International Journal of Spectrum Research in Education & Humanities (IJSREH)
1(4), October-December, Pages 216-225
© Noble City Publishers ISSN: 3092-9539

https://doi.org/10.5281/zenodo.17987335

Performance of Senior Secondary School Students on Mathematics Test Items Arranged in Bottom-Top and Variation in Item Numbering Formats

Goodluck Uwaks Miwari PhD[1] & Friday Ukalinwana Ejimaji[2]

Department of Educational Foundations, Rivers State University
Nkpolu-Oroworukwo, Port Harcourt.
[1]goodluck.miwari@ust.edu.ng & [2]Ukalison27@gmail.com

ABSTRACT

Many scholars have posed a strong argument regarding the effect of test items arrangement on student performances. While some argue that such arrangement does not have any serious effects on students score, others have maintained that such a sensitive aspect of test construction cannot be entirely exonerated from the total outcome of students score. This study investigated students' performance with mathematics test items arranged in bottom-top (descending order of difficulty) and variation in items numbering (VIN) formats. With a quasi-experimental research design, two research questions with associated two null hypotheses guided the investigations. From a population of 52,695 Senior Secondary School two (SS2) students, multistage with simple random sampling techniques was adopted to select 900 participants from 18 intact classes of 50 students each. A test blueprint was added as a guide to construct a 60 objective items Mathematics Achievement Test (MAT) which was used for data collection. MAT was validated by experts in Mathematics and test experts from Measurement and Evaluation. It has reliability indices K = 0.82 and r = 0.79, established using Cohen's Kappa (K) procedure and Kuder Richardson formula 20 (KR_{20}), respectively. Mean score differences and t-test were used for data analyses. Result indicated that arrangements with variation in items numbering (VIN) has significant positive effect on examinees performance in Mathematics while bottom-top had negative effect. It was recommended that as much as possible, bottom-top format should be avoided for its negative consequences as revealed from the study. Arrangement based on variation in item numbering is highly recommended for its merit of positive effect and inhibition of examination sharp practices by students.

Keywords: *Bottom-top format, Examinee, Items arrangement, Traditional (half hazard) format, Variation in item Numbering (VIN) format*

INTRODUCTION

When feedback is given to either the learner or his sponsor, teachers use such expressions as excellent, brilliant performance, good, pass, fail, poor, and the like. These expressions are descriptive of the level of performance based on the measure of achievement of the learner. In other words, a learner's performance in a subject can be excellent, brilliant, good, poor or very poor. These are however, relative and criterion referenced because what may constitute good performance to one evaluator at one point may not be to another. Performance in a subject is assessed through testing and measurement. Testing and measurement are synonymous and are interchangeably applied as the process of determining the characteristics or traits or behaviour in an individual. The tool mostly used for this process is the test. Ukwuije (2009) defined a test as a series of questions given to the testees or examinees to be answered in order to measure performance or knowledge. A test according to Asuru (2016). is a device or instrument for obtaining a sample of behaviour of pupils. Citing Ali (1986), Asuru (2016) posited that in broad terms a test is a systematic procedure for observing a person's behaviour and describing it by means of a numerical scale or a category system. This device or systematic procedure could be oral, paper and pencil or observational. Koul (2018) declared that a test consists of a series of tasks which the

subject is required to perform, and they measure general mental abilities or intelligence, specific abilities or aptitudes, creativity, achievement, personality traits and adjustment, interest and values. Anastasi and Urbina (2006) viewed psychological test as an objective and standardized measure of a sample of behaviour, and further explained that traditionally the function of a psychological test has been to measure differences between individuals and between the reactions of the same individual under different circumstances. According to these scholars, all tests: cognitive or non-cognitive, psychological or physiological are alike insofar as observations are made on a small but carefully chosen sample of an individual behaviour. It is noteworthy, that all the definitions emphasized obtaining a sample of behaviour.

