PSYCHOMETRIC EFFECT OF THE NUMBER OF MULTIPLE CHOICE RESPONSE OPTIONS ON SECONDARY SCHOOL STUDENTS' PERFORMANCE IN ECONOMICS TEST (A CASE OF ITU L.G.A, AKWA-IBOM STATE, NIGERIA)

UDOFIA, I. GODWIN
Dept. of Educational Psychology, Guidance and Counselling, Adeniran Ogunsanya COE, Lagos, Nigeria
&
Mrs. DELHINE ORATOKHAI
Dept. of Educational Psychology, College of Education, Ekiadolor, Benin, Edo State.

Abstract
The study investigated the Psychometric effect of multiple choice response item options on secondary school students' Performance in Economics Test. The population for this study consists of all senior secondary school students in Itu L.G.A of Akwa-Ibom State, Nigeria. 3x3x2 Schematic type of Quasi-experimental design was used for this study. A sample of three hundred (300) students from Six Senior Secondary Schools was selected through a simple random sampling technique. The instrument is an adopted Economics Achievement Test (EAT) with 5 option format type and an adapted format type with 4 and 3 options each with fifty items. Each instrument has its reliability co-efficient value (r) as 0.81, 0.79 and 0.85 respectively. The difficulty and discrimination indices for the options were determined while the use of student t-test analysis was used to test the research hypothesis at 0.05 level of significance. The formulated hypotheses were all rejected. It was found that the use of three option type saves time on the part of the test developer and marking process and ensures adequate content of the subject. It was concluded that during evaluation exercises, test experts should adopt the use of three-option formats as against the conventional four and five-option type. The study recommended that test experts who are charged with the responsibility of developing standardized tests should put into practice, a more evaluative measure to address the issue of guessing especially as it concerns Economics objective test.

Key Words: Number of Options, Performance, Multiple Choice items, Difficulty index, Discrimination index, Psychometric properties.

Introduction

All over the world examination has always been a platform used to select subjects from a pool of entries. Test personnel see examination as a dynamic exercise used to perform functional tasks in a bid to predict students' future abilities, provide guidance and services function, enhance research and development process, experimentation, evaluation exercise among others. The use of test still remains the most reliable means of sourcing for composite and effective quality manpower in any situation. Even in the circular world, it remains a sure means of identifying weak, average and brilliant learners from a group of heterogeneous entities. Okpala, Onocha, and Oyedeji (1993) maintained that examination exists to determine the extent of material and content grasped by a testee in the course of undergoing a test exercise. The examination process allows for qualitative and quantification of test scores in order to take an informed decision.

The multiple-choice type of test has proved effective and efficient in measuring knowledge and skills (Haladyna, 2004; Downing & Haladyna, 2006) and it should continue to be used as a valid measurement of knowledge and skills. Multiple choice tests comprised the problems (questions) which are called the stem and a list of alternative responses (That is the correct answer called the key while the incorrect options are called distracters). In any educational system, the importance of assessing students with multiple choice items cannot be ignored. Therefore, as Nigeria is striving to remain competitive in the world of technology, there have been agitations for test experts to adopt the use of new techniques in constructing multiple test items. Thus, these items should meet the expected psychometric properties. In most Nigeria's secondary schools, it is expected that subject teachers should write or construct test items for their mock examinations in their respective schools because it would help predict the ability of their students ahead of their long awaited summative test i.e senior secondary school examination. Linacre, (2008) observed that test items constructed by subject teachers for the purpose of evaluating the level of students' achievement are not reliable and valid. For any achievement test to measure what it is supposed or purported to measure, it should be valid.

Students all over the world are familiar with test questions either having 4 or 5 option format. A series of studies on testing have been carried out and a majority of them have advanced for a change in the construction pattern of 4 or 5 options to 3 options. Reasons ranging from reasonable reliability quotient, economical in nature, test wise nature of examinees, reduction of guessing because guessing

attracts punishment, saves time for both examinees and examiners, rapt content sampling among others, are some of the findings responsible for a shift, advancement and deviation from the contemporary 4 or 5 option format to 3 option format in the course of test construction.

