UNIVERSITY OF CAPE COAST

EFFECTS OF ITEM SEQUENCING IN MULTIPLE-CHOICE TESTS ON SENIOR HIGH SCHOOL STUDENTS' ACADEMIC PERFORMANCE IN THE KUMASI METROPOLIS: THE MODERATING ROLE OF GENDER

JOHN BAFFOE

2021

UNIVERSITY OF CAPE COAST

EFFECTS OF ITEM SEQUENCING IN MULTIPLE-CHOICE TESTS ON SENIOR HIGH SCHOOL STUDENTS' ACADEMIC PERFORMANCE IN THE KUMASI METROPOLIS: THE MODERATING ROLE OF GENDER

BY

JOHN BAFFOE

Thesis submitted to the Department of Education and Psychology of the Faculty of Educational Foundations, College of Education Studies, University of Cape Coast, in partial fulfilment of the requirement for the award of Master of Master of Philosophy Degree in Measurement and Evaluation

JUNE 2021

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the results of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's signature…… Date ……

Name: ……

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature …… Date ……

Name: ……

Co-supervisor's Signature …….. Date …….

Name:……

ABSTRACT

The study investigated the moderating role of gender in the effect of item sequencing in multiple-choice tests on Senior High School students' academic performance in the Kumasi Metropolis. The study used the quasi-experimental design. Stratified, simple random, proportionate and purposive sampling techniques were used to select a sample of 714 form 2 Senior High School students. Multiple-choice items were adapted from West African Examination Council (WAEC) and the validity of the instruments was ascertained through expert judgement. By using KR20, the reliability of the test items where 0.81 and 0.74 for the Random sequence for both Core Mathematics and Integrated science respectively. The reliability of the Hard-To-Easy (HTE) sequence were also 0.76 and 0.71 for Core Mathematics and Integrated Science respectively whiles that of Easy-To-Hard (ETH) sequence were 0.78 and 0.73 for Core Mathematics and Integrated Science respectively. One-way repeated measures ANOVA, One-Way MANOVA and the process procedure by Hayes (2018) were used to analyse the data that were gathered. The findings of the study indicated that item sequence in multiple-choice test has effect on students' academic performance in Core Mathematics and Integrated Science with Easy-To-Hard (ETH) sequence having the greatest effect on students' academic performance in both subjects. The study also revealed that students' academic performance in the various kinds of item sequence did not differ with respect to gender in Core Mathematics but differed in Integrated Science where females performed better than males. The study further showed that, in Core Mathematics, gender moderates the relationship between Hard-To-Easy (HTE) sequence and academic performance but not Easy-To-Hard (ETH) and Random (RDM) sequence. The results further indicated that, in Integrated Science, gender moderates the relationship between Easy-To-Hard (ETH) sequence and performance but not in Hard-To-Easy (HTE) and Random (RDM) sequence. It was recommended that testing organisations such as WAEC and classroom teachers should use Easy-To-Hard (ETH) sequence in order to increase students' academic performance in Core Mathematics and Integrated Science.

KEYWORDS

Academic performance

Gender

Item sequence

Hard-To-Easy (HTE)

Easy-To-Hard (ETH)

Random (RDM)

ACKNOWLEDGEMENTS

First and foremost, my sincerest gratitude goes to the almighty God through whose grace and mercy I obtained protection and sustenance to complete this study.

I wish to acknowledge the endurance, commitment and constructive submissions of my supervisors, Dr. Kenneth Asamoah-Gyimah and Dr. Eric Anane who devoted their precious time out of their busy schedules providing academic support, guidance and mentorship throughout the conduct of this study.

Again, I am thankful for the indefatigable support of Mr. Daniel Asamoah and Miss. Ruth Anan-Brew for their guidance in making this thesis a success. I wish to acknowledge the invaluable support of all the research assistants who availed themselves as field assistants and facilitated the data collection process.

I wish to also thank Samuel and Emelia Brew-Butler SGS/GRASSAG of UCC for their financial support in the form of research grant which made the conduct of thesis a success.

Special thanks also go to lecturers in the Faculty of Educational Foundations, authorities of schools used for the study and the Ashanti Regional Education Directorate for the support they offered me in the conduct of the study.

DEDICATION

To Mr. John Bernard Baffoe, Mr. Peter Baffoe, Mad. Adwoa Akyere and Florence Akua Baffoe

TABLE OF CONTENTS

LIST OFTABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Background to the Study

In the educational system, it is impossible for anybody to study without being assessed. This shows how critical assessment can be in our educational institutions in Ghana. Through assessment procedures, information is collected about an individual to make decision about how well he or she has achieved the learning targets. According to Nitko (as cited in American Federation of Techers, National Council on Measurement in Education and National Education Association, 1990), assessment is the process of obtaining information that is used for making decisions about students, curricula and programmes and educational policies. The information obtained about students is used to make decisions such as classroom instruction, placing them into different types of educational programmes, assigning grades to them, guiding and counselling them and selecting them for educational opportunities. This means that, the assessment procedures should focus on the information one needs to make educational decision. It is an indisputable fact that educators as well as assessment agencies are able to obtain information about learners through the use of tests because, they supply an important information that would be difficult, if not impossible to obtain from any other means (Nunally, 1978).

Test is explained as any task or series of tasks that is used to quantify specific traits or characteristics in people (Dzakadzie, 2017). In the educational system in Ghana, tests include paper and pencil instruments which help the tester to obtain an estimate of the specific traits being measured. The estimate

of trait of a person is done through observing and describing one or more characteristics using either a numerical scale or a classification scheme (Nitko, 2004). Educators and assessment agencies usually score tests by putting together the "points" a student earns on each question and by so doing characteristics of students are described using a numerical scale. Test results can serve many purposes to teachers and assessment agencies. This is supported. by the assertion of Nitko (2004) that tests serve many useful purposes in the educational systems, in industries and in the world of work. He further listed the purposes of test to include; instructional management decisions, selection decisions, placement decisions, classification decisions, counselling and guidance decisions and credentialing and certification decisions. Amedahe and Asamoah-Gyimah (2016) also highlighted on the uses of various types of test and reported that tests are used for (a) assessment and decision about students (b) planning and organization of instruction (c) instructional management decision (d) motivating students (e) grading students (f) guiding students (g) making selection and placement decisions (h) classification decisions (i) certification decision. Among the purposes of test, certification decision is the one that standardised tests like Basic Education Certificate Examination (BECE) and West African Senior Secondary Certificate Examination (WASSCE) seeks to offer.

Test comes in various forms. Ghanaian teachers as well as assessment agencies construct and use a number of test types. The test item could either be essay or objective types. Stalnaker (as cited in Reiner, Bothell, Sudweeks & Wood, 2002) defined essay test as a test item in which an examinee is required to composed a response made up of one or more sentences where no single

correct responses of pattern can be listed as correct and can its accuracy and quality can be judged subjectively. Another type of test that is usually used by both Ghanaian teachers and assessment agencies because of its objectivity in scoring is the objective type test. Nitko and Brookhart (2007) posited that a test is said to be objective when the procedures by which the correctness of the responses to the item will not in any way differ from one rater to the other. Objective tests are normally made up of a large number of items and the responses are scored without subjectivity, to the extent that experts can agree on the correct responses (Etsey, 2003). Objective test is of two types; thus, the selection type and the supply type. The selection type are the ones that a student selects the correct or the best answer from among a number of options presented to him/her. These consist of multiple-choice type, true or false type and matching type. The supply type has variations such as completion (fill in-the blanks) and short answer.

Among the selection types of objective test, one of the types that is most frequently used in the Ghanaian school systems and external examinations including West African Senior Secondary Certificate Examination (WASSCE) is the Multiple-Choice Test (MCT). The West African Examination Council (WAEC) usually uses multiple-choice tests combine it with the essay test to assess the mastery level of the various topics taught at the Senior High Schools from SHS 1 to 3. However, concerns have been raised by other researchers with regards to the effect that item sequence in MCT has on students' academic performance. For instance, Shepard (1994) asserts that sequence of items in MCT can make a large difference in students' academic performance. This means that for instance, if items are sequenced in Easy-To-Hard (ETH) format,

students are likely to be motivated to answer the items which might improve their performance on the test. In the same way items sequenced in Hard-To-Easy (HTE) is likely to discourage students to progressively answer the items which might also affect students' performance on the test.

It has also been argued in literature that, the effect of item sequence in MCT is influenced by the gender of students. This is supported by Research Division of WAEC, Lagos (1993) who opined that how items are sequenced in MCT as well as students' gender can affect students' academic performance. This means that whiles the sequence of items can have effect on students' academic performance, their gender can also predict their performance on the various item sequence. This was further confirmed by Cobbinah (2016) that students' academic performance in the various item sequence significantly differ with respect to gender.

Item sequencing in MCTs according to Opara and Uwah (2017) involves the various formats through which tests come. Again, it refers to the way classroom teachers and professional test constructors and examination bodies arrange their test items. He further noted that test item sequencing based on ascending order of difficulty implies arranging test items from simple to complex which is also called Easy-To-Hard (ETH). Here, students are presented with simple questions first and progressively, less simple (difficult) item are presented. As Tei-Firstman (2011) noted, presenting students with simple question first will help sustain their interest and zeal towards approaching more questions. Another kind of item sequence is the arrangement of item based on descending order of difficulty (i.e. arranging test items from complex to simple) which is also called Hard-To-Easy (HTE). According to Ollennu (2011), this

form of arrangement applies when students are presented with more difficult items first then followed by moderately difficult then to more simple items.

Again, there is also a random arrangement (RDM) of items where items are arranged from neither Easy-To-Hard not Hard-To-Easy. Ollennu (2011) further posited that, here, items are arranged randomly in order of the syllabus or topic but not based on any difficulty level. Item sequencing is seen to provide cues which may help with memory recall which is consistent with encoding specificity or how information is connected within the brain. This further means that item sequence provides signals that aids in memory recall of information which is associated with the brain. Considering the appropriate way of sequencing items. There is no unified position since many positions have been outlined by researchers. While others are of the view that the ETH is the best (Balch, 1989; Carlson & Ostrosky, 1992; Stout & Wygal, 1992; Jordan 1953; Leary &Doran, 1985), others are also of the view that the HTE is appropriate (Skinner, 1999; Munz & Jacobs, 1971). Meanwhile, Alakayleh (2017) also believe that the random arrangement is appropriate for bringing out the true ability of students. Looking at the predictive ability of item sequence, and the relevance of item sequence in test construction, the study sought to investigate if item sequence in multiple choice test has effect on Senior High School students' academic performance, and further investigate the moderating role of gender in the relationship between item sequence and Senior High School students' academic performance in Integrated Science and Core Mathematics in Kumasi Metropolis.

Statement of the Problem

In Ghana, it is a mandate to pass the four core subjects namely Mathematics, Integrated Science, English Language and Social Studies in external examination at both the basic and senior high levels. This means that every student who goes through second cycle institutions is required to pass these subjects in addition to three other elective subjects before a student can be guaranteed entry into Ghanaian university. Aside Mathematics and Integrated Science being recognized as core subjects in our educational system, they are also seen as an important feature of our education which occupies a high place in our schemes as a nation. The role of mathematics and science education in economic development in the 21st century is not in dispute since the two subjects form the backbone to technological advancements and modern strides. In recent past, several efforts geared towards the advancement of the field of Mathematics and Science have resulted in expansions such as funds. Again, over the years, government and state agencies responsible for education and training, have carried out several activities including equipping teachers with resources and workshops, and packages such as Ghana Partnership for Education Grant, Science Resource Center Project and Ghana Skills and Technology Development Projects aimed at promoting mathematics and science education in general and its teaching and learning.

In spite of these efforts made by the government in the study of the core subjects, especially in Mathematics and Science, students still perform abysmal in their external examination (WASSCE) and the problem of poor performance of students turns into the subjects of discussion whenever results of such examinations are out (Adusei, 2017).

For instance, in Ghana, in the past years, results in May/June WASSCE examinations according to Educational Management Information System (EMIS) (2019) revealed that students' academic performance especially in the core subjects has not been encouraging. According to WAEC (2019), in 2016 May/June results, 247,262 students sat for the WASSCE examination and for the core subjects, out of that, 125,065 students representing 53.10% passed English Language, 77,108 students representing 32.83% passed Mathematics, 113,933 students representing 48.48% passed Integrated Science and lastly, 135,821 students representing 54.93% passed Social Studies. Again, WAEC stated that the 2016 result was not significantly different from 2017 May/June examination results since there was a slight improvement in the performance of English Language from 53.19% to 54.06% although the total number of students who sat for the 2017 May/June examination was 289,210. WAEC (2019) further noted that, Mathematics also showed a markable improvement from 32.83% to 42.73% whiles there was a drop in performance in Integrated Science and Social Studies from 48.48% to 43.66% and 54.93% to 52.25% respectively. Furthermore, the result of May/June 2017 was not far different from the results of 2018 especially with respect to Mathematics, English Language and Integrated Science. There was a marked improvement in the performance of Social Studies over that of 2017 (52.25% to 73.27%), whiles Integrated Science recorded a slight improvement (43.66% to 50.52%). On the contrary, Core Mathematics showed a drop in performance from 2017 (42.73% to 38.33%). Similarly, there was a drop in performance of English Language (54.06% to 46.79%). According to WAEC (2019), although, there was a significant improvement in the performance of students in Mathematics

(38.33% to 65.31%) and Integrated Science (50.52% to 63.17%) in the 2019 WASSCE results, it can be seen from the statistics of performance in the core subjects discussed above that, Core Mathematics and Integrated Science have over the years recorded the worse performance than any other core subjects.