Whether a series of questions to be answered or several independent but related tasks to be performed, and whether the procedure will be oral, paper and pencil or observational, they must be arranged in a particular fashion. Test item arrangement is the positioning of test items in a particular order. It refers to the organization and sequencing of test questions items or task situation based on a particular order. The order, among several others, could be content based arrangement where items are grouped by Content or topics to assess specific knowledge or skills. In this arrangement, it is either items from topics treated first come up or appear first followed by topics treated last, or the reverse. In other words, it is *first topic first (FTF) or last topics first (LTF)*. Arrangement could be Difficulty-based, suggesting *top-bottom* or ascending order of difficulty in which case the less difficult or easier items come first followed by the more difficult ones or *bottom-top* or descending order where the most difficult items appear first followed by the easier ones. It could be Format-based arrangement whereby items are grouped by formats such as short answer, multiple choice or essay questions. Other arrangement types are *variation in items numbering (VIN)* in same test for different examinee, and the usual *traditional type* where items are arranged randomly. The emphasis in this work is the bottom-top and variation in item numbering (VIN).

The arrangement of test items beginning with easiest through the easier to the difficult or hard ones, conceptualized in this work as *Top-Bottom*, can be likened to the structure of a tree where the leaves are lighter than the stem (middle) which is lighter than the base or root which is the heaviest and carrying more load. The easy items are the leaves (top); the less easy ones are the middle (in between) while the hard or difficult ones are the base or roots (bottom), carrying the heaviest load. The conceptualized *Bottom-Top* arrangement is the reverse of Top-bottom and it involves positioning of test items in such a way that the more difficult ones come first followed by the difficult ones, then the easier. This is typically the reverse of the tree structure.

Variation in Items Numbering (VIN) is arrangement of test items in such a way that no two examinees sitting by one another (adjacent examinees), have their items numbered the same. In other words, item 1 for examinee A may be item 15 for examinee B and item number 50 for examinee C. Advocates of this format like Ukwuije and Orluwene (2012) opined that in addition to its influence on test usefulness criteria, it helps to check examinees from copying one another. Random or Mixed Format which is the Traditional format is the arrangement of test item without following a defined sequence. Difficult and easy items as well as topics treated come haphazardly, yet all examinee have the same item serial numbering. This is the commonest and the traditional pattern of item arrangement. Assessment agencies often time adopt this format,

Performance in a subject is how well or badly an examinee or a learner does in that subject. A learner's or testee's level of achievement, accomplishment, and fulfillment in a given subject is measured as the learner's performance, it is the extent to which an individual or an examinee has achieved set objectives in the subject or has demonstrated proficiency in the subject. Performance refers to a person's achievement in a given task including academic area (e.g reading or language arts, mathematics, science and other areas of human learning). Academic performance relates to the academic subjects a child studies in school and the skills the child is expected to demonstrate mastery in (Kathryn, 2021). Academic performance refers to excellence in all academic disciplines in a class as well as extra-curricular activities. including sporting behaviour, confidence in communication skills, and others. Steinberg (2010) posits that academic performance encompasses students' ability and performance, it is multidimensional and intricately related to human growth in cognitive, emotional and social development; it reflects the whole child, it is not related to a single instance, but occurs across time and levels, through a student's life in school and into post-secondary years and working life. Academic performance refers to how well a

student is accomplishing his tasks and studies in school and is evaluated in a number of ways. For regular grading, students demonstrate their knowledge by taking written and oral tests, performing presentations, submission of homework and participating in class activities and discussion. Teachers evaluate students by way of class work, daily or weekly assignments, and weekly or terminal tests and examinations to describe how well a student has done. The arrangement of the items as noted earlier, may be a determining factor to the performance of the testee, in any field of study.