Multiple Choice Tests: Uses and Relevance
Multiple-choice tests are of considerably widespread use as a means of objective measurement. The main reason behind such popularity is the many dominant advantages associated with multiple-choice tests. They can be used for diagnostic as well as formative purposes and can assess a broad range of knowledge. In addition, they are scored easily, quickly, and objectively either by human-beings or by scoring machines. These and many similar advantages make multiple choice tests suitable for a wide range of purposes ranging from classroom achievement testing to large scale standardized tests. Thus, improving the quality of multiple choice test items appears to be of a lot of importance.

One of the most frequently mentioned guidelines with regard to option development deals with the number of options to be written for each item. Haladyna and Rodriguez (2002) claimed that test experts should write as many plausible distracters as possible because the more distracter made available, the more tendency for weak and less test wise students to opt for the wrong distracters. Previous similar studies carried out by Linn, & Gronlund, (2000) on "Measurement and assessment in teaching" and Udofia (2008) on "Effect of number of options on students academic Performance" moved for the use of 3-options format because it allows for adequate content sampling and a more practicable and realistic discrimination index obtained through examinees scores while also reducing the effect of guessing. Technically standardized tests organizations like The West Africa Examinations Council (WAEC), National Examination Council (NECO), Joint Admissions and Matriculation Board (JAMB) among others usually adhere to the rule of 4 or 5-options type. In spite of the widespread use of four or five options per item advocated by many authors and test developers, most of the studies carried out to investigate the optimal number of options have ended with most recommending the use of three-choice items Straton & Catts, (1980); Aamodt & McShane, (1992); Crehan, Haladyna & Brewer, (1993); Haladyna & Downing, (1993); Landrum, Cashin & Theis, (1993); Delgado & Prieto, (1998); Rodriguez, 2005; Shizuka, Takeuchi, Yashima, & Yoshizawa, (2006).

The use of 3-option format as alluded to by test experts is that it is very economical in the realm of measurement and evaluation, scoring is very reliable unlike 4 or 5 option type because the responses are scored according to

predetermined key, allows for identification of difficulty and discrimination index with ease, flexible in assigning scores, reduces guess among testees, affords adequate and comprehensive content sampling among others. In some developed economies of the world today, their examination bodies responsible for conducting standardized tests have advocated for the use of 3 option format because it's a unique way of testing examinees and the face validity of such items is also Uni-dimensional as well. According to Udofia (2008), the 3-option test format tends to have a higher reliability value when compared to test formats that take either 4 or 5-options format.

Psychometric Nature of Multiple Choice Items

Technically during test construction, Multiple Choice Questions, MCQs have two parts, a stem, the question, problem, or task to be answered or solved; and a set of response options or alternatives possible answers or solutions to the question. The options comprises of the correct answer, called the key; and one or more incorrect or less appropriate answers called the distracters or distractors (Onunkwo, 2002). While the stem is an important part of a multiple choice question (MCQ), the options are no less relevant. No matter how well written a stem, a single flawed option can invalidate the item. As a rule, well written options are critical for an MCQ to be adjudged as valid.

For test results to be accurately interpreted and applied, the psychometric features of validity and reliability are necessary and sufficient requirements. Validity is the extent or degree to which the test can measure the qualities, skills, abilities, traits, or information that it was designed to measure; while reliability is the extent to which it can consistently make these measurements. MCTs are a commonly preferred format for large scale testing because they have been found to give valid and reliable results. They can be effectively used to assess a large number of test takers, to cover a wide range of content area (Downing, 2002, cited in Tarrant, Ware & Mohammed, 2009), as well as a wide range of learning objectives (Okoli, 2005), leading to valid results. They can be objectively, easily and accurately scored (Linn & Gronlund, 2000), and they contain a relatively large number of test items, leading to a high degree of reliability.

