INTERNATIONAL EDUCATIONAL APPLIED RESEARCH JOURNAL
E-ISSN: 2456-6713 | Volume: 2 | Issue: 10 | Oct 2018

EFFECT OF NUMBER OF DISTRACTORS PER MULTIPLE-CHOICE ITEM ON DISTRACTOR PLAUSIBILITY IN A REPEATED MEASURE SETTING

A.M. Mamadu

Professor S.O. Emaikwu

O. Adikwu

Department of Educational Foundations and General Studies, University of Agriculture Makurdi, Benue State, Nigeria.

ABSTRACT

This study aims to determine the optimal number of distractors for multiple-choice test items. A Common Person Equating design was employed to compare distractor performance indices across three Mathematics test formats i, ii and iii, which were different only in the number of distractors per item. Format ii and iii were constructed by deleting one distractor from each item in format i and format ii respectively. Format i, ii and iii, each containing 20 multiple-choice Mathematics test items with four, three and two distractors per item, were administered to 169 Senior Secondary II (SSII) students—120 students from Golden College Yagba, and 49 students from Government Secondary School Makurdi, both in Makurdi, Benue State, Nigeria. The three test formats were administered by means of Common Person Equating with Counterbalancing. The Nedelsky Model was applied to compare distractor plausibility across multiple-choice Mathematics test items with four, three and two distractors per item. Distractor plausibility was found to be significantly increased with increased number of distractors, with four-distractor-items yielding more plausible distractors. It is hence recommended that item writers should make it an obligation to include four distractors or more on multiple-choice test items.

Key words: Distractors, Common Person Equating, Nedelskey Model, distractor plausibility, Repeated Measure.

INTRODUCTION:

Preparing the young ones to face future challenges and developing them to meet the manpower needs is one of the objectives of the Nigerian educational system. Schools need to conduct tests as yardstick for assessment as it is the most practical way for evaluation in education.
Tests were first used to assess performance of students in schools by the Chinese who were the first country to appoint civil servants on the basis of their competitive performance in achievement (Borris and Awodun in Agi, Aduloju and Iornienge, 2015). It was explained further by the authorities that by 155BC, Civil Service Examinations were in place and were used to select candidates for employment in the Chinese Imperial Service.

In Nigeria the first public test was conducted by the West African Examination Council (WAEC) for the award of certificates of comparable international standard in 1952 (Anderson as cited in Nwokora, 2010). In the wake of the Universal Primary Education (UPE) project National Teachers Institute (NTI) was established in 1976 to among other mandates award Teacher's Certificate 2 (TC II), National Certificate in Education, and Postgraduate Diploma in Education. In 1978 the Joint Admissions and Matriculation Board was established to conduct Unified Tertiary Matriculation Examination for candidates in tertiary institutions in Nigeria. To demonstrate craft level examinations which were hitherto conducted by City and Guilds, Pittman's and Royal Society of Arts, all of United Kingdom, the Federal Government of Nigeria established the National Board for Business Technical Education (NABTEB) in 1992. The military regime of General Abdusalami Abubakar, in April, 1999, converted the then National Board of Education Measurement (NBEM) to National Examination Council (NECO) to conduct tests such as National Common Entrance Examination (NCEE), Junior School Certificate Examination (JSCE), and Senior School Certificate Examination (SSCE).

Testing is a specific tool or procedure or a technique used to obtain response from the students in order to gain some information which provides the basis to make judgment or evaluation regarding some characteristics such as fitness, skill, knowledge and values. Testing as defined by Okoye (as cited in Nwokora, 2010) is an organised assessment technique which presents individuals with a series of questions or tasks geared towards ascertaining the individual acquired skills and knowledge. Testing is a systematic method of observing certain human behaviours, so that at the end of the observation, figures are assigned to the results of the observation. In other words, testing implies ascertaining the presence, quality, or genuineness of anything. The essence of testing is to disclose the latent ability of a testee. This is the type referred to in this investigation especially as it concerned Senior Secondary II (SS II) students.

PretzelRed (2013) classified tests under the following categories namely: classification based on manner of response (i.e. oral test and written test); classification based on method of construction (i.e. teacher-made test and standardised test); classification based on nature of answer (i.e. intelligence test, personality test, aptitude test, prognostic test, diagnostic test, achievement test, ipsative test, accomplishment test, scale test, power test, placement test, etc); and classification based on mode of scoring (i.e. objective test and essay test) amongst others.