Commenting on items arrangement and examinee performance, Gholam-Reza and Hanieh (2019) had opined that arrangement of items in a test is immaterial relating to any significant effect on the test takers performance but exerts influence on the test usefulness criteria to a varying extent. They maintained that test method, test content and test type are believed to affect test takers performance and have close connection with test characteristics. However, Tei-firstman (2011) cited Mehren and Lehman (1975) who identified a "sequence effect" which occurs in a test taking environment and argued that a testees motivational moral may be altered by negative effect resulting from frustration by a particular item. Having worked so hard and tirelessly preparing for a test only to be challenged from the onset with items he is unable to tackle, a test taker becomes very frustrated. Such a traumatic experience will affect his approach and responses, irrespective of whether the subsequent questions are easy or not.

Abbasian *et al.,* (2019) found out that respective statistical analyses revealed contradictorily that test item arrangement did not have any significant effect on the performance of the test takers. However, that test usefulness criteria proved with varying extent, to be subject to test item arrangement. In the same vein, Item facility (IF) and item discrimination (ID) were also affected in the light of test methods facets. They therefore recommended further studies on the contradictory findings while test developers, classroom teachers and all practitioners are advised to pay due attention to test method facets in their educational assessment decisions as the learners' performance, nature of the measurement devices as well as the nature of the construct itself are all affected by test method factors.

Barbara (2015) examined item arrangement and knowledge of arrangement on test scores. Three item orderings (easy-to-hard, spiral cyclical, and random) and two levels of knowledge of ordering (informed and non-informed) were used to study differences in test results, students' perception of the tests fairness and difficulty, and students' estimation of test performance. Result showed that no significant order effect was found ($\alpha = 0.05$), results were adjudged to be consistent with researches which have no differences in test scores when different orderings are used. No significant effects were found for knowledge of the orderings or for the interactions of knowledge of arrangement and order ($\alpha = 0.05$). Knowledge or lack thereof, of items arrangement is therefore not a factor influencing test scores or students' perception about the test for any of the ordering studied.

However, Ollennu and Etsey (2015) found a contrary result when they studied the impact of item position in multiple choice tests on students' performance at the Basic Education Certificate Examination (BECE) level in Ghana. With a sample of 810 Junior Secondary School (JSS) from different schools, they adopted Quasi-experimental research. A multiple-choice instrument consisting of 40 items in each of English, Mathematics and Sciences was used for data collection. The items of the instrument were arranged using the difficulty order to obtain the 3 treatment groups of random (RDM), easy-to-hard (ETH) and Hard-to-Easy (HTE). At a 0.05 significance level, the data collected were subjected to statistical analysis using ANOVA. The findings revealed that for English language, Mathematics and the Sciences at the BECE level, when item order was altered, the difference in performance was statistically significant. They concluded that the proposition of using re-ordering of items of an objective test to curb examination malpractices may not be the best option after all especially in English Language, Mathematics and Science at the BECE level. The researchers recommended the investigation of other methods.

Statement of the Problem

While examination umpires in Nigeria such as the West African Examinations Council (WAEC), the National Examinations Council (NECO), and the Joint Admission and Matriculation Board (JAMB) in charge of the Unified Tertiary Matriculation Examination (UTME), every season, decry the poor performance of candidates in quantitative subjects, similar reports are also found from other bodies and

for other subjects. Most reports attributed the poor performance to the fact that candidates did not demonstrate manipulative skills.

Toube (2019) observed that many scholars have posed a strong argument regarding the effect of test item arrangement in student performances. While some argue that item arrangement does not have any serious effects on students score, others have maintained that such a sensitive aspect of test construction cannot be entirely exonerated from the total outcome of student's scores (Essien,2001).
Orluwene (2012), while discussing item assembly, prescribed an arrangement pattern in which items dealing with the same content area should be placed together and, in each content, item should be ordered according to their difficulty levels. Hopkins and Antes were quoted in Orluwene (2012) to have opined that the most effective method for organizing items is to arrange sections of the test first by item type, second, grouping within item types by subject matter, and third, ordering in ascending level of difficulty within item types grouped by subject matter. Martuza cited in Orluwene (2012) subscribed to Hopkins and Antes view with the claim that the arrangement based on difficulty level is to motivate the examinees into the test and helps minimize anxieties and frustrations which tend to accompany or encounter with difficult items. Onunkwo (2002) however, advocated arrangement with the easiest items appearing first but in the order by subject matter to ensure that test contains only related materials and enables testees concentrate on a single area of the content at a time. These are laudable propositions but require empirical confirmation without which the gap continues to exist. It is against this background that this study examined the performance of senior secondary school students when Mathematics Test items are arranged in Bottom-top and Variation in Item Numbering (VIN) Formats. In this wise, the study specifically attempts to answer two basic questions:

1. What effect does Bottom-top (hard to easy) arrangement of test items has on examinees performance in mathematics?
2. What effect does test items arrangement based on Variation in Items Numbering (VIN) in the same test for individual testee, have on students' performance in mathematics.

The study further tests two null hypotheses associated with the foregoing questions as follows:

1. Bottom-top test item arrangement has no significant effect on students' performance in mathematics in senior secondary schools in Rivers State.
2. There is no significant effect of test item arrangement based on variation of item numbering in the same test for individual examinee on students' performance in mathematics.

RESEARCH METHOD

The quasi-experimental research design was adopted for the study as advocated by Kpolovie and Kuol, cited in Obilor and Miwari (2023), for situations that do not permit the control and manipulation of all relevant variables, but allows for the waiver of confounding variables that are impossible to eliminate or control, and when the researcher is interested in independent variables which cannot be randomly assigned. Best and Kahn (2006), illustrating with the counterbalanced type of quasi-experimental research design, observed that this design is often used in classroom experiments when experimental and control groups are such naturally assembled groups as intact classes. Experimental control derives from having all the subjects received all the treatment conditions but in different order or form as depicted below:

Table 1: Counterbalanced Type of Quasi-Experimental Designs.

Replication	O_1X_1	O_2X_2	O_3X_3
1	Group A	B	C
2	Group B	C	A
3	Group C	A	B

X= Treatment, O = Observation
Source: Research in Education (page 200)

The participants in the study were students of intact classes in classroom situation, and so it was not possible to randomly assign the independent variables, more over the human element made it

impossible to partial out the effect of confounding variables due to human behaviour, hence the quasi-experimental design was most suitable for the study.

The study was carried out in Rivers State of Nigeria. A state politically stratified into three senatorial districts, and consisting of rural and urban as well as upland and riverine communities. Rivers State is located in the Niger Delta region of the South-South geopolitical zone of Nigeria. Available 2025 statistics from the Rivers State Senior Secondary Schools Board (RSSSSB), shows that the state has a total of 278 Senior Secondary Schools. The population of the study is 52,695 Senior Secondary School Two (SS2) students of all public senior secondary schools in the state. Confident that by SS2, almost all sections of the school syllabus have been covered, the researchers chose this class level believing that it is capable of providing the relevant data necessary and required for this study.

Multistage and simple random sampling techniques were adopted to select a sample of 900 SS2 students from 18 intact classes of 50 students each. Simple random sampling technique was initially adopted to draw six (6) schools from each of the 3 senatorial districts in the state, giving a sample of 18 schools. Thereafter, an intact class of 50 students was randomly drawn from each of the 18 chosen schools at a second stage, which then gave a total of 900 (18x50) students that constituted the sample size for the study.

Using the SS2 Mathematics curriculum, a sixty (60) items Mathematics Achievement Test (MAT) designed by the researches in collaboration with some experienced Mathematics teachers at the senior secondary school level, was used for data collection. An initial pool of 100 items were raised and after pilot testing and item analyses for facility and discriminatory indices, pruned to 60, which is also in accordance with the National Examinations Council (NECO, 2022) standard stipulating a minimum of 40% : 60% ratio coverage for lower and higher order cognitive behavioural objectives. The arrangement of the items was varied according to formats for the two experimental groups (based on difficulty and item number variation), and control groups. Arrangement by difficulty is Bottom -Top (hard-to-easy) and were for experimental groups A while the item number variation type, based on varying item positions is for experimental group B. The random or traditional type based on no particular order is for group C-the control group.