On the whole, the pass rate of Core Mathematics stood at 32.83%, 42.73%, 38.33% and 65.31% for the year 2016, 2017, 2018 and 2019 respectively with an average percentage mean score of 44.80% indicating that 55.20% students still failed in Core Mathematics. Similarly, Integrated Science also recorded a pass rate of 48.48%, 43.66%, 50.52% and 63.17% with an average percentage mean score of 51.46 suggesting that 48.54% failed the Integrated Science. This situation seems not to be different from the performance of students in Core Mathematics and Integrated Science in the Kumasi Metropolis and this has raised a lot of questions among stakeholders of education about the abysmal performance of students (Adusei, 2017). The West African Examination Council's statistics of Senior High School students' academic performance in the core subjects in the Kumasi Metropolis from 2015-2019 is seen in the Table 1.

Table 1- Statistics of Senior High School Students' Academic Performance in the Core Subjects in 2015-2019 in the Kumasi Metropolis in WASSCE

Academic year	Core subjects	Total candidates	Pass rate (A1-C6)
2015/2016	Core Mathematics	22,077	10,516 (47.6%)
	English Language	22,077	15,115 (68.5%)
	Social Studies	22,077	15,457 (70.0%)
	Integrated Science	22,077	14,530 (65.8%)
2016/2017	Core Mathematics	21,812	11,258 (51.6%)
	English Language	21,812	15,116 (69.3%)
	Social Studies	21,812	12,666 (58.1%)
	Integrated Science	21,812	12,556 (57.6%)
2017/2018	Core Mathematics	23,233	10,971 (47.2%)
	English Language	23,233	14,366 (61.8%)
	Social Studies	23,233	18,543 (79.8%)
	Integrated Science	23,233	13,487 (58.1%)
2018/2019	Core Mathematics	23,880	11,701 (49.0%)
	English Language	23,880	16,524 (66.2%)
	Social Studies	23,880	18,101 (75.8%)
	Integrated Science	23,880	12,226 (51.2%)

Source: EMIS (2016-2019); Ashanti Regional Educational Office (2019)

It can be observed from Table 1 that, Core Mathematics and Integrated Science have recorded worse performance than any other core subjects for the year 2016-2019 among Senior High Schools in the Kumasi Metropolis and this low performance has been attributed to a lot of factors. Opara and Uwah (2017) point that individuals' academic performance is relative to a number of factors which could be the subject area itself, family background, economic status and many other factors including biological phenomenon. Other factors according to Asamoah (2018) results from the home, school, students and teachers.

However, very limited attention has been given to the sequence of items in Multiple-Choice Test (MCT) in WASSCE regardless of its' potency of predicting students' academic performance because MCT's form 40% of the total marks in examinations including WASSCE which forms an important portion of determining performance. This further means that, getting 40% of the total marks places one closer to the pass mark. The position of Shepard (1994) was confirmed by Cobbinah (2016) who reported that the sequence or order of test items in MCT has effect on students' academic performance in Core Mathematics in WASSCE. Cobbinah's (2016) findings was consistent with the study of Opara and Uwah (2017) who reported that the Easy-To-Hard and Random sequence had positive and significant effect on performance while the Hard-To-Easy had positive but insignificant effect on students' academic performance. In Greater Accra, Ollennu and Etsey (2015) reported that the position or sequence of an item in MCT in Mathematics, Integrated Science and English Language at the BECE level had impact on students' academic performance. This was further confirmed that item sequencing in MCT has significant effect on students; academic performance (Alakayleh, 2017; Chen 2012; Soureshjani, 2011 & Gray, 2003).

In contrast, Abbasian and Zadsar (2019) found that the sequence of items in a MCT has no significant effect on students' academic performance. The position of Abbasian and Zadsar (2019) was earlier supported by Vandee Schee (2009) who found that the sequence of items has no significant effect on students' academic performance due to lack of time for examinees to complete the test. Furthermore, Togo (2002), Geiger and Simons (1994), Coniam (1993), Chidomere (1989), Leary and Doran (1985) and Gerow 1980 also assert that,

the sequence in which are arranged in MCT has no effect on students' academic performance.

It can be inferred from the aforementioned literature that several studies have been conducted on the effect of item sequencing on students' academic performance with respect to subject areas such as Mathematics (Ollennu & Etsey, 2015; Cobbinah, 2016; Opara & Uwah, 2017 & Alakayleh, 2017), English Language (Coniam, 1993; Soureshjani, 2011; Ollennu & Etsey, 2015; Alakayleh, 2017) Integrated Science (Ollennu & Etsey, 2015 & Alakayleh, 2017), Social Studies, Biology, Economics (Research Division of WAEC, 1993), Physics (Gray, 2003), Marketing (Vandee Schee, 2009), Accounting (Togo, 2002), amongst others. However, the findings of such studies appear to be inconsistent and contradictory. While others posit that item sequence has a significant effect on students' academic performance (Gray, 2003; Soureshjani, 2011; Chem 2012; Ollennu & Etsey, 2015; Cobbinah, 2016; Opara & Uwah, 2017; Alakayleh, 2017), others are of the view that item sequence does not have any significant effect on students' academic performance (Gerow, 1980; Leary & Doran, 1985; Geiger & Chidomere, 1989; Coniam, 1993; Simons, 1994; Togo, 2002; Vandee Schee, 2009; Abbasian & Zadsar, 2019). This creates a research lacuna in literature since there are no certain conclusions as to whether or not item sequencing in MCT has an effect on students' academic performance. With reference to the nature of item sequencing, it has been argued by Cobbinah (2016) and Research Division of WAEC, Lagos (1993) that gender can cause a difference in the academic performance with respect to a particular item sequence. Again, Pettijohn and Sacco (2007) also report that gender can also serve as a moderator in the relationship between item sequence

and students' performance. However, very limited information can be found on the moderating role of gender especially in Ghana, which by a recommendation of Pettijohn and Sacco (2007), there is need to uncover the moderating role of such variable.

Referring to the studies conducted on the effect of item sequencing on students' academic performance, it can be inferred that the studies conducted were done in the western world and very few of the studies on the aforementioned variable focused on Ghana (Cobbinah, 2016; Ollennu & Etsey, 2015). The few studies conducted in Ghana also concentrated on other study areas to the neglect of the Kumasi Metropolis. Making inferences from the background and particularly the low performance of students in Core Mathematics and Integrated Science, there is no evidence as to whether or not the poor performance of students in WASSCE in these two subjects can be attributed to the way the items were sequenced, or whether item sequencing can affect students' academic performance. Again, there is also no readily available information on the moderating role of gender regarding the effect of item sequencing on students' academic performance which has been advocated by Pettijohn and Sacco (2007).

Based on the aforementioned research gaps and with the evidence of low performance in Core Mathematics and Integrated Science in the Kumasi Metropolis, a study of this nature is important as it focused on investigating the effect of item sequencing in MCT on students' academic performance in Core Mathematics and Integrated Science and further determine the moderating role of gender in the Metropolis, it is against this background that the current study sought to investigate the effect of item sequencing in multiple-choice test on

students' academic performance in the Kumasi Metropolis, and further investigate the moderating role of gender in the relationship between item sequencing and students' academic performance.

Purpose of the Study

The main purpose of the study was to investigate the effect of item sequencing in multiple-choice test on students' academic performance and also the moderating role of gender in the relationship between item sequencing and students' academic performance in the Kumasi Metropolis. Specifically, the study sought to investigate whether:

1. the sequence of item in multiple-choice test has effect on students' academic performance in:
    i. Core Mathematics
    ii. Integrated Science
2. a difference exists in students' academic performance in item sequencing in Core Mathematics and Integrated Science with respect to gender.
3. a gender moderates the relationship between item sequencing and students' academic performance in:
    i. Core Mathematics
    ii. Integrated Science

Research Question

What is the effect of sequence of items in multiple-choice test on students' academic performance in:

i. Core Mathematics?
ii. Integrated Science?

Hypothesis

H0: There is no statistically significant difference in students' academic performance in item sequencing in Core Mathematics and Integrated Science with respect to gender.

H1: There is statistically significant difference in students' academic performance in item sequencing in Core Mathematics and Integrated Science with respect to gender.

H0: Gender does not significantly moderate the relationship between item sequencing in Multiple-Choice Test (MCT) and students' academic performance in Core Mathematics and Integrated Science.

H1: Gender significantly moderate the relationship between item sequencing in Multiple-Choice Test (MCT) and students' academic performance in Core Mathematics and Integrated Science.

Assumptions

1. To control the effect of school category having influence in the experiment, only category (A) schools were selected for the study. Category (A) schools stipulated by Ghana Education Service (GES) were selected because it can be assumed that these schools have same or similar characteristics such as facilities, qualification of teachers, WASSCE performance and amongst others.
2. Different groups of students from the sample were used during the administration of the various forms of the test. It was assumed that these groups have the similar characteristics and therefore the outcome of the study was not affected.

3. The test items were adapted from WAEC past questions and therefore it was assumed that the items are of good quality which can be used to get the intended responses from testees.
4. The study assumed that test items are administered in random sequence to determine academic performance (Taub & Bell, 1975)

Significance of the Study

The findings of the study will help to identify or know if the sequence of items (whether from Easy-To-Hard, Hard-To-Easy or Random) has any effect on students' academic performance in Core Mathematics and Integrated Science. This will inform stakeholders in education such as classroom teachers to know when to use and when not to use item sequence when assessing students. In addition, it will help them to know the kind of item sequence to use and not to use when assessing students. Again, the findings of the study will reveal whether or not gender can affect the relationship between students' academic performance and item sequencing. This will help WAEC, classroom teachers and other assessment agencies to construct items that seeks to control the influence of a test taker being a male or a female. The findings of the study would also add up to the limited body of knowledge regarding the effect of item sequencing in multiple-choice test on students' academic performance as well as the moderating role of gender.

Delimitation of the Study

The study focused on variables that relates to the effect of item sequencing in multiple-choice test on students' academic performance and also investigated the moderating role of gender. The scope of the study covered Senior High School Form 2 students in the Kumasi Metropolis but not other

Municipalities or other grade levels. In furtherance, the scope of the study was limited to public Senior High School students but not private schools. The study was also limited to only Core Mathematics and Integrated Science but not other subjects.

Limitations of the Study

One of the major limitations of the study was the use of quasi-experimental design. The design has an inherent limitation because of the lack of random assignment which limited the internal validity. Even though the study used students in the same ability group, random assignment of students into experimental and control group would have increased the external validity of the findings. Another major limitation was the administration of the test items to the students in the experimental group on three different occasions. At a point in time, the students' felt tired and discouraged and that might have affected the reliability and validity of the use of the test results. Lastly, another major limitation was the unenthusiastic attitude of teachers towards the research work especially in the administration of the test items. Some of the teachers were not ready to spare their lesson periods for the administration of the test items. Even though the above were some of the limitations of the study, measures were put in place to control it. In dealing with the first limitation, the authorities in the schools used did not allow the already grouped or intact classes to be disrupted by assigning them into experimental and control group therefore the quasi-experimental design which does not sue randomization was appropriate. More so, the second and third limitations was controlled after explaining the purpose of the study comprehensively to them which encouraged their participation.

Definition of Key Terms

Item Sequence: This refers to the way classroom teachers and professional test constructors and examination bodies arrange their test items

Hard-To-Easy (HTE) Sequence: This is form of sequence where items are arranged from complex or hard to simple.

Easy-To-Hard (ETH) Sequence: This is form of sequence where items are arranged from simple to complex or hard.

Random Sequence: This is form of sequence where items are arranged randomly based on the syllabus or topics.

Academic Performance: This refers to the measurement of students' achievement across various academic subjects.

Organisation of the Study

"The study is organised into five chapters. Chapter one consists of the background to the study, statement of the problem, purpose of the study as well as research questions. It also presented the significance of the study, delimitation and limitations of the study. The second chapter focused on the review of existing literature in relation item sequencing in Multiple-Choice Tests and students' academic performance. The methodology which is captured in the third chapter also deals with the description of the research design, population, sample and sampling procedures and data collection instruments". "The chapter also capture validity and reliability of the instruments, ethical considerations as well as data processing and analysis. Chapter four presents the results and discussion of the study whereas the fifth chapter focuses on the summary, conclusions and recommendations from the findings. The chapter also includes suggestions for further studies".

CHAPTER TWO

LITERATURE REVIEW

Introduction

The study sought to investigate the effect of item sequencing in Multiple-Choice Test (MCT) on Senior High School students' academic performance in the Kumasi Metropolis and further investigated the role of gender in the relationship between item sequencing and students' academic performance. The chapter therefore reviewed relevant literature on the topic indicated. The literature review of the study presented the findings, assertions and observations of several writers or authors on the effect of item sequencing in Multiple-Choice Test (MCT) on Senior High School students' academic performance. Precisely, the review covers the theoretical review, conceptual review which comprised; the concept of assessment, the concept of test, item formats, the Multiple-Choice Test (MCT), conditions for using Multiple-Choice Test (MCT), item difficulty of Multiple-Choice Test (MCT), the concept of item sequencing in Multiple-Choice Tests, the concept of academic performance in Core Mathematics and Integrated Science. In addition, conceptual framework and empirical studies on the topic were also indicated and reviewed respectively.