Two (2) testing formats, based on mode of scoring, have been found useful in measurement namely objective test and essay test. Because essay tests require extensive time to score; and they encourage subjective criteria when assessing answers, there has been a drift towards objective test formats which rather requires less scoring time and encourages objectivity in scoring. Objective format questions are questions that require a specific answer. Pushpangathan (2016) averred that an objective question usually has only one correct answer and it leaves no room for opinion(s). Pushpangathan (2016) further suggested that objective tests could come in the form of: Multiple-choice (MC), True/false, Gap filling, Marching, Transformation, and Cloze (as in word register).

MC test has been found more dependable because it yields higher reliability; covers broader content; costs less for administration; takes less effort for scoring; and enhances prompt reporting and feedback. Structurally, MC questions have two parts: a stem ─ the question, problem, or task to be answered or solved; and a set of response options or alternatives, that is, the possible answers or solutions to the question (Onunkwo, 2002). The options comprise the correct answer called the key and one or more incorrect or less appropriate answers called the distractors.

As mentioned earlier, the construction of MC items takes a lot of time and effort. The most problematic and difficult areas in developing MC items is writing, not the key, but plausible and functional distractors (Haladyna, 2004); (Haladyna & Downing, 1993). Haladyna, Downing, and Rodriguez (2002) recommended writing as many plausible distractors as possible, for instance, four, five options among others. There are very significant arguments in favour of five options with the following assertions.

- That lower number of options, such as three, increase, to an alarming high degree, the chances of successful random guessing and the extent of guessing effects, such as over-estimation of students' achievement over 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 fewer number of options decreases the psychometric quality of the test score.
- And that this psychometric limitation can only be corrected by using five or at least four options per item (Farhady & Shakery; Abad, Olea & Ponsoda; Woodford & Bancroft, as cited in Nwadinigwe & Naibi, 2013).

Many researchers and evaluators have increasingly relied on MC items rather than other forms of tests owing to their higher reliability; broader content coverage; less cost of administration; ease of scoring; objectivity of scoring; prompt score reporting and publishing; fast feedback amongst others (Haladyna, 2004). Even though antagonists of MC items have argued that memorisation/rote learning and guessing come with MC tests, protagonists of MC tests have put forward superlative arguments in favour of the use of MC in measurement. They have suggested the framing of questions that are locally independent, on a total test, as this is capable of bringing the issue of memorisation and/or guessing to a checkmate (i.e. answer to item 5, say, should not lead to answer to item 6). Also, constructing items devoid of flaws like cues and irrelevant difficulty is expedient for placing testees in their ability levels as against the unfounded belief that MC tests would place testees where they ought not to be. They have also suggested the inclusion of plausible distractors only on the responses. A test item with all plausible distractors sets the testee thinking; it demands the testee to have had complete grasp of the subject matter, rather than mere rote, for him to get it right. Responses that distract plausibly would conceal the key for unknowledgeable candidates. This technique if properly harnessed at the development stage of the MC test items is a vital tool for eliminating the guessing parameter which in turn depends on the plausibility of

The objective of this study is to establish whether or not there is a relationship between the number of distractors per multiple-choice item and distractor plausibility of the item.

The following research question is posed to guide the study. What is the relationship between the number of distractors per multiple-choice item and distractor plausibility of the item?

Null hypothesis was formulated to guide the study, and was tested at 0.05 level of significance.

H_o: There is no significant mean difference between five-, four- and three-option Mathematics multiple-choice test forms with regard to option distractor index.

The outcome of this research work would be of immense benefit to the government, school administrators, teachers, students, Ministry of Education, researchers and other stakeholders in the society.

The study looks into number of multiple-choice responses in relation to option distraction. The study uses Mathematics as the subject area while the category of students is Senior Secondary II (SSII). The study covers the eleven (11) Council Wards of Makurdi Local Government Area of Benue State and is carried out in government and grant-aided secondary schools only. The study has 20 Multiple –Choice questions drawn from Logarithm.