The psychometric nature of test items accounts for the fact that the test constructed has satisfied the conditions of validity and reliability. It is expected that before test items are presented to examinees, these test items must meet either face, content or construct validity. And after which the reliability of the items are determined to state the extent of consistency of the test scores. Some test experts like Peter, and Louisa, (2013), see Psychometric nature of a test as incomplete if it ceases to determine the discrimination and difficulty level of a test.

Discrimination index is used to indicate the extent to which response to an item could distinguish between the strong and weak students. It measures the extent to which items are capable of measuring individual differences. The discrimination index indicates the effectiveness or power of an item in discriminating between good and poor students. It varies between 0 to 1.00 while difficulty index is an indication of the extent to which an item is difficult for respondents. It is usually defined as the proportion of students responding correctly to an item. The higher the proportion is, the easier the item is. The maximum value of P, 1.00 occurs when everyone responds correctly. The lowest value of P, 0.00 means that everyone missed it or failed to respond to it. An item difficulty level is determined by estimating the percentage of examinees that are likely to get the item right when it is administered Ogunjimi (2009).

Arguments For and Against the Adoption of 3, 4 and 5 Option Test Type
There are significant arguments in favour of Four and Five option type against 3-options type. Woodford and Bancroft (2004), Abad, Olea and Ponsoda (2001), and Farhady and Shakery (2000) have asserted the following:
- That lower number of options, such as three, increases, to an unacceptably high degree, the chances of successful random guessing and the extent of guessing effects, such as over-estimation of student achievement or ability (with five options, the degree of chance success is 20%; with four options, it is 25%; and with three options, it is 33.3%),
- That this decreases the psychometric quality the test scores, making it less reliable and consequently less valid and
- This psychometric limitation can only be corrected by using five, or at least four options.

The preference for three options, over four and five option format was based on a study by Costin (1970) who claimed that three options type sometimes gives similar and superior outcome on many testing criteria, and are thus preferable for some testing purposes. He claimed they are less arduous and time consuming to develop and administer, and have reduced completion time. The time and resource thus saved could be used to develop more items for the test, boosting reliability. Consequently, this time, cost, and energy saved, and used for other relevant activities, increase efficiency of assessment without compromising test quality and hence, can be equated with increase in content validity and test reliability.

Studies have substantiated the efficacy of three options and their suitability for educational tests. Theoretically, using mathematical formulae-driven perspectives, it has been contended that three options optimize the discrimination ability of a

test and the information that it could provide (Ebel, & Frisbie,(1991) especially with moderate item difficulty (Lord, 1977); that reliability significantly increases when the number of option is increased from two to three, but not by much when increased to four, and that any increase beyond three would only give marginal increase in the range of 0.02 to 0.05 and the three option type gives higher reliability and discrimination than the others (Grier, 1975). From the empirical perspective, studies have found three option items to have higher mean scores and to be less difficult and more discriminating (Landrum et al.,1993); (Tversky, 1994); Trevisan et al., 1994).

It is noteworthy that despite the claims by traditionalists, there is a dearth of contemporary research evidence to support the superiority and continued use of five or four options. Yet, most achievement and ability testing programmes and examinations still use five or four options, while test and measurement textbooks typically recommend this, based on the belief that the greater the number of options, the higher the reliability. Test experts are of the belief that three options are sufficient in most instances and that the effort of developing the fourth or fifth option type is probably not worth it. In a landmark review research study by Haladyna et al. (2002) they reviewed twenty seven measurement textbooks and twenty seven research studies and reviews. They found significant research support for three options, but revealed that most of the assessment textbooks surveyed were still divided on the issue, with about 70% advocating the prevailing guideline, and a few others advocating a middle of the road 4 option, as an industry standard.