For content validity, a test blue print was used to ensure the sampling adequacy of items and spread across the topics and cognitive behavioural objectives. Two (2) experienced Mathematics experts (teachers) matched the items with the topics and cognitive behavioural objectives, and any item not properly matched was dropped. Finally, a specialist in measurement and Evaluation and a test expert scrutinized the instrument for face validity and grammatical syntax before the final draft and arrangements.

The reliability of the instrument was established using Cohen's Kappa procedure and Kuder-Richardson formula 20 (KR_{20}). The items matched with the topics by the two Mathematics experts were analyzed using Cohen's Kappa inter-scorer reliability (K) and found to be 0.82 index. Furthermore, 20 copies of the instrument arranged in no particular order (traditionally or randomly) ware administered to students that are not part of the sample of this study. Their responses were subjected to KR_{20} analyses and a coefficient of 0.79 was attained. In both cases, the indices indicated reasonable reliability for the instrument to be used for the study.

The instrument was administered on the subjects by direct delivery method (DDM) Twelve (12) of the eighteen (18) schools constituted the experimental groups of A: Bottom-top (hard to easy) and B: Varied item number (mixed types) while six schools of group C is the control group (traditional type). Each of the groups was assigned six intact classes of 50, making 300 students per group. With the help of research assistants who were experience classroom teachers, cheating was checked and standard testing environment was maintained during administration and retrieval of instruments which was accomplished within the specified time limit stated in the instrument.

Mean score differences and standard deviation were used to answer the research questions while t-test statistic was used to test hypotheses at the 0.05 level of significance. A mean score difference greater than zero point one (X_i-X_j > 0.1) between a group and that of the control group was considered positive effect, equal to zero or less than 0.1, (0 =X_i - X_j, < 0.1) was considered no effect or negligible effect, while less than zero (X_i- X_j, < 0) was considered negative effect or impact. A null hypothesis was retained if the observed t-values was below its critical equivalent, and was not retained or dropped, if

otherwise. Mean differences used to compute t-values were each of the experimental groups versus the control group.

RESULTS AND DISCUSSION

Research Question 1: What effect does Bottom-top (hard to easy) arrangement of test item has on examinees' performance in mathematics?

Table 2: Descriptive Statistics on the Effect of Test items arranged in Bottom-Top (Hard to Easy) on Testees' Score in Mathematics.

Group	N	$\overline{X}$	SD	$\bar{X}_A - \bar{X}_C$	Remarks
Group A	300	27.06	7.15		
				-4.88	Negative effect
Group C	300	31.94	8.67		$(\bar{X}_A - \bar{X}_C < 0)$

Table 2 reveals that examinees in Group A who were tested with items arranged in bottom-top (hard-to-easy) format have a mean score of 27.06 while those (group C) tested with items arranged in random (or traditional) format have a mean score of 31.94. Their respective standard deviations were 7.15 and 8.67, while a mean difference of -4.88 was obtained of group A over group C. The mean difference is a glaring negative value, obviously indicating a negative effect. In other words, test items arranged in bottom-top (hard-to-easy) format have negative effect on testees' performance in Mathematics.

Research Question 2: What effect does test items arrangement based on variation in item number (VIN) in the same test for individual testee have on examinees' performance in mathematics?

Table 3: Descriptive Statistics on Effect of Test Items arranged based on Variation in Item Number (VIN) for individual Testees, on Examinees' performance in mathematics.