The Concept of Distractor Plausibility:

Distractors are usually checked to see whether they are plausible or functional enough to be retained. This activity is a component of what is referred to as item analysis. For a distractor to be functional or plausible, it must present confusion to those who are not sure of them (Emaikwu, 2011). The distractor has to appear to some examinees as the correct answer and also has to appeal more to the low ability group than to the high ability group.

A distractor is any of the incorrect responses in a multiple-choice question while its index is the disparity between the proportion of the candidates choosing the wrong option in the upper group and those choosing the wrong option in the lower group mathematically written as:

$$Distractor\ Index = (W_U - W_L)/\frac{1}{2}N_{U+L}\ Where;$$

$W_U = Upper\ group\ candidates\ selecting\ the\ particular\ distractor$

$W_L = Lower\ group\ candidates\ selecting\ the\ particular\ distractor$

$N_{U+L} = Total\ number\ of\ candidates\ in\ both\ upper\ and\ lower\ groups$

The plausibleness of a distractor is determined by its movement along the negative-positive axis. A distractor which index is positive is termed implausible while a distractor with zero (0) index is said not to distract at all. Conversely, a negative distractor index posits that candidates who are not knowledgeable in the tested concept will not get an item correctly, while the knowledgeable candidates will get the item right.

In empirical item analysis, distractor indices are to be computed for all incorrect options in a given test. It is not certain how many options per question would yield a plausible distractor! This is the crux of the current investigation.

Theoretical Basis:

The study is based on the Nedelsky's Unlikely-Distractor Model. The model was advanced by Leo Nedelsky in the year 1954.

The model states that if a MC item i has $J_i + 1$ responses arbitrary indexed $0,1, \ldots, J_i$ then 0 indexes the correct response. The borderline test-taker responds to a Multiple-Choice (MC) question by first isolating the responses he recognises as wrong and then guesses at random from the remaining responses. The first entry is fixed at 0 because it is assumed that the correct response is always eventually selected.

If S_i $represents\ options;$

$J_i\ represents\ distractors;$

$S_{ij}\ represents\ unlikely\ distractor;$

$S_{i0}\ represents\ the\ key\ by\ random\ guessing;$

$\sum_{j=1}^{J_i} S_{ij}\ represents\ sum\ of\ all\ unlikely\ distractors;$

$\zeta_{ij}\ represents\ difficulty\ to\ recognise\ unlikely\ distractor;$

$and\ \zeta_{i0}\ represents\ plausible\ distractor; then$

$P_r(S_{ij} = 1|\theta) = \frac{exp(\theta - \zeta_{ij})}{1+exp(\theta - \zeta_{ij})}$ is the probability of selecting the key upon isolation of all unlikely distractors S_{ij}.

The implication is that the closer the S_{i0} component of responses on SSMT tends to zero ($S_{i0} = 0$) the less unlikely the responses get. Corollary, the higher the ζ_{i0} component of SSMT responses ($\zeta_{i0} = \infty$) the more difficult it becomes to recognise unlikely distractors in SSMT. $\zeta_{i0} = \infty$ eliminates unlikely distractors from the pool of responses for the unknowledgeable testees. This presupposes the strengthening of the ζ_{i0} component of the Nedelsky Unlikely-Distractor Model by ensuring the inclusion of plausible distracters only on SSMT.

Nedelsky's (1954) Unlikely-Distractor Model advanced the index $0,1, \ldots, J_i$ as responses for each item i and suggested that the unknowledgeable SSII student will never reject the correct response 0; but will always select it because he/she firstly removes the responses he/she deems wrong from the pool of responses and then eventually selects the corrects response by sheer guess. Nedelsky's approach presupposes that distractors should be made as plausible as possible by test constructors in order to ameliorate guessing effect.

METHODS:

The study adopted Quasi Experimental Research design. Specifically, it was One-group Posttest-only Quasi Experimental Research design. This is because the study has just one participating group, just one case of intervention, and just one phase of observation.