Theoretical Review

Validity of assessment Results

Validity according to Nitko and Brookhart (2007), "is the soundness of one's interpretations and the use of students' assessment results'' (p. 38). Again, according to America Education Research Association, America Psychology Association and National Council on Measurement in Education (1999) as cited

in Amedahe & Asamoah-Gyimah (2003), validity refers to the degree to which evidence and theory support the interpretations of test scores. Considering the above definition, validity can be explained as the extent to which assessment results are meaningful and appropriate or suitable interpretations and uses are assigned to students' assessment results. In furtherance, for a student's result to be valid several factors surrounding the test should be taken into consideration not the test instrument in question. For a test considered to have high validity, the items in the test should closely related to the test's intended focus. For many certification and licensure tests, this means that the test items will be closely related to the specific job or occupation. However, if a test is considered to have poor or low validity, then the test in question does not measure the job-related content or competencies it ought to.

Linking this to item sequencing in a test, if items in a test are sequenced in a way to have effect on students' performance, the test eventually measures what it ought not to measure. When this becomes the case, there is no justification for using the test results for the purpose in which it was constructed for. Several pieces of evidence can be used to support the validity of a test including content validity evidence, criterion validity evidence and construct validity evidence.

Content Validity-Evidence of Achievement Tests

The main purpose of content validation is to assess whether the item adequately represent a performance domain of a psychological construct of specific interest. This is because the validity of a classroom assessment results largely depends on how well an assessment samples the learning targets (Nitko & Brookhart, 2007). In content validation, the items on the test should

adequately sample the domain of concern or interest. This can be achieved by clearly identifying the important learning targets and ensuring that they are well sampled by the assessment procedure. According to Nitko and Brookhart, the tasks included in an assessment should reflect the important content and learning outcomes specified in the schools' and state's curriculum framework. They further indicated that the content included in an assessment should be of great value or significance to students' further learning or life skills.

Airasian (2005) stated that to ensure content validity, test items should match with course objectives, instruction and reflect adequate sampling of instructional materials. The items must be fairly distributed on the content of the curriculum and what a teacher teaches in class. The test constructor must be certain that the assessed content relates directly to important learning targets. To determine the content validity of test, the items are, therefore, compared with the table of test specification. Test experts or judges may also be asked to assess the extent of content validity in a particular test. According to Amedahe and Asamoah-Gyimah (2016), one must first define content domain and/or universe of situation in assessing the content validity of a particular test. They further explained that for the above to be achieved the test constructor should consider the subject matter-content and the type of learning target desired from students. This again indicates that without the table of specification, it will be very difficult to have a wider content coverage of the main issues in the syllabus and the main content of the test. In the classroom assessment, the curriculum and instruction determine the domain of achievement tasks. Amedahe and Asamoah-Gyimah again, stated that to ensure content validity of teacher-made tests, one must clearly specify the domain of instructionally relevant tasks to be

used to measure students' achievement. Notar, Zuelke, Wilson and Yunker (2004) have confirmed that the means to achieve content validity of teacher-made tests is to develop a table of test specifications for the test. The table of specification will ensure that content tested agrees with the content taught. Miller, Linn, and Gronlund (2009) have indicated that the table of test specifications ensures that teacher-made tests produce scores that represent both the content areas and the learning outcomes that the test constructer desires. This allows the tester to identify the learning content at each level of Bloom's Taxonomy. The obstacle associated with planning a table of test specifications for each test is that these tables require considerable time and effort to develop (Parr & Bauer, 2006). However, the time and effort used to develop a table of test specification can ensure that the test is valid and reliable (Notar et al. 2004).

In conclusion, content validity remains one of the important areas that teachers as well as assessment agencies must not overlook in the test construction process of assessment. If the content validity is low as a result of the test constructors' failure to set questions which forms part of the important domain or match very well the instructional objectives to the behaviours, it becomes very difficult to tell if students have mastered the important domain. The poor performance of most students in external examinations may be due to the fact that they are familiar with tests of low content validity and which, therefore, do not challenge them to bring out their true competencies.

Criterion Validity-evidence of Achievement Tests

Amedahe and Asamoah Gyimah (2016) indicated that "criterion-related validity is concerned with the empirical method of studying the relationship between the test scores or other measures and some independent external

measures" (p. 53). Thus, criterion-related evidence serves as a basis for using the test scores to predict an individual's standing on a criterion measure of interest. This is achieved when the performance of the student on the test is checked against a standard measure, which is a direct and independent measure of the specific behaviour which the test is designed to predict. In other words, a test's criterion related validity is measured by how well it relates with some accepted criteria of the behaviour being measured. The literature distinguishes between two types of criterion-related validity evidence referring to whether the external criterion is obtained at the time of the administration of the test, or at some time in the future. Kinyua and Okunya (2014) described the two main types of criterion-related evidence (concurrent and predictive validity evidence) as follows:

Concurrent Validity Evidence: This refers to the extent to which an individual's current status on a criterion can be predicted from their prior performance on an assessment instrument (Nitko & Brookhart, 2007). For example, the newer State of Anxiety Scale can be administered at the same time as the older and much more established Taylor Manifest Anxiety Scale so that if the results of the former are comparable to the later, the former test will have passed criterion validity test.

Predictive Validity Evidence: This is where the performance of one test is used to predict the potential performance in another test. For example, the performance of a student's in the West Africa Senior School Certificate Examination (WASSCE) used in predicting the Grade Point Average in the University at the first semester. From the review, internal examination results may affect the predictions of assessors about students' performance in specified

situations, if the test comprises of ineffective items. For instance, if test constructors do not strive hard to set questions that control the effect of item sequencing especially in a multiple-choice test, tests scores maybe used wrongly to predict students' performance in a specified situation. Students' results from such tests may not reflect their true competencies on the trait been measured.

Construct Validity-Evidence of Achievement Test

According to DeVellis (1991), construct validity-evidence is directly concerned with the theoretical relationship of a variable (e.g., a score on some scale) to other variables. It is the extent to which a measure 'behaves' the way that the construct it purports to measure should behave with regard to established measures of other constructs (p. 46). Also, Messick (1989) defined construct validity as a unified evaluative judgment of the degree to which empirical evidence and theoretical rationales support how adequate and appropriate references and actions are made based on test scores or other approaches of assessment. Considering the purpose of construct validity, Moss (1992) pointed out that the essential purpose of construct validity is to justify a particular interpretation of a test score by explaining the behaviour that the test score summarises (p, 233). This means asking whether the interpretation given to the test score truly summarises the behaviours. This mean that a construct must be operationalized and syntactically defined to be able to measure it effectively. Here, the operationalization of the construct involves developing series of measurable behaviours or attributes that correspond to the latent construct. In the same way, defining the construct syntactically means establishing assumed relationships between the construct of interest and other

related construct or beahviour (Benson, 1998; Crocker & Algina, 1986; Gregory, 1992).

In sum, construct validity seeks to ensure that the test is actually measuring the intended attribute and no other extraneous attributes. For example, if a social studies test is designed using difficult vocabulary beyond the level of the learner then such a test may likely have an effect on the main construct being measured because it is measuring other constructs (vocabulary besides the intended construct of sociability. Test of this nature is said to have low construct validity. Linking the concept to item sequencing, students may likely perform poorly on the intended construct if the test items are sequenced to have effect on the measurement of the intended construct.

Criteria for Evaluating Validity of a Test

Whatever the type of validity a tester is intending, Linn, Baker and Dunbar (1991) proposed eight criteria for evaluating validity in assessment that cross-cut the above types of validity. These are the:

1. Consequences, that is, on the effects of the assessment on the learner the test constructor will be asking questions regarding intended purpose of test as and to what extent the learner is prepared to live by this purpose.
2. Content quality focuses on the consistency with current content conceptualization.
3. Transfer and generalizability mean the validity focuses on the assessment's representatives of a larger domain.
4. Cognitive complexity focuses on whether the cognitive level of knowledge assessed is corresponding with the learner’s experiences.

5. Meaningfulness addresses the aspect relevance of the assessment in the minds of students.
6. Fairness deals with aspect of extent to which the test items are taking into account potential individual differences among learners.
7. Cost and efficiency focuses on the practicality or feasibility of an assessment in terms of the cost of producing and administrating the test and time required to complete the tasks.

From the above criteria given by Linn, Baker and Dunbar, a valid test should embody these factors. Making reference to the second criteria and linking it to item sequencing, the question that the test items are asking should reflect the purpose in which the test items are written other than sequencing the items to have effect on the students' responses. Furthermore, the sixth criteria also spell out that fairness should be considered in writing the test items in order to cater for the individual differences among learners. This makes it prudent for tests constructors to consider the effect that item sequencing can have on students' responses in ensuring fairness at the test construction stage.

Factors Affecting Validity of Assessment Results

A number of factors affect the degree of validity of assessment results. These factors tend to reduce the degree of validity of the results affecting its uses and appropriate interpretations. Amedahe and Asamoah-Gyimah (2016, p. 59) outlined some of these factors as:

1. Unclear directions: To help the student to respond meaningfully to test items, there is always the need to provide clear directions to them. Directions that do not clearly indicate to the testee how to respond to the tasks and how to record the responses will tend to reduce the validity of

the results. This is because students may get confused over how to respond and how to record their responses that may in turn affect their performance.

2. Reading Vocabulary and Sentence Structure: The sentence structure should not be too complex for the level of the students. When the vocabulary and sentences are too difficult and complicated for the students taking the assessment, it will result in the assessment measuring the student's reading comprehension ability rather than the student's achievement in a subject matter content. In this case, the use and interpretation of the test scores may have low validity.
3. Ambiguity of Items: When test items are ambiguous, they can be interpreted in different ways. This can result in misinterpretation and confusion. This will intend reduce the validity of the students' results.
4. Inadequate time limits: Students need to be given adequate time to complete a test. Too short duration to complete tests may deprive testees the opportunity to reason and respond accordingly while too long duration may also cause testees to finish very early and misbehave by sometimes cheating or changing the right answer and which therefore introduces a lot of biases to their results.
5. Poor Construction of Items: A poor construction can take the form of the items providing clues to the answer. In the selection type of test items the clues are provided by certain word(s) in the stem of the item (e.g, is/an) or grammatically inconsistent options. Poorly constructed items makes the tester deviate from his purpose of testing and this therefore reduce the validity of the results.

From the review, if assessors and other stakeholders can give sound interpretations to students' assessment results and even put their scores into appropriate use, I believe that assessors in constructing tests items should be cautious since a poorly constructed items can reduce the validity of the tests. This also tolls in line with the fact that test constructors must be in the know that sequencing the test items in a particular order can duly produce scores which might not reflect the true ability of students and thereby reducing the validity of the tests.

Reliability of Assessment Results

Reliability refers to the consistency of the scores obtained. That is, how consistent the scores are for each individual from one administration of an instrument to another and from one item to another. Reliability again is defined as the measure of stability dependability, trustworthiness and consistent a test is in measuring the same thing each time (Worthen, Borg & White, 1993). In making reference to the reliability of teacher-made tests, Amedahe (1989) indicated that high reliability of teacher-made tests is required because of the important decisions based on these tests. This assertion is very crucial particularly in Ghana where standardised tests is non-existent and major decisions about students and learning are made based on the scores of teacher-made tests.

For the students' test scores to be consistent in both internal and external examinations then, internal examinations like the teacher-made test must be crafted very well to elicit the desirable behaviour. In crafting the items in order to obtain stable, dependable, trustworthy and consistent test scores as spelt out by Worthen, Borg and White, item sequencing remains one of the important

factors to consider. This is because when items in a test are sequenced in a particular order to have effect on students' responses and for that matter their scores, scores obtained from one administration of the test to another may not reflect their actual competencies.

Methods of Estimating Reliability of Test results

Amedahe and Asamoah-Gyimah (2016) identified and explained the various methods used in estimating reliability as:

Test-Retest reliability: As the name implies, this method of estimating reliability or stability of test scores involves the administration of test on a group of students two times with a given time interval between the two administrations of the test. The situation where a person comes out with the same or similar results in both tests, the test is said to have higher reliability and vice versa. The test-retest method seems to yield the most reasonable estimate of test reliability but some measurement experts agree that this method is not without problems (Allen &Yen, 1979; Crocker & Algina, 1986; Tamakloe, Atta & Amedahe, 2005). The most serious problem with this method of estimating reliability is the potential for a carry-over effect between testings: the first testing may influence the second testing (Allen & Yen, 1979, p. 77)

Alternate/Equivalent forms reliability: This method provides a measure of the degree to which generalizations about student performance from one assessment to another are justified. Alternate/equivalent forms of a test are made up of tasks carefully developed from the same table of specification. This, therefore means that the alternate form method requires constructing two similar forms of a test with the same content and level of difficulty and administering both forms to the same group of examinees. Crocker and Algina (1986) added

that the two forms should be administered within a very short time period, allowing only enough time between tests so that examinees will not be tired. It is considered desirable to balance the order of administration of the forms so that the examinees are randomly assigned to the first test form followed by the second test form, whereas the other half take the second followed by the first.

The correlation coefficient between the two sets of scores is computed, usually with the Pearson Product Moment formula. This correlation coefficient is called the coefficient of equivalence. The higher the coefficient of equivalence, the more confident test users can be sure that scores from the different test forms may be used interchangeably.

Split-Half reliability: This method of reliability estimates the internal consistency of a test. Internal consistency estimate provides information on how items in a test measure the same dimension of an attribute of concern. Again, internal consistency is achieved when performance on each item by testees correlate well with their total performance on the test.