However, backed by contemporary research evidence, counter arguments have been put forward. Haladyna and Downing (2004) have stated that in a majority of cases, it is often difficult and time consuming to write up to four or even three plausible functional distractors for an MCI, so that not more than two of the distracters are usually functional and additional options after the third one is often always implausible. They add that the additional distracters are merely fill-ins which are not plausible enough to distract weak students, who immediately see through them and easily guess the correct answer. They advocate three options (i.e. two distracters and the correct option) as a natural limit under most circumstances.

Proponents of five or four options have countered these claims with the assertion that if the extra options are functional, (if they are well-written and plausible), the overall benefit of reducing guessing outweighs any extra time that may be gained by using 3 options, and constructing more items to boost reliability (Woodford & Bancroft, 2004); and that since the use of more options reduces guessing effects,

which increases reliability, it also increases validity (Thorndike & Thorndike-Christ, 2010). These scholars have described these effects as overrated and negligible respectively, arguing that most serious student's who have adequate time to write a test, use partial knowledge and educated guessing, rather than the psychometrically deleterious random guessing.

In terms of reliability, Stratton and Catts (1980) found similar reliability (and higher standard error of measurement) for three options; while Landrum et al., (1993), Trevisan et al., (1994), Sidick et al. (1994), Rogers and Harley (1999) all found non significant or little increase in reliability. While similar studies by (Shizuka, Takeuchi, Yashima and Yoshizawa, 2006) neither found any significant difference in difficulty, nor in both difficulty and discrimination but (Rogausch, Hofer and Krebs, 2010) all found increase in item difficulty.

Economics in Senior Secondary Schools
Economics is one of the subjects taught at the senior secondary school level in Nigeria. Its teaching helps students to understand and develop basic economic skills needed for engendering national development as well as involving economic techniques for solving societal problems. Economics is important in giving learners a more comprehensive understanding of contemporary issues facing an economy and the world economies at large. As Nigeria is undergoing rapid economic advancement, establishing closer links with the world technology and striving to remain competitive among the world economies, the study of Economics is highly relevant. Economics as a subject also helps students to understand the principles and practice needed in undertaking their everyday livelihood activities and for participating in those activities that engender the well being of the society.

The importance of teaching and learning Economics to any nation can only be ignored at its peril, as it enables both leaders and citizens to understand basic concepts, principles as well as appreciate and seek to improve emerging economic dynamics for their own social good. According to Shizuka, Takeuchi, Yashima, & Yoshizawa, (2006) management and business management derive their existence from Economics. This implies that Economics is the root on which business/management and all management related disciplines such as insurance, accounting, purchasing and supply, marketing, personnel management among others, are hinged. Thus, an attempt has been made by stakeholders to integrate the theoretical baseline of the subject curriculum with practical applications. The Nigerian senior secondary school Economics curriculum proposes that the subject should prepare students for an entrepreneurial career in future. The 21st century has witnessed landmark economic reforms and stratification globally. In Nigeria,

the last decade has also witnessed unprecedented economic reforms that have influenced different sectors of the economy. There is, therefore, the need according to Nigerian Educational and Research Development Council (NERDC), to make the Economics curriculum responsive and relevant to the Nigerians quest to be among the top 20 players of the world economies by 2020 (NERDC, 2008). In order to actualize this, there is a need to assess learning achievement with a standard instrument.

Economics achievement test is designed to measure the extent to which learners exposed to the subject have acquired or mastered the content as well as achieved the objectives of the subject taught. The individual's performance on the course content is an index of his mastery of the executed piece of instruction (Rogauch, A., Hofer, R., & Krebs, R. (2010). Achievement tests are designed to find out whether the required behavioural objectives needed in the learner have been achieved as a result of learning. Achievement test serves the purpose of measuring what the individual has learnt, educational progress and specific school experiences. Studying students' achievements can serve as predictive value to educators, researchers and evaluators. However, constructing Economics objective items using 3, 4 or 5 option format can further juxtapose between discriminatory and difficulty index of learners especially when used for a heterogeneous group.