Group	N	$\overline{X}$	SD	$\bar{X}_B - \bar{X}_C$	Remarks
Group B (VIN)	300	37.44	8.94		
				5.50	Positive effect
Group C (Control)	300	31.94	8.67		$(\bar{X}_B - \bar{X}_C > 0)$

The control group (Group C) is compared with the experimental Group B in the above table. The experimental group, those tested with test items arranged in mixed format (based on Variation in Item Numbering in the same test for different examinee) have a mean score of 37.44 with a standard deviation of 8.94 while the Control Group (Group C) those who were tested with items arranged in the traditional or random order, have a mean score of 31.94 and a standard deviation of 8.67. A positive mean score difference of 5.50 between the two groups was observed in favour of Group B and by implication, it was concluded that test items arrangement based on Variation in Item Numbering in the same test for individual testees have positive effect on the performance of the examinees in Mathematics.

Hypothesis (Ho) 1: Bottom-top test items arrangement has no significant effect on examinees' performance in mathematics in senior secondary schools Rivers State.

Table 4: t-test Analyses on Effect of Test Item Arrangement Based on Bottom-Top (Hard-to-easy) on Testees performance in Mathematics

Group	N	$\overline{X}$	SD	t_{-cal}	df	α	Decision
Group A	300	27.06	7.15				

(Bottom-Top)								
				4.34	198	0.05	1.96	Reject Ho
Group C (Control)	300	31.94	8.67					

The t-test analysis from table 4 revealed an observed t-value of 4.34, and with degrees of freedom of 198, a critical value of 1.96 at the level of significance of 0.05. Since the computed value of t is higher than its critical equivalent (ie. $t_{cal} > t_{crit}$), at the given level of significance of 0.35, the null hypothesis that "Bottom-top test items arrangement has no significant effect on examinee performance in Mathematics" is rejected and the alternative, accepted. This suggests that the computed t-value is too significant to be attributed to sampling error. The implication is that test item arrangement based on bottom-top (hard-to-easy) format has significant effect on examinees performance in Mathematics.

Hypothesis (Ho) 2: There is no significant effect of mathematics test item arrangement based on variation in item numbering (VIN) in the same test for individual testee, on examinees performance.

Table 5: t-test analysis on the effect of test item arranged based on Item Number Variation in same test for individual testee on examinee performance

Group	N	$\overline{X}$	SD	t-cal	df	α	t-crit	Decision
Group B (VIN)	300	37.44	8.44					
				3.30	198	0.05	1.96	Reject Ho
Group C (Control)	300	31.94	8.67					

Table 5 above shows that a significant effect was found on testees performance in mathematics when test items were arranged based on item number variation. This arrangement is such that no adjacent Examinees have their item numbers the same, in other words, item number 1 for examinee A may be item number 25 for examinee B. Under such mixed format arrangement, the t-test analysis revealed a calculated t-value of 3.30 which is far and above the critical value of 1.96. With a degree of freedom of 198 at a level of significant of 0.05, the computed value of t was significant and could be due to chance. Consequently, the null hypotheses of no significant effect of test item arrangement based on variation of test item number on examinee performance was not retained. The implication of this is that test item arrangement based on item number variation or mixed number format has significant influence on examinees' performance in mathematics than when items are arranged in the traditional or random format.

Summary of Findings

The findings in this study can be summarized into the following:

1. Test item arrangement based on bottom-top format (descending order or hard-to-easy) has a significantly negative effect on examinees performance in Mathematics
2. Test items arrangement based on Variation in Item Numbering (VIN) in the same test for individual testees has significant positive influence on examinee performances in Mathematics

Discussion of Findings

A major result of the study was that test items arrangement based on bottom-top (descending order of difficulty or hard-to-easy) format has a significantly negative effect on examinees performances in mathematics. A negative mean difference of -4.88 was found between group A, tested with items arranged in bottom-top format and group C, tested with the traditional format (items arranged haphazardly). The negative mean difference was an obvious indication that group A underperformed compared to those in group C; suggesting that items arrangement based on bottom-top (hard-to-easy) format have but negative impact on examinees performance. This is contrary to the claim of Abbasian

and Zadar (2019) that test item arrangement does not have any significant effect on the performance of test takers but only on test usefulness criteria with varying extent.