Kuder-Richardson reliability: This method of reliability estimates is also concerned with the internal consistency of the test. They focus on the consistency with which students perform one task to another. K-R20 is suitable for determining the reliability of dichotomously scored items, that is, items scored either 0 or 1 such as multiple-choice items. The formula is as follows:

$KR20 = [n/n -1] [1- \Sigma pq/SD_x^{\,2}]$

where n = the number of items

$SD_x^{\,2}$ = the total variance of the test

p= proportion of examinees who got an item correctly

q= proportion of examinees who got an item wrong

The above formula was later modified to become more generalized in order to cover that of constructed response items. This general version which was developed by Cronbach was named coefficient alpha. The formula is as follows:

Coefficient alpha (α) = [n / n-1] [1- ΣSdi) ^{2}/ (Sd_x ^{2}]

Where; n = the number of items

Sdi^2 = the variance of the item i

With each of the two formulae, a single test is administered and the required statistics are determined and substituted into the formulae to determine the reliability coefficient. The reliability coefficient simply indicates the percentage of the score variance that can be attributed to the construct being measured.

Inter-Rater reliability: This type of reliability estimate is normally used when dealing with essay type test. Here, the same test is scored by two raters to determine the consistency of errors in both tests. In the situation where both raters come out with either both high or low scores, it shows that the errors in both tests are consistent and therefore decisions by users will be reliable.

Factors Affecting Reliability of Assessment Results

Again, Amedahe and Asamoah-Gyimah (2016. p. 79) identified some factors that may affect the consistency of assessment results as follows:

Characteristics of a Test: The weaknesses of the items in a test can take the form of poor wording, unclear directions/instructions or ambiguity of the items. These factors generally create difficulty for the student with regard to what exactly is either being measured or what to do and tend to lower the consistency of their performance. Ambiguity of test items, for example, may lead to differences in how an item is interpreted and may give rise to guessing which reduces reliability.

Test Difficulty: When a test is difficult, students may be induced to guess the answers to the items (selection type items) or bluff (e.g., essay type items) hence introducing errors into the scores.

Test Length: A test with limited number of items is not likely to measure the abilities or behaviours under consideration exhaustively and accurately and therefore errors may occur in the measurement.

Time Allocated to the Test: When the time allotted for taking a test is too short, testees would not have enough time to read and think about the problems before answering them. The test is completed in a rush. This could lead to fluctuation in performance from one occasion to another. It could also lead to guessing. On the other hand, if the time is too long, the fast students would finish and be tempted to assist their colleagues and friends leading to irregularities. It is obvious that cheating during tests cannot result in consistent scores from one occasion to another.

Subjectivity in Scoring: If a test is subjectively scored, inconsistencies are allowed to create random errors within the scores that in turn lower the reliability of the test.

Testing Condition: When uniformity of the testing conditions is not ensured during test administration inconsistencies are likely to be introduced into the performance of the students which would affect the scores.

Group Variability: Group variability influences reliability because reliability coefficients are directly influenced by the spread of scores in the group assessed. Other things being equal, the larger the spread of scores, the higher the estimate of reliability will be. Because larger reliability coefficients result when individuals stay in the same relative position in a group from one assessment to

another, it follows that anything that reduces the possibility of shifting positions in the group also contributes to larger reliability coefficients. In general, if the group tested is heterogeneous the reliability of the scores tends to be high.

In line with the validity and reliability theories reviewed, Afful (2014, p. 45) asserts that one would appreciate why assessors have over the years yearned for assessment scores of students that are consistent over time; they partly contribute to the measurement of a construct and the validity of assessment results. Making reference to the characteristics of a test as one of the factors that affect reliability of tests scores, it is an undeniable fact that sequencing of test items remains an aspect of this characteristics. That is to say, when test items are sequenced or arranged in an order that can have effect on students' responses and hence their test scores, test scores that will be obtained may remain consistent from one administration of test to another but might reflect the actual trait intended to be measured by the test.

Furthermore, with regards to test difficulty as another factor that can affect reliability of test scores, especially when test items are sequenced or arranged in Hard-To-Easy (THE), it is likely to be progressively difficult for students to answer and hence may be induced to guess the answers to the items especially in the multiple-choice test which can eventually introduce errors into the scores to affect the reliability of the scores. In addition, the subjectivity in scoring factor among the factors affecting reliability may be avoided specially in using multiple-choice tests which can bring about high reliability but, in a situation, where the sequence or the arrangement of the test items can have impact on students' responses and hence their test scores, high reliability in this case may not be achieved.

In conclusion, the theory of validity and reliability have been comprehensively explained and related to why it bothers the concept of the effect of item sequencing in multiple-choice tests and students' performance.

Conceptual Review

Nature and Relevance of Core Mathematics and Integrated Science

In general, education is considered as the most important factor for development. This is a fact, which has been highlighted by a World Bank statement that the economic returns to investments in education (as measured by productivity and income) are higher than these in physical capital and are essential to the successful investments in other sectors of the economy (World Bank, 2004).

In the study of Mathematics, knowledge and competences which are gained can be applied in numerous fields of study as well as some aspect of human life (Mensah, Okyere, & Kuranchie, 2013). In support of the above assertion, Anthony and Walshaw (2007) as cited in Mensah et al., 2013) also indicates that Mathematics plays an important role with regards to shaping how an individual deal with their private, social and civil lives. According to Mbugua, Kibet, Muthaa and Nkonke (2012) as cited in Asamoah (2018), Mathematics is seen by every society as the basis of scientific and technological knowledge that is very important in the socio-economic development of every nation. Considering the importance of Mathematics to other fields of study, Umameh (2011), Tshabalala and Ncube (2013) indicates that Mathematics as a subject has a direct relationship with other subjects that are technical and scientific in nature and thus, mathematics is seen as the bedrock for scientific, technological and economic advancement of any country.

Mefor (2014) also comments that mathematics in general and especially core mathematics relates or connect to everything in the universe which is connected to the daily life of the individual. He further states that, the subject is very important to human life and humans cannot function meaningfully without. UNESCO (1999) as cited in Asamoah (2018) points out that core knowledge in mathematics holds the key to Africa's industrialisation and poverty alleviation.

In the Ghanaian perspective, core mathematics has been accorded the needed importance in the curriculum and content in all educational policies from the basic to the secondary level (Mereku, 2012). This is principally important because, core mathematics is essential in the development of basic computational skills and knowledge which consequently fosters the ability to be accurate in solving problems as well as doing further mathematics (Sa'ad, Adamu & Sadiq, 2014). They further state that a careful study of core mathematics develops logical and abstract thinking skills, solve problems with related mathematical knowledge and expose the individual to creativity and curiosity in discovering solutions to their day-to-day problems. Similarly, Akinyi (2003) contends that core mathematics is very important in the sense that, people working in the public and other sectors require some basic knowledge of mathematics regarding their daily endeavours. Akinyi further notes that it is in this regard that core mathematics is made a prerequisite subject of study in all tertiary institutions that offer scientific and business courses as well as a requirement for most employment opportunities.

The relevance of Mathematics can be likened to science since it is also one of the most valued subjects and therefore made a pre-requisite for both first and second cycle institutions. The development and progress of any nation

largely depends on the quality of scientific and technological research. Science education is one of the important areas of the curriculum to enable people to make sense of the world and to use the resources at hand (Sjøberg, 2002). Science should therefore be given high priority in the school curriculum.

Development and application of science and technology is globally recognised to be vital for a country's developmental efforts. In Ghanaian perspective, science education is needed to produce the necessary human resource and skilled labour force to manage our local industries. A country's development rests on science and its application in the world of work and industry and competent workers and citizenry need a sound understanding of science and mathematics (Shadreck & Mambanda, 2012; Anamuah-Mensah, August 2004). A strong science and technology base therefore constitute the currency for social and economic transformation of nations (Anamuah-Mensah, August 2004).

Currently, many jobs in Ghana demand advanced skills, a skill which requires people to be able to learn, reason, think creatively, make decisions, and solve problems. A high quality of science education can play a major role in contributing to these skills. Yet education including science education gets, according to UNDP report, a small share of the economy, especially in developing countries, and that the proportion of foreign assistance going into this area is rather small for most donors (UNDP, 1993-2005). Ghana as a developing country is not an exception to this phenomenon. Science and Mathematics offer such a foundation for students to be critical and analytical in their thinking. It is, therefore, gratifying that Ghana's constitution mandates that all efforts should be made to ensure education up to pre-tertiary levels become

free (Republic of Ghana, 1993). Development of economies is closely linked to the advancement in the fields of mathematics and science education to occasion accelerated expansion. Secondary Education Improvement Project (SEIP) is but one of the interventions to drive the agenda forward.

The above expositions make the relevance of Mathematics of Science undisputable hence their inclusion as the pre-requisite subjects for both basic and secondary education. However, students' low performance in these subjects becomes disturbing. Therefore, the study sought to provide trends in students' performance in Mathematics and Science and offer insight into some of the reasons causing the trends including tests item sequence, especially in the MCTs.

The Concept of Assessment

The concept of assessment has been defined by different authors depending on their perspective and goals about the concept. According to Nitko (2004) as cited in American Federation of Teachers, National Council on Measurement in Education and National Education Association, assessment is a method of obtaining information that is used to make decision about students' curriculum and programme and national policy. This assertion considers assessment in the context of getting information about students with regards to their ability in order to make decisions about them. Similarly, Amedahe and Asamoah-Gyimah (2016) see assessment as the process of obtaining information which is used to for making decision about students, curricula, programmes and educational policies. The decision component of their assertion highlight how important assessment is since its process helps to bring out the true reflection of traits of students in order not to make wrong decision

about their abilities. In support, Goodrum, Hackling and Rennie (2001) also notes that classroom assessment is an essential component of the teaching and learning process as it helps teachers in classroom decision-making.

From the various perspective of authors, assessment is generally seen as a way of gathering information about students in order in make decisions concerning the students' wellbeing in terms of the curriculum and programme and national policies on education. Considering the nature of information that can be gathered about students in order to make decision about them, teachers as well as test constructor gather information such as students' knowledge, ability, understanding, attitude and motivation (Ioannou-Georgiou, 2003). This however means that assessment is purpose driven since it is used to make decision and according to McMillan (2001) there are number of "essential" assessment concept that teachers must be abreast with in order to make a valid decision about students as well as using various means to obtain information in order to avoid biases in assessment. McMillan is of the view that whatever decision made from the assessment procedure should be as a result of combining different sources evidences that can be inferred from students' ability.

In education, assessment is conducted for several reasons and nature of the assessment process usually mirrors the purpose with which the assessment is being carried out. For instance, Dun, Morgan, O'Reilly and Parry (2004) posit that assessment is used to achieve several different purposes which is far broader than just measuring students learning and maintaining pre-set standards of achievement. They also proposed that good and efficient assessment procedure is that which focus on diagnosing students' difficulties, measuring students 'achievement over time, motivating students, judging mastery skills,

evaluating teaching efficiency and providing feedback to students. In addition, Salvia and Ysseldyke (1978) also outlined five purposes of assessing students. These include screening, placement, programme planning, programme evaluation and assessment of individual progress.

Similarly, Amedahe and Asamoah-Gyimah (2016) also emphasized some educational purposes of assessment as follows:

Planning and organization of instruction: Educational assessment helps teachers to plan and organize their teaching activities. Before one can do any meaningful teaching, one needs to have a clear idea of the entry behaviour of one's students. The entry behaviour is basically what your students know and what they can do as well as they cannot do. It can also include their abilities, interest and deficiencies. Understanding the entry behaviour will assist the teacher to determine what should be taught and it is through assessment that you will know students' entry behaviour in order to plan instructional activities for them.

Instructional management decision making: Educational assessment also helps to make decisions including instructional management in the instructional process. Through assessment, instructions are diagnosed and remediated. This means that assessment is used to diagnose students' difficulties in the teaching and learning process in order to devise appropriate and alternative methods for maximizing students' understanding. Educational assessment is necessary for instructional measurement decision making. It is not sufficient to plan, organize and deliver your instruction. There is the need to be on the lookout for how to help your students to improve as well as make sure you (the teacher) improve upon your instruction.

The purpose of assessment with regard to instructional measurement decision-making has several aspects. One aspect of the instructional management decision making is feedback to the teacher. As a teacher you need to diagnose your instruction and remediate the aspect(s) which have not been very effective. Effectiveness of your instruction is partly determined by student responses to questions you pose to them. When the assessment of their understanding does not provide the expected answer(s) that is an indication that the instruction has not been very effective.

For managing and improving your instruction, your pupils also need information on how well they have been doing in relation to the set instructional objectives and standards. This is what is called feedback to pupils. Your pupils need feedback on their performance. Unless you assess them in some form you will not be able to provide the necessary feedback to individual pupils. Thus, one purpose of assessment of pupils is to provide feedback to them to facilitate their learning. It is, however, important to note that for the feedback to the pupil to be effective, the pupil should not be given only a mark but shown where he/she has gone right or wrong. The pupil should be helped to correct his or her errors before going on to a new material.

Motivating Students: Another significant purpose of assessment which unfortunately many teachers don't pay attention to is motivating students. Motivation is simply concerned with that which initiates and sustains one in an act. Motivation, therefore, can activate and direct students learning by sustaining their interest. Assessments in the form of tests and quizzes motivate pupils to learn. Imagine, you will not be assessed at the end of an instructional session. Will you really be motivated to complete the session in time and

diligently? This situation applies at all levels. Your pupils will not be motivated to learn, if they know that they will not be assessed at the end of units, course or programme. There is no doubt that classroom assessment directs the learning of students.