Statement of Problem
Construction and validation of multiple choice tests have been in existence for decades in both developed and developing countries. In Nigeria, construction and validation of achievement test has been based mainly on classical test theory framework until recently, when public examining bodies began to shift from the mere theories surrounding its formation to the number of options to maintain in order to minimize guessing, differentiate between quality of learners, justify the impact of distracters in number of options among others.

Structurally, examination bodies have over flogged the use of five and four number of options in test construction with respect to its difficulty and discriminatory index. Technically, emphases are moving towards the use of three option format (That is more efficient in terms of time to develop, administer, and complete) as a catalyst to effective evaluation of learning. How valid is this assumption in the light of the present fall in student's performance in Economics tests?

Research Hypotheses
Ho1: There is no significant difference between discriminating and difficulty level of Pre-test scores of the experimental and Control groups in Economics test
Ho2: There is no significant difference between discriminating and difficulty level of Post-test scores of the experimental and Control groups in Economics test

Objective of the study
1. To investigate the extent of differential between discriminatory and difficulty levels of students' performance in Economics test designed using three, four and five number of options
2. To identify the extent to which differences exist when three, four and five options formats are used in test construction especially in Economics examinations.

Significance of the study
1. The study will be of importance to test experts in that it will give them better information as to which number of options to adopt during test construction.
2. To teachers, the outcome of this work will guide them in taking an informed decision as to the appropriate number of option to use in order to reduce or minimize guessing among students especially in objective tests.

Methodology
This study adopted a 3x3x2 Schematic Pre-Post test type of Quasi-experimental design because no variable was manipulated or matched at any point in time throughout the study while treatment was at one level (Experimental group). The essence of this design was to find out how the Number of Options in a Multiple Choice Item affects students' Performance in Economics Test. The study covered Public Senior Secondary Schools, (SSS) in South-South, Nigeria. The reason for the choice of this study area is on the premise that most states in the region record low pass mark in their students' annual reports (WAEC) but recently records significantly positive performance especially in Economics examinations, according to a Chief Examiner's Report, (2015).

A total of three hundred (300) students participated in this study as sample. These students were from intact classes. Three senior secondary schools were selected for this study through the adoption of a simple random sampling technique. With the aid of table of specification, a 50-item Economics Achievement Test (EAT) with 5-option format, 4-option format and 3-option formats constructed by the researcher was used as instrument for the study. After meeting face and content validity by experts in the field of Test, Measurement and Evaluation, twenty students from one of the schools not part of the study but have all the

characteristics with those of the main study was used as pilot study. After using test-retest method of reliability, the derived co-efficient for 5-option, 4-option and 3-option formats were 0.81, 0.79 and 0.85 respectively. At the first phase the researcher pre-tested the experimental groups along side the control groups, thereafter the experimental groups were exposed to treatment, but the control groups were not exposed to any form of treatment. After a time lag of two weeks, the same Economics Achievement Test, EAT, with 5, 4 and 3-options format were re-administered to both groups again (Experimental and Control groups) in order to get their Post- test scores.

Method of Data Analysis
Multiple regression analysis was used to analyse the formulated hypothesis and Scheffe Post-Hoc analysis was used to determine the effect of treatment on which option format used to test the students in Economic Achievement Test. The hypothesis was tested at 0.05 level of significance.