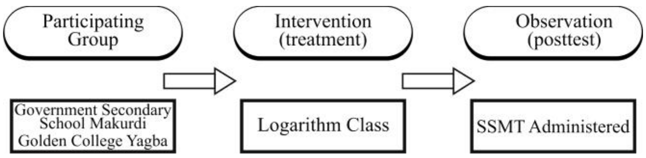


Figure 1: One-group posttest-only Quasi experimental design

The population is 1,791comprising all Senior Secondary II (SS II) students from 19 public secondary schools in Makurdi metropolis (Benue State Ministry of Education, Science and Technology, 2016). All SS II students offer Mathematics since the subject is a compulsory one. The study has a sample size of 169 students comprising 49 students from Government Secondary School Makurdi and 120 students from Golden College Yagba. The sample size of 169 is intact classes of 49 and 120 in the respective two (2) schools sampled. The researcher employed purposive sampling technique to select Government Secondary School Makurdi and Golden College Yagba. The researcher adopted purposive sampling technique because the subjects form intact classes, and also this category of the population has been taught Logarithm.

The instrument used for data collection for the study was a self-constructed teacher-made achievement test on Logarithm. The test is known as Senior School Mathematics Test (SSMT). As a first step towards development of SSMT, 24-item trial test was written (see appendix F) based on the Nigerian Educational Research and Development Council (NERDC) syllabus, with strict adherence to the test blue-print (see appendix E) developed to guide the SSMT construction process. The researcher carried out classroom teaching in Second Term to be sure the students had the required exposure to the topic (Logarithm) (see lesson plan on appendix D). Modified Common Person Equating was adopted to control for differential examinee proficiency. Precisely, Modified Common Person Equating was applied using horizontal linking because in this investigation all the items are anchor items across the three (3) forms of SSMT; and also previous ability is fixed. However, instrument was not designed using common item equating because all items are common (same) across all the three (3) forms of SSMT, and it thus implies that there are no anchor items.

The pilot test contained 5 response options on each item (A-E). It also had eleven (11) items drawn from place value of numbers and logarithm; nine (9) items were constructed from logarithm of numbers greater than 1; while logarithm of numbers less than 1 contributed four (4) items in development of SSMT (see appendix E). To develop SSMT, a 24-item multiple-choice test was administered on 90 SSII students of Lobethas Unity College Daudu. This was done in an attempt to adhere to a general item-writing guideline suggested by Haladyna (2004) which states: "use typical errors of students when you write distractors".

Item analysis was performed on the trial test containing 24 items (see appendix M). Anastasi; Thorndike and Hagen as cited in Alonge (2004) opined that items with extreme low or high difficulty index contribute nothing to the reliability and validity of the test. More so, the difficulty level of the items that constitute a test determines not only the mean difficulty level of the test but also the spread of the test scores; and the maximum spread of the total score is obtained with items whose difficulty levels hover around 0.5. This is the basis for the following item analysis.

Items 6, 12, 18 and 24 were discarded for being either too difficult or too easy. Item 6 had difficulty index of 0.21, item 12 yielded a difficulty index of 0.17 while item 24 had difficulty index of 0.23 which shows they (6, 12 and 24) were too difficult for SS II. Yet, items 12, 18 and 24 presented with a compounded case by possessing ineffective (non-functional) distractors as well as being too difficult or too easy for the level and hence they were discarded. Item 12's option D was ineffective/non-functional; option E of item 18 did not distract at all (distractor index equal 0) while item 24 had all its options except option E as ineffective/non-functional (i.e. having positive distractor indices). In another scenario, item 18 was too easy, with difficult index of 0.77. However, item 11 was retained because its option E that did not distract was modified. Twenty (20) items were selected to form the final form of SSMT (see Appendix N).

The final form of SSMT contained 20 items that were selected after item analysis was carried out. This form of SSMT is sub-divided into three (3) forms (forms i, ii and iii) and is the actual tool for data collection in this investigation. Form i has five (5) responses $(A-E)$; form ii has four (4) responses $(A-D)$; while form iii has three (3) responses $(A-C)$, contained on each item. This agrees with the opinion of Grier as cited in Gerald (2008) that having greater than or equal to fifty-four (≥54) responses in the total test is sufficient. So a 20-item multiple-choice test is workable since, at least, 3 options would yield >54 $(20 \times 3 > 54)$. The five options in the trial test were transferred to form the five options of the 5-option form SSMT. Item Stack Shift was applied to get the 2nd form (4-option form) i.e. the first eight (8) items of the 5-option form were pulled to the bottom of the stack.

