* University of Nigeria, Nsukka.

Okoye, R. O. (1996). *Educational and psychological measurement and evaluation*. Ikeja, Lagos: ED-Solid Foundation Publishers.

Onah, F. E. (1994). Content validity of teacher made Agricultural Science tests usedin secondary schools in Enugu State (Unpublished Master's Thesis). University of Nigeria, Nsukka.

Oosterhot, A. C. (1976). Similarity of various item discrimination indices. *Journal of Educational Measurement,* 13,(2) 145 – 150

Salman, M.F, Mohammed, A.S, Ogunlade, A.A, & Ayinka J. O. (2012) Causes of Mass Failure in SSC Mathematics Exams as viewed by secondary school teachers and students in Ondo, Nigeria. *Journal of Education and Practice*, 3 (8) 79-88.

Sataridona, L. S., Pornel, J. B. &Valleiyo, A. (2003).Some application of item response theory to testing. *The Philippians Statistician*, 52 (14) 82 – 92

Sidhu, S.K. (2005). *New approach to measurement and evaluation*. New Delhi: Sterling Publishers Ltd.

Temitope, O. (1999, July 8). Welcome NECO. *The Punch*. Lagos.

WAEC (1995). An Empirical Check on the effect of the number of options of multiple choice items on the difficulty and Reliability of ordinary level Questions papers. Completed research project Report,11, 2, 63 – 72.