Grading students: As a teacher, you are expected to assign scores or grades to your pupils based on how good their performance or achievement is, taking into consideration your objectives and standards. Although as a teacher you continually assess your pupils' progress in many ways, it is necessary to formally evaluate them using grades. The marks or grades that you assign to your pupils represent your evaluations about the quality or worth of achievement of the important learning objectives. Usually, the tests used in the assessment process give your pupils help you to assign the marks or grades.

Guiding students: Assessment has an important role to play in the guidance of the pupils. Guidance is one of the pupil-personnel services provided in a non-instructional setting to cater for the needs of the pupils including educational, emotional, moral and adjustment needs. Guidance is a form of a systematic assistance aside the regular instruction to pupils to help them assess their abilities and capabilities and to use that information effectively in daily living.

To be able to provide effective guidance to pupils, there is the need to obtain the necessary and relevant information on the pupils to aid you. Thus, assessment results are useful in providing guidance to pupils. At the Junior Secondary School level where students will be making choices of subjects, they need guidance from the teacher. You can provide this guidance based on the performance of the individual students and their interest. In a nutshell, one purpose of assessment is to facilitate pupil guidance.

Making selection and placement decisions: another important purpose of assessment is making selection and placement decisions. In your work as a teacher, we guess, you were at one time or another involved in selection and placement decision making. Assessment provides the basis for the grouping of children with reference to their ability to profit from different types of school instruction and the identification of the intellectually retarded and the gifted (Cunningham, 1986). Nitko (2004) has pointed out that sometimes, based on assessment results, a decision is made that result in a person being assigned to one of several different but unordered categories of programmes. According to Cronbach and Glaser (cited in Nitko, 2004), these types of decisions are called classification decisions. These decisions result in either assigning students in the same classroom to different groups for effective instruction or assigning students to special education classes. Cunningham (1986) however cautioned assessment experts about the over reliance on assessment results in assigning students to special education classes by pointing out that, intelligence tests are only one component of the assessment of students referred for possible placement in special classes (p. 11).

On the issue of placement decisions, Cronbach (1960), Kubiszyn and Borich (1984) and Nitko (2004) have pointed out that placement decisions are made after an individual has been accepted into an educational programme. They continued by noting that placement decisions basically involve using assessment results or test data to determine where in a programme an individual is best suited to begin work. Such decisions are characterised by assigning individuals to different levels of the same general type of instruction or education based on their ability, with no one rejected by the institution (Nitko,

2004). Promotion in Ghanaian schools from one class or form to another which in most cases is based on the performance in tests of the previous class is an example of a placement decision

Certification decisions: On credentialing and certification decisions, Nitko (2004) and Amedahe and Gyimah (2003) explained that they are concerned with assuring that a student has attained a certain standard of learning. Credentialing and certification may be mandated by state legislation as in the USA and executed by an external examining body at the state level. In Ghana, certification and credentialing of students is done by the WAEC. With the introduction of the practice of continuous assessment as a result of the educational reforms in 1987, Ghanaian classroom teachers contribute 30% of the total marks for certification of students at the JSS and SSS levels (Amedahe, 2000; Pecku, 2000).

Watering, Gijbels, Dochy and Rijt (2008) are also of the view that assessment is a tool that is used for improving learning. They further explain that assessment as a process does not only assign grades and determine whether students achieve objectives but also become a learning tool. Their perspective about assessment does not only focus on assignment of grades to determine whether the objectives of the instructional process have been achieved but also using the assessment process as a learning tool which seeks to improve upon students' learning. Nenty (2005) also emphasizes the importance of assessment in learning by positing that learning is an ultimate dependent variable in education and everything done in the name of education impacts positively or negatively on learning. He further said it is through assessment that learning and hence quality of education is defined and any improvement in these depends

ultimately on the quality of and improvements in assessment. This means that assessment is at centre of all educational activities and every activity in education looks on to assessment to establish its validity and effectiveness. The quality of assessment bears on the quality of educational inputs, processes and products and hence on the quality of education enjoyed by the society.

During the assessment process, teachers interact with learners in order to obtain information that identifies and diagnoses their learning difficulties. This is likened to the assertion of Tamakloe, Amedahe and Atta (2005), assessment is a process where one person engages in some form of interaction with another person in order to obtain and interprets information about that other person with regards to the person's knowledge, understanding, abilities and attitudes. This means that assessment is also seen as a communication tool which gives some form of feedback to the learning process (McAlpine, 2002). Emphasizing on using assessment as a tool for feedback, Cockcroft (1982) also added that assessment serves several functions including reporting students' progress to interested clients. In using assessment as a feedback to students, feedback facilitates and improves their learning provided they are integrated in the instructional process. It also gives students the opportunity to correct their own errors before a new instruction (Nitko, 2004). Nitko added that feedback in assessment is also of great benefit to the teacher because it provides feedback about how well students have learnt and how well the teacher has taught.

Assessment is also seen by other authors as the process through which students' traits or characteristics are observed in order to describe them using a classification scheme. For instance, Ormrod (2008) defined assessment as the procedure of observing a part of student's beahviour and making inferences

about his or her knowledge and abilities. This means in the assessment process, only a sample of a student's behavior can be obtained in order to describe their knowledge and abilities. In making inferences from students' behavior, Delclos, Vye, Burns, Bransford and Hasselbring (1992); Poehner (2007) posits that the knowledge, skills, attitude and beliefs of the learner must be in measurable terms in order to be able to quantify using scale. Capraro, Roe, Caskey, Strahan, Bishop Weiss and Swanson (2012) added that in the classroom level, the teacher decide which specific knowledge, skills, attitudes and beliefs warrant; at what point and for what specific purpose they should be assessed; and which tools might best accomplish these classroom-based assessments.

Assessment as a concept as elaborated above is also done in the learning of Mathematics and Science as subjects, this is to say that in the study of Mathematics and Science in our Senior High Schools in Ghana, teachers employ assessment procedure in order to obtain information about students with regards to their traits. This information is used by teachers to makes decisions about students.

The Concept of Test

The use of testing (paper and pencil tests) in schools officially started with the introduction of paper in the 12th century (Dubois, cited in Cunningham, 1986). Assessment by means of written tests was first used by the Jesuits at St Ignatio (Cunningham, 1986). He stated that the development of academic tests was pioneered in Britain, particularly in the University of London. Under its initial chapter, testing and awarding of degrees were recognised as a legitimate basis for decision making. It is important noting however that, prior to this period, academic testing (oral testing) in schools had already begun. As stated

by DuBois (cited in Anastasi, 1982), among the ancient Greeks, testing was an established adjunct to the educational process.

Tests were initially used to assess the physical as well as intellectual skills. Anastasi (1982) pointed out that the Socratic method of teaching with its interweaving of testing and teaching has much in common with today's programmed learning. On the account of Ebel (cited in Amedahe, 1989) and Anastasi (1982), from their beginnings in the middle ages, European universities relied on formal examinations in awarding degrees and honours". These examinations, were however, largely oral. The birth of psychology as a science in testing took place in the laboratory of Wilhelm Wundt in Leipzig, Germany. However, Galton (1967) developed psychological testing by studying individual differences. Many of the methods in psychological testing can be traced to Alfred Binet who focused on the assessment of human abilities. Thorndike and Alfred Binet contributed immensely to the development of the current principles in testing. Thorndike (1903) published the first book in educational measurement. In this book, he indicated that whatever exists at all exists in some amount.

It is worth noting that achievement testing in Ghanaian schools today involves the use of multiple-choice and other objective type tests. Discoveries, innovations and development continued in the field of educational measurement over the years and by 1945 many of the theories and principles used in educational testing today had been developed (Amedahe, 1989). Today, the concept of test is seen as a concept which embodies the use of an instrument or systematics procedure in observing and describing one or more characteristics of students using a numerical scale or a classification scheme. Tests are

administered to students, marked and scores assigned to students based on the degree of correctness of the answers provided (Amedahe & Gyimah, 2016). He further noted that tests, therefore, describe the student using a numerical scale. Tests are used to quantify students' characteristics such as performance or achievement in a subject area and to compute the marks in order to describe a characteristic, one must add the scores on each item. If a student has more of the characteristics, it is expected that the students get higher marks and vice versa. The role and use of tests in assessment of students in our todays educational is immeasurable based on its potency to reveal to assessors some information about students in order to take decisions about them. Just like assessment, tests also serve many useful purposes and according to Ohuche and Akeju (1976) tests are useful in the educational system, in industry and in the world of work. He further outlined some of the purposes of tests as follows:

1. Stimulus for Studying: Tests and test results provide impetus for learning and the stimulus which an average pupil or student needs for studying.
2. Administrative Decisions in Education: Test results are used in making administrative decisions about students, teachers and the curriculum.
3. Diagnosis: Test results are used in identifying weaknesses and strengths in a class of pupils. This aids in the desirable effort of giving pupils individual and remedial attention. One other diagnostic use of tests is for guidance and counselling to the most appropriate course to undertake.
4. Selection and Placement: Selection tests are very useful instruments for picking round pegs for round holes and square pegs for square holes. Selection picks among many people the best for a course, career or

training and it is institution-centred. Placement chooses among many qualified career persons, the best for a position.

5. Certification: The most popular among the uses of examination is to determine who will receive what certificate. These certificates serve as passports to job, higher institution and instrument for social mobility.
6. Maintenance of Standards: Standards represent the minimum degree of excellence which society can accept. For example, professionals like lawyers and medical officers must meet certain standards before they are allowed to practice their profession. Usually these standards are enforced through examinations which may be written, oral or practical.
7. Research: Most researchers in education depend somehow on tests. Tests are thus very useful tool in the hands of researchers, in curriculum work, teaching methods and learning theories.

Ohuche and Akeju's perspective on the purpose of the tests is similar to the perspective of Amedahe and Asamoah-Gyimah (2016) on the purposes of assessment. This affirms that tests are the main and commonly used instruments in the assessment procedure. This is because in gathering information about students in order to make decisions about them in the assessment process, test becomes one of the main devices for obtaining such information. Furthermore, the purposes of tests according to Salvia and Ysseldyke (1978) is in agreement with the assertion of the above authors especially the assertion of Asamoah-Gyimah (2016) that, tests are used for screening, placement, programme planning, programme evaluation and assessment of individual progress.

In conclusion, the above literature has elaborated on the historical development as well as the concept of test in educational assessment. It was

noted in the literature on the historical development of tests that, test in the form of paper and pencil in schools officially started with the introduction of paper in the 12th century (Dubois, cited in Cunningham, 1986). Today tests come in various types and forms. It was however also noted that tests come with many purposes in the assessment process. These purposes were found to be similar to the purposes of assessment.

Item Formats

Choosing a test item format for assessing students' achievement is considered to be one of the greatest elements in test item construction. According to Allam (2007), item format differs in the degree of freedom given to students to express themselves and the skills and knowledge they acquired. This means that the test constructor must choose the best item format to assess students' achievement because they vary in their typical cognitive demand and in the range of conditions they sample (Martinez, 1999). Classroom teachers as well as assessment agencies construct and use a number of test item formats depending on the kind of learning outcome, they want to measure in the assessment process. Test item formats are broadly classified into two; namely the objective-type test and the essay-type test.

Objective-type test is the kind of test which is characterized by its total objectivity and unambiguity in scoring. An objective test requires a respondent to provide a briefly response which is usually not more than a sentence long. According to Nitko and Brookhart (2007), objective type test is the kind of test whose procedure by which the correctness of the responses to the items will not in any way differ from scorer to the other. This means that the scoring of correct responses of the objective tests will not in any way change from one scorer to

another. In addition, Amedahe and Etsey (2003) posit that the objective tests items normally consist of large number of items and the responses that are scored objectively to the extent that competent observers can agree on how responses should be scored. They further noted that, there are two major types of objective tests. These are the selection type and the supply type. The selection type consists of the multiple-choice type, true or false type and matching type. The supply type has variations as completion, fill-in-the blanks and short answer. In their explanation to the various forms of tests under the two major types of objective tests, they noted that a true or false test item consists of a statement marked true or false. Here the respondent is expected to demonstrate his command of the material by indicating whether the given statement is true or false.

One of the limitations in constructing this type of objective test is the probability of getting right answer by guessing is 0.5. It can be used to assess only a limited number of educational objectives, and can be used to evaluate definitions, facts, meaning of expressions, recognition, and interpretation of charts / graphs. One advantage of the true-false format is that it is suitable for classroom short–time evaluation.

Another form of test under the supply type of the objective test formats that mostly eliminates guessing is the completion type. The short answer is the type of objective test item that is known as the supply, completion, and fill-in-the blank. It consists of a statement or question and the respondent is required to complete it with a short answer usually not more than one line. It is useful for testing knowledge of facts or recall of specific facts (e.g. “knowledge objective” in Bloom’s taxonomy of educational objectives). One demerit of the format is

that it can be such that there is more than one answer, and that situation makes the scoring of the test subjective.

According to Amedahe and Etsey, a matching-type test is a special form of the multiple-choice format which is not very flexible, though it is useful for testing variables that are compatible. The matching type of objective consists of two columns. The respondent is expected to associate an item in column A with a choice in column B on the basis of a well-defined relationship. Column A contains the premises and column B the responses or options.

As already explained in the concept of MCTs, Amedahe and Etsey also added that a multiple-choice test item is a type of objective test in which the respondent is given a stem and then is to select from among three or more alternatives (options or responses) the one that best completes the stem. The incorrect options are called foils or distracters. Considering the usefulness of the objective-type test items, Amedahe and Etsey continued that, the tests are mostly useful when class sizes are very large and when there is limited time to submit the results of the test. Mehrens and Lehmann (1991) are however of the view that objective tests are more susceptible to guessing; and the distribution is determined almost completely by the test.