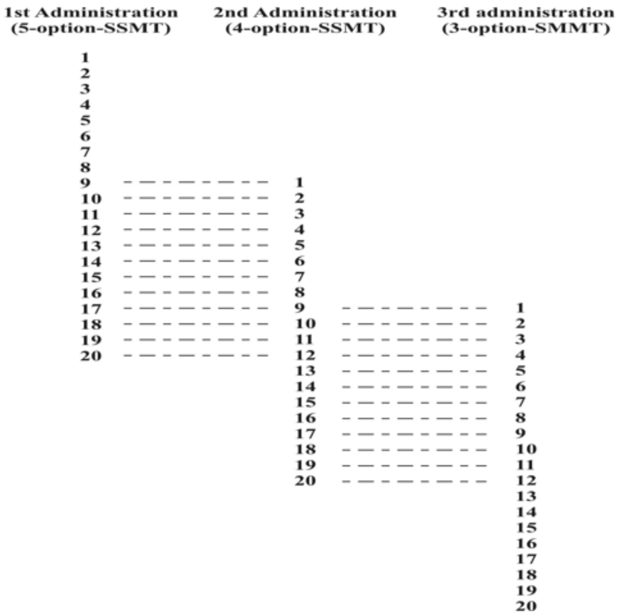


Figure 2: Item Stack Shift showing how the three forms of SSMT were constructed.

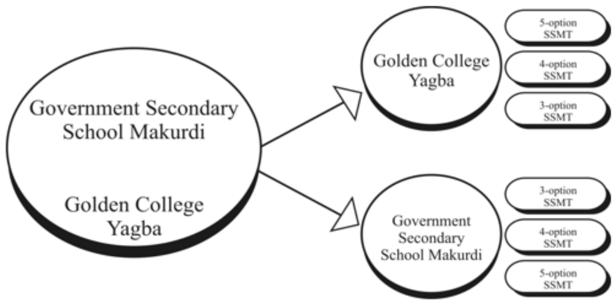
Also, the fifth (5^{th}) option of the 5-option form was eliminated where the option was a distractor. However, where the 5^{th} option was the key, it was moved one place up, to be the forth option; and the existing (forth) option was discarded. Item Stack Shift technique was again used to get the 3-option form except for the special approach: the options were eventually shuffled. Item Stack Shift is a special and peculiar technique that employs the pushing (adding to the top), and pulling (taking way from the bottom) of a particular number of items to/from the pool of items (item stack), and subtly omitting a distractor and/or moving a key one step up.

SSMT was presented to one expert in the field of Mathematics and two experts in Measurement and Evaluation. These experts were requested to judge the worth of each item against the following criteria:

- Whether or not items on SSMT conform to test blue print;
- Whether or not items on SSMT conform to all the statistical indices of item analysis;
- Whether or not items on SSMT are devoid of flaws like clues and irrelevant wordings;
- Whether or not solution to one item on SSMT leads to solution to another.

Kudder-Richardson ($K - R_{20}$) formula was used in computing the reliability of SSMT. This formula is applicable to dichotomously scored tests (one point each for correct answer, and zero for incorrect answer); and does not assume equal item difficulty (Alonge, 2004; Emaikwu, 2011). $K - R_{20}$ was found to be 0.66 which is moderate enough for the instrument to be regarded as internally consistent. This is in agreement with the opinion of Salvucci, Walter, Conley, Fink, and Saba as cited in Tan (2009) that between 0.50 and 0.80 the reliability is moderate.

Common Person Equating with Counter balancing approach was used during the data collection stage. Common Person Equating with counterbalancing approach was used to annihilate the potential for order effects which is usually associated with equating procedures. To achieve Common Person Equating with counterbalancing in this study the researcher purposively segmented the two schools sampled into groups (random samples) and administered SSMT in the order shown in figure 3.



Empirical Item analysis was carried out to answer the research question while Repeated Measure One Way Analysis of Variance (ANOVA) statistical technique was used for testing the null hypothesis at 0.05 level of significance. The choice of Repeated Measure ANOVA is borne out of the fact that the same instrument was administered three (3) times to the same group of examinees under three (3) different conditions; and the technique removes variability due to individual differences.

RESULTS:

Research Question:

What is the mean differential effect of five-, four- and three-option Mathematics multiple-choice test formats with regard to option distractor index?

In order to answer this research question distractor indices of all possible distractors on each item were computed and compared across the three SSMT formats as presented in Table 1.