Testing of Research Hypothesis
The following research hypotheses were formulated for the research study:
Ho1: There is no significant difference between discriminating and difficulty level of Pre-test scores of the experimental and Control groups in Economics test

Table 1: Simple regression table showing Difficulty and Discriminatory index of Pre-test scores for Experimental and Control group

Forms	No of Options	N	Diff. level	Discr. Level	Mean	SD	DF	t-cal.	Sig.	Decision
Experimental group	5-options	150	0.49	0.45	0.64	0.46				
	4-options		0.15	0.32	0.51	0.59	298	0.678	0.013	Significant
	3-options		0.68	0.41	0.69	0.83				
Control group	5-options	150	0.63	0.33	0.59	0.51				
	4-options		0.36	0.37	0.66	0.39				
	3-options		0.31	0.47	0.47	0.66				

Source: Data output, 2017

Table 1 shows an averagely low difficulty index meaning for the experimental and control group except for the 5-option format for control group which shows a relative increase while the relative index depicts that the items do distinguish between weak and strong students. However, t-cal. value is 0.678, significant at 0.013 where $P<0.05$ meaning that the null hypothesis stated is rejected and the alternative hypothesis which states that there is a significant difference between discriminating and difficulty level of Pre-test scores of the experimental and Control groups in Economics test is upheld.

Ho2: There is no significant difference between discriminating and difficulty level of Post-test scores of the experimental and Control groups in Economics test

Table 2: Simple regression table showing Difficulty and Discriminatory index of Post-test score for Experimental and Control group

Forms	No of Options	N	Diff. level	Discr. Level	Mean	SD	DF	t-cal.	Sig.	Decision
Experimental group (PostTest score)	5-options	150	0.64	0.62	0.45	0.64	298	0.851	0.046	Significant
	4-options		0.55	0.56	0.68	0.51				
	3-options		0.71	0.78	0.52	0.83				
Control group (PostTest score)	5-options	150	0.72	0.58	0.65	0.76				
	4-options		0.46	0.49	0.54	0.33				
	3-options		0.44	0.67	0.43	0.61				

Source: Data output, 2017

Table 2 shows a relatively high difficulty index for both experimental and control group except for the 4-option format for control group which shows 0.46. While a significant increase was also observed for the discrimination index which means that for both experimental and control groups, the items discriminate between the weak and strong learners. However, t-cal was 0.851, significant at 0.046 where P<0.05 meaning that the null hypothesis is rejected and alternative hypothesis which states that there is a significant difference between discriminating and difficulty level of Post-test scores of the experimental and Control groups in Economics test is upheld.

Table 3: Scheffe Post-Hoc tests analysis of experimental and control test scores only according to 5, 4, and 3-options

Groups	N	Mean	5options	4options	3options
Experimental Group	150	76.41	***	***	****
Control Group	150	49.59	***	**	***
TOTAL	200				

Source: Data output, 2017

Table 3 showed that students in experimental group who preferred 3-Option format contributed significantly to the mean while in the control group, students who preferred 5-options and 3-options contributed significantly to the mean score. However, it further reveals that students in the experimental group have the highest mean of 76.41 while students in the control groups have 49.59 respectively, stating that the experimental group contributed significantly to the change in scores of the students due to the treatment administered.

Discussion of findings

The result is not in consonance with the findings of Peter and Louisa (2013), who argued that the number of options in an item does not affect the test scores greatly. The work revealed that three-choice items were indicated to be only slightly easier than four and five-choice items respectively, takes significantly less time to construct by test experts and easy to attempt by students. So it was inferred that instead of four or five choice items, more of three-choice items should be adopted in an examination because it does not contribute in any way to an increase in testing time and thus more content valid tests are being guaranteed of being developed.

The study noted that Rodriguez (2005) reported a decrease in item difficulty is observed when there is a reduction in the number of options while Landrum, Cashin, and Thesis (1993) asserted that either an increase or no change in item difficulty can arise as a result of reducing the number of options per item. Yet, Shizuka, Takeuchi, Yashima, and Yoshizawa (2006), as well as the present study detected a significant difference regarding item difficulty between tests administered to experimental and control groups. However, considering a number of psychometric properties together as well as practicality related issues, all the assertions put forth by psychometricians converge on the use of three options.