Specifically, on investigating the relationship between test item arrangement and performance and test usefulness criteria, Abbasian *et al.,* reported that test validity, reliability, interactiveness, authenticity and practicality were subject to test item arrangement with varying extent but the performance of the testees were not affected by the arrangement of items. Although the current study used quasi-experimental design while Abbasian *et al.,* used Ex-post facto, and the subjects used for investigation were different (mathematics in the current and English studies in the former), the item arrangement format were similar as both examined order of difficulty Moreover, similar statistical procedures (means, standard deviation and mean differences) were used in both studies. The difference in findings may be due to the different subjects used and the designs adopted. Another contrary finding to the current study was issued by Barbara (2015) who examined item arrangement and knowledge of arrangement on test scores. He used the three items ordering (easy-to-hard, spiral cyclical and random) and two levels of knowledge of ordering (informed and non-informed) to study differences in test results, student's perception of the test fairness and difficulty, and students' estimation of test performance. With quasi-experimental design and various statistical tools including mean differences, using science subjects, Barbara (2015) findings revealed no significant order effect and no significant effects were found for knowledge of the ordering or for the interaction of knowledge of arrangement and order, which is contrary to the current findings that arrangement of item with particular reference to bottom-top or hard-to-easy has significant negativity effect on examinee test scores.

Nevertheless, results similar to the findings of the current study were obtained by Tei-Firstman (2011) and Ollennu and Etsey (2015). Ollennu and Etsey studied the impact of item position in multiple choice tests on student's performance at the basic education certificate examination (BECE) in Ghana, adopting Quasi-experimental design and using a sample of 810 junior secondary school (JSS) from different schools. Items in English, mathematics and sciences arranged using difficulty order were used for data collection while analysis of variance (ANOVA) was the statistics adopted. Findings from Ollennu et al. (2015) study revealed that for the three subjects at the BECE level, when items order was altered, the difference in performance was statistically significant which agrees with the findings of the current study. In other words, the sharp deflection of performance following the alteration of items arrangement suggests a significant effect of items arrangement on test performance of examinees. While the current study used only mathematics at the senior secondary, the former used English, Mathematics and Science in the junior, ruling out possible influence of level of class and subjects. Moreover, the complex analyses procedure adopted in the former study and yet result are in line with the current study, further strengthens the authenticity of the findings of the current study. The variation in the approach of the two studies ruled out possible influence of extraneous variable, while the similar strategies sustained the focus on what should be the outcome.

The study further revealed that Test items arrangement based on item number variation in the same test for individual testees has significant positive influence on examinee performances in Mathematics. Item number variation for individual testee is an arrangement format whereby no adjacent testees have the same or similar items numbering, in other words, item number 1 for examinee "A", may be item number 20 for student "B" and 48 for "C", etc. Under this arrangement for the current study, the experimental group had a mean score of 37.44 with a standard deviation of 8.94 while the Control Group had a mean score of 31.94 and a standard deviation of 8.67. A positive mean score difference of 5.50 between the two groups was observed in favour of the experimental group and by implication, it was concluded that test items arrangement based on variation in item numbering in the same test for individual testees have positive effect on the performance of the examinees in Mathematics. This is contrary to the results of Uwah (2016) who obtained a significant but negative effect on testees performance when examined with varied item number format, and conjectured that testees may have not been independent of themselves during testing, a situation under this arrangement that helps to check examination malpractices. For the current study, this format is most applauded since it checks examination malpractices and yet examinees performed better compared to the traditional format. Ukwuije (2012) subscribed to this arrangement as he re-echoed its merit of checking students copying from one another, even though he did not comment on whether it enhances performance or not. This arrangement of item number variation is worth embracing because other scholars and researchers like

Mehren and Lehman (1975) and Tei-Firstman (2011) had advocated for this format, advancing that it also reduces negative sequence effect in the individual testee.

CONCLUSION

Arrangement of test items is a very sensitive and key aspect of the testing process. Whatever the format, it has been established that it has effect on the performance of the examinee. In the two arrangement formats considered in this work, item number variation for individual examinee is most preferred to the Bottom-top and traditional type arrangements for its positive influence on performance and its ability to inhibit examination sharp practices.