Table 1: Overall option distractor indices across all the three formats of SSMT

Item	Distractor Index											
	1st Administration (5-option form)					2nd Administration (4-option form)				3rd Administration (3-option form)		
	A	B	C	D	E	A	B	C	D	A	B	C
1	-0.15	-0.22		-0.20	-0.17		-0.13	-0.10	-0.15	-0.02	-0.04	
2		-0.20	-0.28	-0.17	-0.15	-0.09		-0.28	-0.09		-0.10	-0.04
3	-0.13		-0.20	-0.07	-0.17	-0.11		-0.07	-0.22	-0.10	-0.13	
4	-0.28	-0.17	-0.15	-0.20		-0.11	-0.11		-0.07	-0.02		-0.11
5	-0.20	-0.15		-0.17	-0.22	-0.04	-0.20		-0.11	-0.11	-0.10	
6	-0.15	-0.22		-0.20	-0.17	-0.09	-0.09	-0.07			-0.02	-0.04
7	-0.24	-0.22	-0.24	-0.20		-0.22		-0.04	-0.07	-0.02		-0.02
8		-0.20	-0.24	-0.20	-0.15		-0.13	-0.11	-0.10	-0.11	-0.07	
9		-0.15	-0.20	-0.20	-0.30	-0.15	-0.10		-0.10	-0.10	-0.04	
10	-0.17		-0.15	-0.24	-0.24	-0.17	-0.15		-0.13	-0.10	-0.07	
11	-0.22		-0.22	-0.24	-0.28		-0.17	-0.11	-0.10	-0.04	-0.07	
12	-0.24	-0.20		-0.22	-0.20	-0.02		-0.20	-0.07		-0.02	-0.02
13	-0.17	-0.20		-0.20	-0.13	-0.11	-0.13		-0.15	-0.02		-0.04
14	-0.24	-0.24	-0.17		-0.22		-0.10	-0.13	-0.15	-0.02	-0.02	
15	-0.22		-0.17	-0.15	-0.17	-0.10		-0.10	-0.11		-0.10	-0.07
16		-0.30	-0.11	-0.20	-0.17	-0.11	-0.13		-0.02	-0.04		-0.07
17	-0.24	-0.17	-0.15	-0.20		-0.13	-0.07		-0.20		-0.07	-0.02
18	-0.15	-0.24		-0.24	-0.17	-0.10	-0.17		-0.11	-0.04		-0.02
19	-0.26	-0.20	-0.17	-0.26		-0.13		-0.15	-0.10	-0.10	-0.10	
20	-0.17		-0.24	-0.22	-0.15		-0.07	-0.09	-0.11	-0.04		-0.04
Option Distractor Totals	-3.23	-3.08	-2.69	-3.78	-3.06	-1.68	-1.75	-1.45	-2.16	-0.88	-0.95	-0.49
Mean Option Distraction	-0.16	-0.15	-0.14	-0.19	-0.15	-0.08	-0.09	-0.07	-0.11	-0.04	-0.05	-0.03
Mean Totals	-0.79					-0.35				-0.12		
Mean Differentials of Distractors			0.44					0.23				

From Table 1, option distractor index of 5-option items was smaller in magnitude than the option distractor index of items with 4 options. Similarly, option distractor index of items with 4 options was seen to be smaller in magnitude than that of items with 3-options. This result is indicated by Mean Option Distraction and Mean Totals respectively. It was observed through inspection that Mean Totals of 5-option items was -0.79; that of 4-option items was -0.35; while that of items with 3 options was obtained to be -0.12. Mean differentials of Distractors between 5- and 4-option items was 0.44 while the Mean differentials of Distractors between 4-option items and 3-option items was 0.23. It is important to note that the blank spaces in Table 3 represent the key, which ought not to appear in the table showing distracters.

Ho: There is no significant mean difference between five-, four- and three-option Mathematics multiple-choice test formats with regard to option distractor index. The ANOVA result of this research hypothesis is shown in Table 2.

Table 2: Repeated ANOVA result of five-, four- and three-option Mathematics MCT format with regard to Distractor Index.