Another revelation made was that three-choice test items were more difficult than or at least as difficult as their four-choice counterparts. As any improvement in students' performance was associated with the superiority of three-choice test items, or by the format in which the test expert state the three-choice items specifically to measure student knowledge" (Landrum, Cashin, & Thesis, 1993).

Criticism is usually made of using fewer options per item due to enhancing the probability of guessing. However, as the results of the current study revealed, multiple choice tests, regardless of their number of options per item, would remain almost immune to the effect of guessing factor when the items are appropriately targeted for the group of test-takers, and enough time has been

allotted. Furthermore, using three options would allow for the inclusion of more items in a fixed period of time, thus increasing the reliability and validity of the test (Udofia, 2008).

Moreover, the Wright map showed that items were of appropriate level of difficulty for the test-takers and were very well-targeted. This could be another reason why the 4 and even the 3-option items were as good as the 5-option items. Guessing usually occurs when the items are above the ability level of the test-takers or when the test is administered under time constraints (speed tests). In this study, the time allotted to complete the test was enough. Therefore, one can argue that if 3-option multiple choice test items are carefully constructed, are at the appropriate level of difficulty and are administered as power tests and not speed tests, they can be as difficult as 4 and 5-option tests and as efficient in terms of discrimination, reliability and fit to the Rasch model.

Adegoke (2012) posited that treatment in relation to any study plays a significant role in testing learner's ability. Studies have shown that the importance of treatment is so relevant that it contributes significantly to the outcome variable. Due to treatment package, it tends to reduce the probability of guessing any item during testing. However, the three-option format seems to have a high propensity to reduce guessing because the powers of distracters are easily identified (Rogauch, A., Hofer, R., & Krebs, R. (2010).

Conclusion
1. Three option format is very easy to construct and allows for the coverage of a wide content area with less distracter unlike four or five option format.
2. Furthermore, the study also revealed that irrespective of the nature of subject to be set, the use of three-option format enhances a swift reduction of guessing prevalent among students globally.

Recommendations
From the study the following recommendations were made:
- Psychometric nature of multiple choice items test should be tailored towards the use of item response theory (IRT) analysis as against the conventional classical test theory (CTT) approach.
- Teachers should be encouraged to use three-option format in both formative and summative classroom assessment to enhance the phasing out of the 5-option convention.

- Test experts should put into practice, the corrective measure to address guessing especially in objective tests
- Seminars, symposiums and workshops on new testing techniques should be made available by test experts and psychometricians to teachers and other stakeholders in education especially at the secondary school level.

References

Aamodt, M. G., & McShane, T. (1992). A meta-analytic investigation of the effect of various test item characteristics on test scores and test completion times. *Public Personnel Management, 21*(2), 151-160.

Abad, F. J., Olea, J., & Ponsoda, V. (2001). Analysis of the optimum number of alternatives from the Item Response Theory. *Psicothema*, 13*(001),* 152-158.

Adegoke B. (2012) Statistical methods for behavioural and social science research, Everlasting Prints, Ibadan.

Costin, F. (1970). The optimal number of alternatives in multiple-choice achievement tests: Some empirical evidence of a mathematical proof. *Educational and Psychological Measurement,* 30, 353-358.

Crehan, K. D., Haladyna, T. M., & Brewer, B. W. (1993). Use of an inclusive option and the optimal number of options for multiple-choice items. *Educational and Psychological Measurement*, 53, 241-247.

Delgado, A. R., & Prieto, G. (1998). Further evidence favoring three-option items in multiple-choice tests. *European Journal of Psychological Assessment, 14,* 197-201.

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement.*(2nd Ed). Englewood Cliff, NJ: Prentice-Hall.

Farhady, J. & Shakery, S. (2000). Number of options and economy of multiple-choice tests. *Roshd Foreign Language Teaching Journal*, 15*(57),*132-141.

Grier, J. B. (1975). The number of alternatives for optimum test reliability. *Journal of Educational Measurement, 12*(2), 109-113

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed). Mawah, NJ: Lawrence Erhbaum.