RECOMMENDATIONS

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

Item number variation for individual examinee and the Top-bottom arrangement should be used while compiling test items. Items should be arranged beginning with the simple ones through the fairly difficult to the more difficult ones, bearing in mind that no adjacent examinees have the same item numbering.

REFERENCES

Anastasi, A. & Urbina, S. (2006). *Psychological testing (7th Edition).* India: Dorling Kindersley.

Anyarathne, A. H. S. S. S. (2013). Factors affecting students test scores. *ST306-Data Analysis and Report Printing.* Department of Statistics and Computer Science. Faculty of Science, University of Paradeniya

Asuru, V. A. (2015). *Measurement and evaluation in education and psychology (2nd Edition).* Port Harcourt: Pearl Publishers International Ltd.

Barbara, (2015). Item arrangement and knowledge of arrangement of test scores. *In Journal of Experimental Education,* 49(1). https://dei.org/10.1080/00220973. Retrieved Jan. 2, 2020.

Best, J. W. & Khan, J. V. (2010). *Research of Education* (2nd ed.). India: Dorling Kindersley.

Essien, I. T. (2001). *Fundamental of Educational Measurement and Evaluation*. Uyo: Afahaide & bros Printing & Publishing Company.

Gholam-Reza, A. & Hanieh, Z. (2019). Relationship between Test Item Arrangements and Testee's Performance and Test Usefulness Criteria. *International Journal of English language Teaching.* Sciedupress.com, https://doi.org/10.5430/ijelt.v6n2p52 Retrieved September. 28, 2024

Hopkins, C. D. & Antes, R. L. (1978). *Classroom measurement and evaluation. Itasca, IL: F.E. Peacock Publishers Inc.* https://www.jstor.org. Retrieved Feb. 23, 2020

Kathryn, (2021). Teachers' regular supervision and students' academic performance in public secondary schools in Ekiti State, Nigeria. *International Research in Education. 9(2), 54*

Koul, L. (2018). *Methodology of educational research (4th ed.).* India: Vikas Publishing House PVT Ltd.

Kpolovie, P. J. (2012). *Advance Research Methods.* Port Harcourt: Springfield Publishers.

Obilor, E. I & Miwari, G. U. (2023). Effect of Mathematics test items arranged in Top-bottom and by Order of Topics, on the performance of senior secondary school students in Rivers-South senatorial District. *International Journal of Research Publication and Reviews, 4(110), 1149-1154*

Ollennu, S. N. & Etsey, Y. K. A. (2015). The impact of test item position on multiple-choice test on students' performance at the basic education certificate examination (BECE) level in Ghana. http://www.edu.ghana.WAEC/outline.org. Retrieved, Sept. 18, 2019

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

Orluwene, G.W. (2012). *Introduction to test theory and development process.* Port Harcourt: Chris-Ron Integrated Services.

Steinberg, L. (2010). A Dual Systems Model of Adolescent Risk-Taking. *Developmental Psychology 52(3), 216-224*

Toube, A. C. (2009). Assessing students critical thinking ability. *Journal of educational evaluation.* 4(16), 322-328

Tei-Firstman, R. I. (2011). *Test Item Arrangement on Students Test Scores*. Unpublished Thesis, University of Port Harcourt.

Ubulom, W. J., Uzoechi, K. C., Amini. C. M & Vipene, J. B. (2011). *Fundamentals of Measurement and Evaluation.* Port Harcourt: Celwils Publishers.

Ukwuije, R. P. I. & Orluwene, G, W. (2012). *Peanuts Educational Statistics.* Port Harcourt: Celwils Publishers.

Uwah, I. V. (2016). *Effect of test items arrangement on performance in mathematics among junior secondary school students in Obio/Akpor local government area of Rivers state.* Unpublished M. Ed dissertation, University of Port Harcourt, Nigeria.