Another important item format which teachers and assessment agencies use to prevent students from guessing is the essay-type tests. According to Reiner, Bothell, Sudweeks and Wood (2002), an essay test requires a response composed by the examinee, usually in the form of one or more sentences, of a nature that no single response or pattern of responses can be listed as correct, and the accuracy and quality of which can be judged subjectively only by one skilled or informed in the subject. Tamakloe, Atta and Amedahe (2005) have

also stated that, an essay test is a test which gives the student/testee the freedom to compose his/her responses to the items usually in the form of a number of logically arranged and related sentences. Its nature is such that no single response in terms of a word, phrase, or clause can be listed as correct. No plausible answers are given the testee and consequently no selection of correct answers is done by the student. Essentially, an essay test provides the student/testee with the opportunity to construct and compose his/her own responses to the items within the given limits of each item.

In conclusion, it has been explained that, the major types of item formats are the objective-type test and the essay-type test. The objective-type test requires students to select from the various options to answer the stem whiles the essay requires students construct their own responses to the demands of the items.

The Multiple-Choice Test (MCT)

Multiple Choice Tests (MCTs) are tests which basically require examines to respond or answer a question or an incomplete statement followed by a list of suggested answers or options or alternatives from which answer must be selected. Case and Swanson (2001); Paniagua and Swygert (2016) defined a multiple-choice test as a test that consists of a stem and two or more options in which the examinee is required to respond by selecting a single correct option where the items are scored dichotomously as corrector incorrect. The assertion of the above authors means that an examinee is entitled to selecting only a single correct alternative from lists of alternatives given and it is scored dichotomously thus scoring wrong or right.

According to Hohensinn and Kubinger (2009), conventional multiple-choice formats consist of a single correct answer option and two to seven distractors. Here, Hohensinn and Kubinger are highlighting on the range of distractor that multiple-choice tests must have. The distractor here simply means the alternatives that are plausible to the correct option.

They also posited that other multiple-choice format contain more than a single solution which the examinee has to select multiple solutions. In this case, scoring rules then differ between counting every correctly marked correct answer option and scoring an entire item as solved only if all correct answer options and no distractors have been marked.

Because the scoring of multiple-choice items is economical and objective in nature, this format is the predominantly used response format in psychological aptitude or achievement tests and educational tests.

Thissen, Wainer, and Wang (1994) however highlighted the concerns raised by other authors on whether using different response formats in assessment alters the measured latent trait and to what extent different response formats influence the difficulty of an item. They further stated that there are some evidences that multiple-choice as well as constructed response formats measure the same proficiency but multiple- choice items are supposed to be generally easier than constructed response items.

Amedahe and Asamoah-Gyimah (2016) posit that multiple-choice tests are considerably widely used as a means of objective measurement of examinees trait. They continue to say multiple-choice tests are widely used because of many advantages associated with its uses such as, they are used as a

diagnostic tool for students' diagnosis, for formative assessment purposes as well as for assessing a broad range of knowledge.

Similar to the assertion of Hohensinn and Kubinger (2009), Amedahe and Asamoah-Gyimah (2016) postulate that "the nature of multiple-choice permit scoring to be done easily, quickly and objectively either by human-beings or scoring machines. These advantages make MCTs suitable for a wide range of purposes ranging from classroom achievement testing to large-scale standardised tests". In addition to the uses and accuracy of multiple-choice tests results, Haladyna, Downing and Rodriguez (2002) also added that MCTs are commonly used test formats used in large-scale educational testing because it is easy to administer, score and it is highly reliable. This means that MCT are widely used due to the accurate results it produces which gives a true reflection of students' ability that can be sued to make sound decision.

Despite the numerous advantages or the benefits that MCTs provide, it is however, worth noting, according to Adebule and Awodele (2015) that such tests are prone to blind guessing which enables examinee to be credited with undeserved scores where an academically poor students stand the chance of getting more marks than knowledgeable students in a particular subject. This means that such a test sometimes makes it difficult to discriminate between high achievers and low achievers. In addition, they further stated that multiple-choice exams are mostly less time-consuming and effortless to score than essay type exams.

In spite of the disadvantages of the nature of multiple-choice tests, Haladyna, Downing and Rodriguez (2002), Hogan and Murphy (2007) are also of the view that some of the pitfalls are from teachers or test constructors in that

sometimes they don't take in consideration construction biases where teachers or test constructors over-represent or under-use certain learning outcome in response options. In addition, Attali and Bar-Hillel (2003) added that both test takers and test makers appear to be biased in favor of answer choices located centrally among the alternatives. This bias will produce answer keys that are unbalanced in the sense that not all answer choices are equally represented. This however makes the construction of good multiple-choice tests generally more complex and time-consuming to create as compared to other test formats. Again, the ability to construct good multiple-choice tests require certain amount of skill and knowledge that teachers and test constructors must develop.

In conclusion, MCTs if well-constructed, provide wide range of benefits to teachers as well as test constructors in the assessment process especially because of the objectivity nature of it which helps in obtaining the true reflection of students' traits. Furthermore, attention must also be paid to how items within the MCTs is sequenced since studies have proven to have effect on students' responses to such items.

Conditions for Using Multiple-Choice Test (MCT)

In the assessment procedure, it is important that the form of test used must be suitable for achieving the intended educational objective if it is the aim of the assessor to obtain scores that accurately represent the degree to which students have reached that objective. The use of MCT as a form of test to achieve an educational objective mostly requires or tests students' knowledge of facts or recall of specific facts (e.g. "knowledge objective" in Bloom's taxonomy of educational objectives) where students are presented with a question known as a stem followed by some alternatives from which students

are made to select the best alternative to answer the question. There have been assertions by different authors on the various conditions that call for the use of Multiple-Choice Tests (MCTs). For instance, Amedahe and Asamoah-Gyimah (2016) are of the view that MCTs are applicable and widely adaptable for the measurement of most educational outcomes and objectives. Burton (1991) also believes that MCTs are only good for measuring simple recall of facts. He further contends that MCT are mostly used for measuring lower-level objective such as knowledge of terms, facts, methods and principles. he however stated that the real value of MCTs are their applicability in measuring higher-level objectives such as those based on comprehension, application and analysis. Etsey (2003), Linn and Gronlund (1995), Nitko and Brookhart (2007), Oosterholf, (2001) Tamakloe, Atta and Amedahe (2005) also added that multiple-choice test can also be appropriate for measuring variety of learning outcomes. This however means that, multiple-choice tests are not only used for measuring recall of facts as stipulated by Burton (1991) but can also be used for the measurement of multiple learning outcomes. They further contend that some of these learning outcomes would be appropriately assessed with this assessment tool (MCT) than others.

According to Linn and Gronlund (1995), multiple-choice items are best used to measure learning outcomes in the knowledge area which important in the school subject. They added that, knowledge of terminology, knowledge of specific facts, knowledge of principles and knowledge of methods and procedures as classic examples of learning outcomes in the knowledge area which can be measured using multiple-choice tests. They continue to explain that knowledge of terminology is how well students are able to understand a

given term by asking them to select from among alternatives another term which is similar in meaning to that given word in the stem. Knowledge of specific facts also requires to reproduce certain facts presented in their learning materials.

Linn and Gronlund further noted that questions such as 'who', 'what', 'when', and 'where' are most commonly used in measuring knowledge of facts. According to them, knowledge of principles also assesses students' ability to remember the basic ways in which something works. There are certain conventional methods and procedures of doing things and when a teacher assesses students on any of these laid down methods and procedures, then knowledge of methods and procedures are being sought. Apart from the learning outcomes explained above, Linn and Gronlund also believe that multiple-choice items could also be used to measure higher learning outcomes. What they contend is that, these learning outcomes would be appropriately measured when outcomes in the knowledge area have been earlier measured.

The above literature clearly specifies the conditions which call for the use of multiple-choice tests. Whiles some authors are of the view that multiple-choice tests are used for measuring lower learning outcomes, others are also of the view it can equally be used for measuring higher learning outcomes.

Item Difficulty of Multiple-Choice Test (MCT)

Multiple-Choice Test (MCT) is a form of objective-type test which allows assessors to easily study its psychometric properties. One of these psychometric properties includes item difficulty. In explaining item difficulty, Crocker and Algina (2008) contend that when an item is dichotomously scored just like multiple-choice test, the mean item score corresponds to the proportion of examinees who answer the item correctly. They further stated that, the

proportion of examinees who answer the item (i) correctly is usually denoted as (Pi) which is called item difficulty. This means that the item difficulty (Pi) for item (i) is defined as the proportion of examinees who get that item correct. Following the definition of item difficulty, Allen and Yen (2002) contend that, although the proportion of examinees passing an item traditionally has been called the item difficulty, this proportion logically should be the item easiness because the proportion increases as the item becomes easier. Allen and Yen further stated that the words "difficulty" and "correct" are best suited for discussions of ability or achievement tests. For instance, if a personality test is being developed, a "correct" response is a response that counts towards the traits and the "difficulty" of an item reflects the popularity of the "correct" responses-that is, the proportion of examinees who choose this response. Item difficulty (Pi) is very useful in the analysis of MCT as it gives information on the kind of item (s) which is/are easy or difficult and should be altered or discarded.

Item difficulty of an item is usually given in range of values. The range of values of item difficulty is called difficulty indexes. A difficulty index according to Case and Swanson (2001) and Paniagua and Swygert (2016) is the percentage of examinees who answered the item correctly; expressed in values from 0 to 1.00. Similarly, De Champlain (2010), Haladyna and Rodriguez (2013), Livingston (2006) and Thorndike & Thorndike-Christ (2010) defined difficulty index as the percent of examinees who answered an item correctly and is usually represented in decimal format. They also added that it is commonly referred to as the p-value and it ranges from 0.00 to 1.00. Again, they contend that difficulty index is useful to especially the instructors and administrators as it alerts them on students' performance on an individual item

and therefore the content it measures. Nitko (2004) is also of the view that, the total fraction of the total group answering the item correct is called item difficulty index. Just like other authors, Nitko also stated that item difficulty index ranges from a low of 0.00 to a high of +1.00.

In relation to the range of values of item difficulty and its interpretations, a lot of submissions have been made by different authors. For example, according to Amedahe and Asamoah Gyimah (2016), for multiple-choice test, the optimum difficulty level of an item is about 0.50 for norm referenced tests. They also added that the recommended range of item difficulty is 0.40 and 0.80. This means item difficulty index of 0.40 or below means that the item is difficult and difficulty index of .80 and above shows that the item is easy. However, item difficulty of 0.50 does not mean every item should have difficulty level of 0.50, it simply means that the average of all items should be 0.50. Again, Allen and Yen (2002) also believe that, item difficulties indexes of about .30 to .70 maximizes information the test provides about differences among examinees (p. 121). This means that if an item records a difficulty index of .70 and above, the item is described as "easy". For an item to be moderately difficulty, the difficulty index should range from .50 to .69 while a difficult item records a difficulty index of .30 or below. In addition, Matlock-Hetzel (1997) also indicated that, if an item records a difficulty index of 0.85 and above, the item is described as "easy". Again, for an item to be moderately difficulty, the difficulty index should range from 0.51 to 0.84 while a difficult item records a difficulty index of 0.50 or below.

Furthermore, Musa, Shaheen, Elmardi and Ahmed (2018) also contend that item difficulty index is a measure of the proportion of the total examinees

who answered an item correctly and is most commonly referred to as the p-value. They further stated that item difficulty index ranges from 0.0-1.0. It is often interpreted as follows: (0.30 = too difficult); (0.30 - 0.70 = recommended, good or acceptable) and (0.70 = too easy). Even though, they believe that an item difficulty indexes between 0.30 to 0.70 is recommended or accepted, assessment expert also consider an item difficulty index of 0.50 – 0.60 to be more ideal and recommended (McAlpine, 2002; Tavakol & Dennick, 2012; Matlock-Hetzel, 1997, Ananthakrishnan, 2000).

In conclusion, the above elaborations have clearly explained the meaning of item difficulty of MCTs and its indexes. It has been noted also that, the indexes of item difficulty and its interpretations vary from author to author depending on the angle each of them is coming from.

The Concept of Item Sequencing in Multiple-Choice Tests

Item sequencing in MCTs is one of the important areas to researchers, instructors as well as assessment agencies because of the effect it may have on student grades or standardized test scores. Item sequencing as a concept is used to explain various forms of arrangement of items based on item difficulty indexes usually in multiple choice tests. In this regard, Opara and Uwah (2017) defined item sequencing as the various formats through which tests come. They further stated that it is a way classroom teachers and professional test constructors and examination bodies arrange their test items. Inferring from the assertion of Opara and Uwah, teachers, test constructors as well as examination bodies or agencies use the various item sequencing formats to arrange test items. This item sequencing is determined by finding the item difficulty indexes of each and every item and afterwards arranging the indexes to suit a particular

form of arrangement or format. Talking about the importance of item sequencing or arrangement, Baxter (1998) noted that, test item arrangement plays a vital role in determining the performance of students in examination.