Haladyna, T. M. & Downing, S. M. (2004). How many options is enough for a multiple-choice item? *Educational and Psychological Measurement*, 53, 999-1010.

Haladyna, T. M., Downing, S. M., & Rodriguez M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15*(3),* 309-334.

Landrum, R. E., Cashin, J. R., & Thies, K. S. (1993). More evidence in favour of three option multiple-choice tests. *Educational and Psychological Measurement*, 53, 771-778

Linn, R. L. & Gronlund, N. (2000). *Measurement and assessment in teaching* (8th ed). Columbus, OH: Merrill.

Linacre, J. M. (2008). A user's guide to WINSTEPS-MINISTEP: Rasch-model computer

programs. Chicago, IL: winsteps.com.

Lord, F. M. (1977). Optimum number of choices per item – a comparison of four approaches. *Journal of Educational Measurement*, 14, 33-38.

National Educataional and Research Development Council (NERDC), (2008) Annual Journal.

Okoli, C. E. (2005). *Introduction to educational and psychological measurement*. Lagos: Behenu.

Onunkwo, G. I. N. (2002). *Fundamentals of educational measurement and evaluation*. Owerri: Apple Publishers.

Ogunjimi M.O. (2009). Essentials of Educational Test, Measurement and Evaluation, Joytal Prints, Ibadan, Oyo State.

Peter I. N. and Louisa N. (2013): The Number of Options in a Multiple-Choice Test Item and the Psychometric Characteristics. Journal of Education and Practice, Vol.4, www.iiste.org,ISSN 2222-1735

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: a meta- analysis of 80 years of research. *Educational Measurement: issues & practice*, 24*(2),* 3-13.

Rogauch, A., Hofer, R., & Krebs, R. (2010). Rarely selected distractors in high stakes medical examinations and their recognition by item authors: a stimulation and survey. *BMC Medical Journal*, 10, 85-91.

Rogers, W. T., & Harley, D. (1999). An empirical comparison of three and four-choice items and tests: Susceptibility to test-wiseness and internal consistency reliability. *Educational and Psychological Measurement, 59,* 234-247.

Sidick, J. T, Barret, G. V., & Doverspike, D. (1994). Three alternative multiple choice tests: An attractive option. P*ersonnel Psychology*, 47, 829-835.

Shizuka, T., Takeuchi, O., Yashima, T, & Yoshizawa, K. (2006). A comparison of three- and four- option English tests for university entrance selection purposes in Japan. *Language Testing,* 23*(1)*, 35-57.

Straton, R. G., & Catts, R. M. (1980). A comparison of two, three, and four-choice item tests given a fixed total number of choices. *Educational and Psychological Measurement, PP*. 357- 365.

Tarrant, M., Ware, J., & Mohammed, A. H. (2009). An assessment of functioning and non- functioning distractors in multiple-choice questions a descriptive analysis. *BMC Medical Education*, 9 *(40)*. 1-8.

Tversky, A. (1994). On the optimal number of alternatives at a choice point. *Journal of Mathematical Psychology,* 1,*(2),* 386-391

Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and Evaluation in Psychology and Education* (8th Ed). Upper Saddle River, NJ: Pearson/Merril Prentice Hall.

Trevisan, M. S., Sax, G., & Michael, W. B. (1994). Estimating the optimum number of options per item using an incremental option paradigm. *Educational and Psychological Measurement*, 54*(1),* 86-91.

Udofia, I. G. R. (2008) on "Effect of number of options on students academic Performance", an Unpublished B.Ed. Project, University of Ibadan, Oyo State.

Woodford, K., & Bancroft, P. (2004). Using multiple choice questions effectively in Information Technology education. In R. Atkinson, C. McBeath, D. Jonas Dwyer & R. Philips (Eds), *Beyond the Comfort Zone:*