Tests of Within-Subjects Effects							
Measure: MEASURE_1							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Distractor_index	Sphericity Assumed	.188	2	.094	48.447	.000	.718
	Greenhouse-Geisser	.188	1.839	.102	48.447	.000	.718
	Huynh-Feldt	.188	2.000	.094	48.447	.000	.718
	Lower-bound	.188	1.000	.188	48.447	.000	.718
Error(Distractor_index)	Sphericity Assumed	.074	38	.002			
	Greenhouse-Geisser	.074	34.950	.002			
	Huynh-Feldt	.074	38.000	.002			
	Lower-bound	.074	19.000	.004			

The assumption of sphericity of the three forms of SSMT was used to test hypothesis 2, and it was tested with Mauchly's test of sphericity (see appendix ZB). Sphericity was assumed at $Sig. > 0.05$. From the Mauchly's test of sphericity, $Sig. = 0.44$. Therefore, Mauchly's test: $\chi^2(2) = 1.64, P = 0.44$ did not violate the assumption of sphericity. Since data from Mauchly's test seem spherical, the results of Greenhouse-Geisser, Huynh-Feldt and lower bounds in Table 4 were ignored, and the basic (uncorrected) results were simply interpreted. The difference between the means of $5-, 4-$ and $3-$option SSMT indicated by $F(2,38) = 48.45, P = 0.00$ is statistically significant at 0.05 level of significance and hence, the null hypothesis is rejected.

DISCUSSION:

Table 1 reveals that options on SSMT exhibited incremental weakening distractor plausibility. For instance, option A of item 17, 9 and 1 was -0.24, -0.15 and -0.02 respectively. This result shows the distractor index of option A kept on increasing in magnitude. The distractor indices increased by 0.09 units when 4-option item-SSMT was administered; and amplified by 0.13 units when 3-option item-SSMT was administered (see figure 2 for item shift). By implication, the distractor index of option A was proven to be drifting away from negative upon 2nd and 3rd administration, and hence, it distracts more plausibly when 5-option SSMT was administered. Option B of item 17 exhibited similar distractor behaviour as its distractor index increased by 0.07 units from -0.17 in the 1st administration, to -0.10 in the 2nd administration, and further increased by 0.06 units from -0.10 in the 2nd administration, to -0.04 in the 3rd administration. This positivity tendency is an indication that 5-option items possess more plausible distracters. The difference between means indicated by $F(2,38) = 48.45, P = 0.00$ shown in Table 4 further proves that there is a difference in distractor plausibility between 5-,4- and 3-option item-SSMT.

The findings revealed that option distractor index becomes larger upon administering 4- and 3-option SSMT. By implication multiple-choice options distract more plausibly when 5-option-test items are used. This finding negates the finding of Haladyna and Downing (1993) that the number of effectively performing distractors per item was approximately 1, as the finding was in support of 3-option items.

CONCLUSION:

Based on the findings of this study the researcher concludes that 5-option multiple-choice responses distracted more plausibly in a repeated measure setting. It is therefore recommended that item writers should make it an obligation to include four distractors or more on multiple-choice test items.

References:

1. Agi, C.I., Aduloju, M.O., & Iornienge, T.M. (2015). Computer-aided testing: a panacea for examination malpractice. African Journal of Theory and Practice of Educational Research, 1, 76-90.
2. Alonge, M.F. (2004). Measurement and evaluation in education and psychology. Ado-Ekiti: Adedayo.
3. Emaikwu, S.O. (2011). Fundamentals of test, measurement and evaluation with psychometric theories. Makurdi: Selfers.
4. Haladyna, T.M., & Downing, S.M. (1993) How many options is enough for a multiple-choice test item? Educational and Psychological Measurement, 53, 999-1010.
5. Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A review of multiple-choice item writing guidelines for classroom assessment. Journal of Applied Measurement in Education, 15(3), 309-334.
6. Nwadinigwe, P.I., & Naibi, L. (2013). The number of options in a multiple-choice test item and the psychometrics characteristics. Journal of Education and Practice, 4, 28.
7. Nwokora, B.I. (2010). Assessment of the implementation of government measures for controlling examination malpractices in Ebonyi State secondary schools (PhD thesis).
8. PretzelRed. (2013, February 28). Educational measurement and evaluation [SlideShare]. Retrieved from https://www.slideshare.net/mobile/edtechred/educational-measurement-and-evaluation
9. Pushpangathan, K. [Kavukavya]. (2016, June 14). Types of test items [SlideSh are]. Retrieved from

https://www.slideshare.net/mobile/kavukavya/types-of-test-items