The various forms of item sequencing or arrangement according to Opara and Uwah are the Easy-To-Hard (ETH), Hard-To-Easy (THE) and Random (RDM). In explaining the various forms of item sequencing or arrangement, the ETH implies arranging test items from easy or simple to hard or complex. This mean that items are sequenced or arranged in such a way that the easier questions begin and ends with hard or difficult questions. Highlighting the importance of arranging items from ETH, Tei-Firstman (2008) noted that presenting students with simple question first will help sustain their interest and zeal towards approaching more questions. In addition, Balch (1989), Carlson and Ostrosky (1992) and Stout and Wygal (1992) supporting the assertion that students perform significantly well on ETH sequence than the other forms of item sequencing, they noted that ETH sequence gives advantage to students which enables them to score high scores. This however means that when the ETH sequence of items are used, students stand the chance of performing well in the examination. Jordan (1953) and Leary and Dorans (1985) are also of the view that students significantly perform better when items are organised from easiest to hardest. ETH item sequence is also believed to give students some level of motivation to proceed answering the items and in view of that Gronlund (1985) posited that "Except for a few items at the beginning of the test, for motivational purposes, none of our items should be so easy that everyone answers it correctly" (p.137).

Another form of the item sequencing is the Hard-To-Easy (HTE). In this form of sequence, students are presented with test whose item difficulty indexes range from lower to higher. In other words, the HTE is a form of arrangement or sequence where items are sequenced from hard to easy. Considering why test items should be arranged from HTE, Skinner (1999) noted that students may actually perform better if tests begin with difficult questions and students are given immediate feedback. Supporting the assertion of Skinner, Munz and Jacobs (1971) as cited in Ollennu (2015) also added that, considering the various categories of item order arranged by difficulty, the hard-to easy item order encourages examinees to make better achievements. This however, create inconsistencies as to whether the ETH or HTE is the best form of item sequence for teachers and assessment agencies to use in order to obtain true reflection of students' abilities.

In addition, random (RDM) arrangement of item is also a form of item sequencing or arrangement where the difficult indexes of items is scrambled. In other words, items in this form of arrangement is not either sequenced from ETH or HTE but randomly arranged based on their difficulty levels. Literature on this particular form of item sequence is even though scarce, but the findings derived from the studies conducted revealed that the random arrangement of items has no significant effect on students' responses to the items which can eventually affect their performance. For instance, Peters and Messier (1970) conducted a study on the effect of random arrangement multiple-choice tests on their performance using a class of graduate students studying research methods. At the end of the study, their findings revealed that no differences in performance on random question order multiple-choice tests in a class of

graduate students studying research methods. This however means that random sequence or arrangement of items could not predict students' performance.

The findings of Peters and Messier (1970) however contradicted the findings of Taub and Bell (1975) who considered the positioning of test questions and concluded that a truly random arrangement of questions results in lower examination scores than questions which were ordered to follow the topics taught at lectures.

It is evident from the above literature that, item sequencing remains one of the important areas in the construction of multiple-choice tests which teachers as well as assessment agencies need not to compromise if their main aim is to obtain the true reflection of students' traits.

The Concept of Academic Performance in Core mathematics and Integrated Science

Academic performance is one of the concerns mostly discussed by stakeholders of education. In our school setting, academic performance is believed to have a lot of predictors or factors that affect it. Academic performance as a concept has been defined by numerous authors with different perspectives. For instance, Academic performance according to Dimbisso (2009) is viewed as a successful attainment or accomplishment in a specific subject area which include the assignment of grades, marks and scores aimed at describing a trait possessed by students. This basically means that one's academic performance is measured by the person's attainment in a particular study area through assignment of scores in order to describe the person's characteristics. Also, as stipulated by Ferla, Martin and Yonghong (2009), academic performance depicts the perception and knowledge someone has with

regards to his or her academic ability in performing successfully a given academic task at a designated level. Ferla et al. basically explains that in a given task, one's thinking regarding his or her academic ability to those tasks successfully at a designated level is academic performance. In addition, academic performance according to Ankomah (2011) refers to the measure of students' output at the end of series of assessment tasks. In a similar definition given by Tetteh (2011), he noted that, academic performance is seen as the reach of gain by effort or accomplishment of one's goal. The definition given by Tetteh is not far different from that of Danso (2011) who also sees academic performance as the series of action of a person on a learning task.

In education, sometimes performance is mostly presented as synonymous to achievement or attainment. However, the two concepts are different. Academic achievement can be defined as the process in which students' show their ability to pursue tasks. In other words, what a student is able to achieve when he or she is tested on what has been taught usually in formal education. Thus, at the end of a period of carefully planned mode of instruction, students may have to be assessed after having undergone such series of instruction in order to enable educators measure and categorize the output of the students and eventually establish their performance. Performance on the other hand is referred to as the one requiring student to demonstrate skills and competencies by performing or producing something (Wisconsin Education Association Council, 1996). Researchers through their finding have revealed that, in vitualising academic performance, one has to segregate between an action (i.e., behavioural) aspect and an outcome aspect of performance (Campbell, McCloy, Oppler, & Sager, 1993; Roe, 1999). The behavioural

aspect of performance refers to what students can portray or do in the task situation. It encompasses student behaviours such as attending to classes, performing class assignment, presenting home work on time and any other school activities as well as teachers teaching basic reading skills to school children (Kanfer, 1990). Considering what students can do, Motowidlo, Borman and Schmit (1997) are of the view that, actions which can be measured are considered to represent performance.

Linking the assertion of Kanfer (1990) to the performance of SHS students in Ghana on the behavioural aspect of performance explaining what students can do at the end of a given task, it is evident that what students are able to produce in the core subjects of WASSCE for the past four years has generally not been encouraging. This evidence is provided in the statistics of their performance in the core subjects in the statement of the problem. The results of students' performance in core subjects shows inconsistencies with Core Mathematics and Integrated science emerging the lowly performed subjects (EMIS, 2019). This situation of low performance in Ghana with regards to the core subjects is surprisingly not different from the performance of SHS students in the study area (Kumasi Metropolis) within the same four-year period. This low performance in Kumasi Metropolis according to Asamoah (2018) is attributed to factors coming from the home, school, students and teachers. However, very little attention is given to the form of item sequencing in the multiple-choice tests of the WASSCE questions that the students answered as a factor causing changes in their performance. This is supported by the assertion of Shepard (1994) that tiny changes with respect to the sequence of items in a Multiple-Choice Test (MCT) can also make large difference in

students' performance. It is in this regard that the study sought to confirm or refute the position of the author to serve as an empirical evidence on whether items sequencing in multiple-choice tests in WASSCE could have caused the changes in their performance especially in Mathematics and Science.

Conceptual Framework

The conceptual framework of the study was designed based on the concepts as well as the variables derived from the empirical studies reviewed. The rationale was to clarify the relationship between the independent variables (item sequencing with three levels (Easy-To-Hard, Hard-To-Easy and Random) and the dependent variable (students' academic performance in Core Mathematics and Integrated Science). The relationships between the individual variables are shown in Figure 1. Figure 1 depicts the relationship between the independent variables and the dependent variable. It is evident that the independent variables which originate from the way test items are constructed are directly related to academic performance of students in Core Mathematics and Integrated Science with gender moderating the relationship between the independent and dependent variables. It can also be inferred from Figure 1 that gender influences the sequence of items in MCT. The study seeks to clarify these relationships among the study variables in the study area.

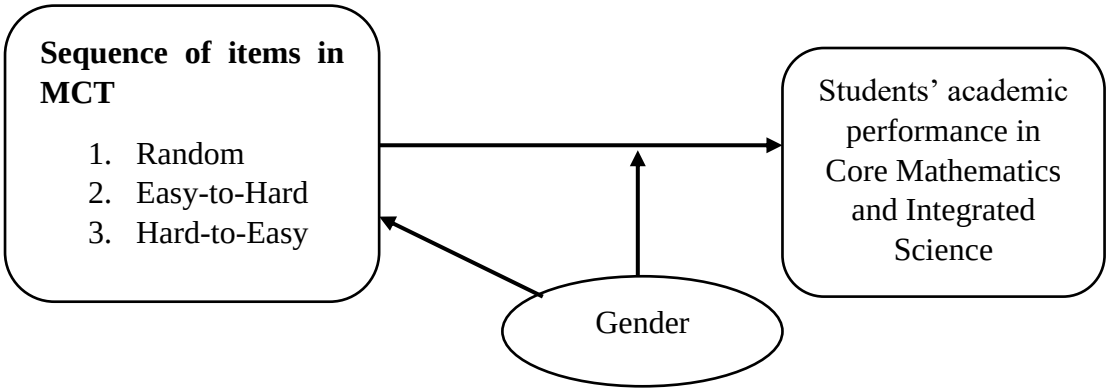


Figure 1: A conceptual framework on the effect of item sequencing in multiple-choice test on students' academic performance and the moderating role of gender

Source: Author's own construct, 2019.

Empirical Review

Several studies or works have been carried out internationally on the effect of item sequencing or arrangement on students' academic performance in different subject areas. In addition, quite a number of works have also been done on the moderating role of gender in item sequencing.

The Effect of Item Sequencing on Students' Academic Performance

Opara and Uwah (2017) investigated the effect of test item arrangement on performance in Mathematics among Junior Secondary School Students in Obio-Akpor L.G.A of Rivers State Nigeria. In the attempt to investigate the above variables, they used the quasi-experimental design. A sample of 100 Junior Secondary Schools were drawn using simple random and proportionate sampling out of a target population of 6.777 students. Students who took part of the study were given Mathematics Performance Test (MPT) to respond to. The instrument had four types (type A-D). Type A of MPT contained 40 items arranged in ascending order of difficulty, Type B equally contained 40 items

arranged in descending order of difficulty. Type C was arranged in order of topic presentation while type D was arranged in no particular order. Type A was administered to 25 students in the first school likewise Type B and C. Type D was administered to the control group. In other words, the test instrument was administered to three experimental groups and one control group through the help of a research assistant and the classroom teachers. Mean, standard deviation as well as t-test analysis were used to compare the mean of each of the group against the control group and test. The findings of the study were that item arrangement based on ascending order of difficulty has a positive and significant effect on students' academic performance in mathematics at 0.05 alpha level while item arrangement based on descending order has a positive but insignificant effect on student' performance in mathematics. Finally, item arrangement based on no particular order of difficulty has a positive and significant effect on students' academic performance. It is evident from the findings of Opara and Uwah that item sequencing or arrangement based on ascending and descending order were positive and significantly affected students' academic performance which means that the ascending and the descending orders can have effect on students' academic performance. The random or no order arrangement was also positive but not significant which means that arranging items in random order had no effect on students' academic performance.

Cobbinah (2016) also conducted a similar study in Ghana where he investigated items sequencing on difficulty level and students' achievement in Mathematics test in Cape Coast Metropolis of Central Region, Ghana. He made an attempt to find out whether item difficulty index and item ordering of the

W.A.E.C multiple choice test items has influence on Senior High Students achievement in Core Mathematics. The quasi-experimental design was adopted for the study and the purposive sampling technique was used to select six (6) schools which comprised 250 form two students from six intact classes in the Central Region of Ghana. Forty Core Mathematics multiple-choice items were adapted from W.A.E.C past questions. Cobbinah also arranged or sequenced the items based on the item difficulty index where he subjected the responses of the students to a two-way analysis of variance. The reliability of the instrument was established by the use of Kuder-Richardson formula 20 and it was 0.78.

The findings of Cobbinah revealed that the change in item order or arrangement on Senior High School students' achievement in Core Mathematics was significant. This means that item sequencing has effect on students' achievement. His findings were consistent with the findings of Opara and Uwah that item sequencing or arrangement can affect students' academic performance with respect to Mathematics as a subject.

Again, Ollennu and Etsey (2015) also conducted a similar study in Ghana, their focus was to investigate the impact of item position (arrangement) in multiple-choice test on student performance at the Basic Education Certificate Examination (BECE) level. The sample consisted of 810 JSS 3 students selected from 12 different schools. A quasi-experimental design was used. Students for the study were drawn from schools representing public and private schools in urban and rural areas. A multiple-choice test consisting of 40 items in each of English Language, Mathematics and Science was used. The items were arranged or sequenced using the difficulty order to obtain the three treatments i.e. Random (RDM), Easy-to-Hard (ETH) and Hard-to-Easy (HTE).

The data collected were subjected to statistical analysis using ANOVA at $p \leq 0.05$.

The results of the analysis indicated that for English Language, Mathematics and Science at the BECE level, when item order was altered, the difference in performance was generally significant. However, there was no significant difference in performance between the ETH and HTE treatments of English Language. Similarly, there was no significant difference in performance between the RDM and ETH, and RDM and HTE of Science. The results for Mathematics were consistent, in that significant difference in performance was observed between all the three treatments for the subject.

Alakayleh (2017) investigated if differences in item arrangements can affect 10th grade students' academic performance in multiple-choice tests in Mathematics, Science, and English language final exams in Jordan. The study sample consisted of 764 students selected from twenty regular government schools. The study adopted quasi- experimental design and the study instrument used was a multiple-choice type of 48 items for Mathematics, Science and English. The test items were arranged according to their difficulty for three processors (RDM), Easy -To- Hard (ETH) and Hard-To-Easy (HTE). The data collected were analyzed statistically using the analysis of mono-variance at the statistical significance level of 0.01.

The results of the analysis indicated that for Mathematics, Sience and English courses, and in the order of items, the change in performance was statistically and morally significant, and it was found that the proposal to rearrange the test items for final achievement tests to control the penetration and misbehavior of the test rules and controls may not be optimal, The study of the

English language, which did not show significant differences and suggests the use of other methods such as random model and parallel and split- half tests and further studies of other levels of study.

The findings of the above authors however contradict the findings of some other authors. For instance, Abbasian and Zadsar (2019) conducted a study on the relationship between test item arrangements or sequence and testee's performance and test usefulness criteria. Their study shed lights on the issues like test item arrangements in relation to test taker's performance, test usefulness criteria including test Validity, Reliability, Impact, Interactiveness, Authenticity, and Practicality, and also test characteristics such as Item Facility (IF), Item Discrimination (ID), and Choice Distribution (CD). The ex-post facto design study was conducted with a group of Iranian EFL learners. For the purpose of the study, they randomly selected some university students majoring in English Language Teaching, English Translation Studies, and English Literature and learners from different language institutes and gave them a version of the Nelson proficiency test. They later administered two different versions of the same researcher-made test in terms of item arrangements to 116 students obtaining an acceptable score in the Nelson proficiency test. Respective statistical analyses revealed contradictorily that test item arrangement did not have any significant effect on the performance of the test takers. In addition, their findings with respect to the first null-hypothesis (Do test item arrangements have any significant effect on testee's performance on a test?) postulated that the arrangement of items did not have any significant effect on the examinees' performance on the three tests which means that when

the arrangement or the sequence of the items were altered, pupils performance remained the same.

Gender Difference in Students' Academic Performance in Item Sequencing in Multiple-Choice Test (MCT)

Cobbinah (2016) investigated into items' sequencing on difficulty level and students' achievement in Mathematics test in Cape Coast Metropolis of Central Region, Ghana where his main focus was to find out whether item difficulty index and item ordering of the W.A.E.C multiple choice test items has influence on Senior High Students achievement in Core Mathematics. The study adopted the quasi-experimental design where purposive sampling technique was used to select six (6) schools comprised 250 form two students from six intact classes in the Central Region of Ghana. Cobbinah's main instrument was a Forty Core Mathematics multiple-choice items which was adapted from W.A.E.C past questions. The findings of the study established among others that the effect of change of item order on Senior High School students' achievement in Core Mathematics was significant. His findings indicated that differences in the item sequencing achievement of students with respect to gender was significant which proved that gender has role to play as far student achievement in Mathematics and the order of items are concern. The findings further indicated that male students generally performed better no matter the order of items than their female counterpart.

Similarly, in Nigeria, The Research Division of WAEC, Lagos (1993) investigated the effect of item position on performance in multiple-choice objective tests at the level of the West African Senior School Certificate Examination (WASSCE). The subjects used were Agricultural Science,

Biology, Economics, English Language and Mathematics. A forty-five multiple-choice objective test was developed in each of the selected subjects. The test consisted of three sub-tests indicated as sections in the question paper with each section having fifteen items. The items were arranged to reflect the following order:

(i) Easy-to-Hard

(ii) Hard-to-Easy and

(iii) Random order

The research sample consisted of one thousand, one hundred and twelve (1,112) SSS form 2 students randomly selected from 35 schools in eight states in Nigeria. The data collected were subjected to statistical analysis using ANOVA and t-test to determine if any significant differences existed in the performance of students in the three arrangements of items and the pattern of performance among the five subjects. As expected, there were significant differences in the performance of students. Easy-to-Hard arrangement of items was the most effective in enhancing the performance of students in English Language and Biology. Again, the Hard-to-Easy arrangement was significantly the most effective in Mathematics and Economics while the Random Order was effective in Agricultural Science, among the three arrangements.

In a follow-up research, the Research Division of WAEC, Lagos (1993), investigated the effect of sex, ability group and school type on students' academic performance in the three formats of arranging multiple-choice test items were investigated. In the findings, it was reported that the general performance of students in the three arrangements was independent of sex except Hard-to-Easy in Mathematics and Random Order in English Language.

This indicates that irrespective of the arrangement adopted in Economics, Biology and Agricultural Science, the pattern of performance was influenced by gender. They further noticed that the Hard-to-Easy arrangement in Mathematics disadvantaged the female candidates while the Random-Order in English Language seem to have reverse effect though it is assumed that the occurrence is by chance.

Chapter Summary

Looking at the empirical studies reviewed, it can be observed that a number of studies have been conducted internationally on item sequencing in multiple-choice tests and students' academic performance. From the empirical review, it can be observed that little information can be obtained locally on the topic. However, majority of the studies conducted especially in the western world on the topic focused on the effect of item sequencing in multiple-choice tests and students' academic performance in various subjects to the neglect of the moderating role of gender in item sequencing in multiple-choice tests and students' academic performance. It can also be observed that, no empirical studies were reviewed on the moderating role of gender in item sequencing since it appears no study has been conducted on it.

Referring to the low academic performance of Senior High School students in Core Mathematics and Integrated Science in the Kumasi Metropolis and the research gap in literature on the topic, there is the need to find evidence of the effect of item sequencing in multiple-choice tests on students' academic performance and also find out if gender moderates the relationship between item sequencing and students' academic performance.

CHAPTER THREE

RESEARCH METHODS

Overview

The methodology section describes the research design and the study area, the population, sample and sampling procedures, data collection instrument, data collection procedure and data analysis procedure. Ethical issues that were considered in the study were also discussed.

Research Design

According to Sarantakos (2005), a research design is the prescription of the logical sequence in which a study is to be carried out, as well as elements of the study, its methods of data collection and analysis and administrative procedures that need to be considered for the study to be carried out without problems or delays. Amedahe and Asamoah-Gyimah (2018) posit that research design is the researcher's overall plan for obtaining answers to the research questions or for testing the research hypothesis. They further noted that research design refers to the blueprint that specifies how data relating to a given problem should be collected and analyzed. The study adopted the quasi-experimental research design. The quasi-experimental design is one of the types of experimental designs. An experimental design as explained by Amedahe and Asamoah-Gyimah (2018) refers to a design that presents a situation in which a researcher objectively observes a phenomenon which are made to occur in a strictly controlled situation where one or more variables are varied and the others are kept constant. This means that in an experimental study, there is a constant/control group and experimental group where the effectiveness of an intervention is checked on the experimental group.

As a component of the experimental design, Ary, Jacobs and Razavieh (2002) explain that quasi-experimental design is applicable in situations whereby subjects cannot completely be assigned to groups. Amedahe and Asamoah-Gyimah (2018) also noted that quasi-experimental design are used when it is impossible to randomly assign participants to comparison groups and when a researcher is faced with a situation where only one or two participants can participate in the research study. They further stated that quasi-experimental design can be used when it is impossible to control for all potentially confounding variables in most cases because the participants cannot randomly be assigned to the groups. Again, they noted that the quasi-experimental design is so called because it does not include all features of a true experiment as a complete experimental control is not always possible, especially concerning the randomization and application of intervention. The results of quasi-experimental research do not have the same validity as those obtained in true experiments since the absence of randomization of the subjects in the experiment and control groups cannot guarantee equivalence between the groups at baseline (Amedahe & Asamoah-Gyimah, 2018). This is supported by assertion of Cook and Campbell (2001) that the prefix quasi means "resembling". This means that, the quasi-experimental research looks like an experimental research but not true experimental research. In quasi-experimental research, even though the independent variable is manipulated, participants are not assigned randomly to conditions or odder of conditions.

The study used the quasi-experimental design because there is need to find out if the sequence of item can predict students' academic performance using an experimental and control group where an intervention was given to the

experimental group to check the effectiveness of the intervention. In this regard, the various item sequence (HTE, ETH and RDM) were the interventions which were administered to the experimental group to check their effect on students' academic performance. Additionally, the study adopted this design because it was difficult to assemble all the testees at one place and randomly assign them into control group and experimental groups which is a strength of the quasi-experimental design. More so, the authorities in the schools used (Kumasi High SHS, Opoku Ware SHS, Prempeh College, Yaa Asantewaah Girls SHS, St. Louis SHS and T.I Ahmadiyya SHS) did not allow the already grouped or intact classes to be disrupted by assigning them into experimental and control group therefore the quasi-experimental design which does not use randomization was appropriate. The quasi-experimental research design was appropriate for the study since the researcher could not possibly control all the confounding variables that might have rendered the study's results invalid.

Any design adopted for a study is supported by a particular research paradigm. Therefore, the quasi-experimental research design is supported by the positivist paradigm which under the objectivity epistemology is the main methodological philosophy in quantitative research (Crotty, 1998). According to Cohen, Manion and Marison (2011), in explaining the positive paradigm, in the process of studying the phenomenon, the relationship between an independent variable and one or more dependent variables will be discovered by causal inferences which is as a result of the experimental designs and be fully be determined by how researchers maximise the influence of the independent variables on the dependent variables using this process. Even though, the quasi-experimental design was appropriate for the study, it lacks randomization of

participants into control and experimental groups which makes it impossible to control for all potentially confounding variables. Nevertheless, since the randomization of participants was beyond my control, the quasi-experimental design which does not allow for randomization was appropriate.

Study Area

The study was conducted in the Kumasi Metropolis in the Ashanti Region. According to Ghana Statistical Service (2014), Kumasi Metropolitan Assembly (KMA) is known to be the second largest most populous city in Ghana, following Accra; the national capital. Its strategic location has also endowed it with the status of the principal inland transport terminal, thus giving it an essential role in the vast and profitable business of the distribution of goods in Ghana and beyond to other West African countries. The relationship between Kumasi and its adjoining districts are interdependent. Whiles these adjoining districts provides farm produce for the Metropolis, Kumasi provides marketing avenue for these produce to traders not only from other parts of Ghana but to those from those from the West African sub-region. In view of the fact that Kumasi provides an avenue for marketing farm produce, it is also seen to be a place where traders meet for commercial activities. Additionally, the Metropolis provides services related to healthcare, education, inter-city transport, financial services and wholesale and retail trading among others to traders from the adjoining districts and beyond. The population of Kumasi Metropolis is 1,730,249 which represents 36.2 percent of the total population of Ashanti Region, from the 2010 Population and Housing Census. The Metropolis has a sex ratio of 91.4. The population of the Metropolis depicts a broad base population pyramid which tapers off with a small number of elderly persons (60

years and older). The age dependency ratio for the Metropolis is 59.9 for males and 56.3 for females. The percentage of population aged 11 years and older who are literate is 89.5 percent while 10.5 percent are not literate. Among the literates, 72 percent can read and write in both English and a Ghanaian language. About 9.1 percent of the population 3 years and older have never attended school while 40.5 percent are currently in school and 50.4 percent have attended school in the past. Kumasi Metropolis is one of the thirty (30) districts in Ashanti Region which is located between 6.35 ºN and 6.40 ºS and longitude 1.30 ºW and 1.35ºE and elevated 250 to 300 meters above seal level. The also shares boundaries with Kwabre East and Afigya Kwabre Districts to the north, Atwima Kwanwoma and Atwima Nwabiagya Districts to the west, Asokore Mampong and Ejisu-Juaben Municipality to the east and Bosomtwe Districts to the south. It about 270km north of the national capital, Accra. It has a surface area of nearly 214.3 square kilometers which is about 0.9 percent of the region's land area. However, it houses about 36.2 percent of the region's populations.

Population

Nitko (2004) defined a population as the entire aggregation of cases that meet a designated set of criteria. Bryman (2001) defined population as any set of persons or subjects that possess at least one common characteristic. The target population comprised all public Senior High School (SHS) Form two students in the 22 SHS public schools within the Kumasi Metropolis in the Ashanti region of Ghana with a total population of 29,960 (Educational Management Information System, 2019). Out of the 22 public SHS, there are 4 single sex boys' schools, 5 single sex girls' schools and 13 mixed schools. The breakdown of the schools is seen in Table 2.

Table 2: Distribution of SHS Schools in The Kumasi Metropolis

Schools' Status	Name of School	SHS form 2 Enrollment		
		Male	Female	Total
Boys	1. Kumasi High School SHS	1356		1356
	2. Opoku Ware SHS	1823		1823
	3. Prempeh College SHS	1984		1984
	4. St. Hubert Seminary Secondary Sch.	502		502
Girls	1. Kumasi Girls SHS		1241	1241
	2. St. Louis SHS		1331	1331
	3. Kumasi Wesley Girls SHS		1320	1320
	4. Serwaa Nyarko Girls SHS		789	789
	5. Yaa Asantewaah Girls SHS		1256	1256
Mixed	1. Adventist SHS	666	688	1354
	2. Anglican Secondary School	1035	775	1810
	3. Armed Forces Sec/Tech. School	795	461	1256
	4. Asanteman SHS	886	726	1612
	5. Kumasi Sec/Tech School	1771	1145	2916
	6. Osei-Kyeretwie SHS	906	706	1612
	7. Al-Azhariya Islamic SHS	293	352	645
	8. T.I Ahmadiyya SHS	1189	1041	2230
	9. KNUST SHS	667	598	1265
	10. Agric-Nzema Comm. SCH	171	165	336
	11. Islamic SHS	1219	1484	2703
	12. Pentecost SHS	298	125	423
	13. Uthmaniya Islamic SHS	105	87	192
Total		15,670	14,290	29,960

Source: EMIS (2016-2019); Ashanti Regional Educational Office (2019)

However, the accessible population comprised all SHS form 2 students in all the 6 category "A" public SHS in the Kumasi Metropolis. The study did not used Form 1 and 3 students because, as at the data collection period, Form 1 students have not covered much topics which could have affected the results of the study. Furthermore, the Form 3 students were also seriously preparing to write WASSCE and so their study habit could have affected the results of the study since the study used WASSCE past questions. The Form 2 students were

appropriate for the study since they have been on campus for quite a while and might have covered more of the SHS topics, as potential WASSCE candidates. Finally, only category A schools were used for the study because it is assumed that these schools have same or similar characteristics such school facilities, qualification of teachers, WASSCE performance, amongst others which may not have influence in the experimental process. The total population of the Form 2 students in all the category "A" schools was 9, 980 as at the year 2019. The breakdown of these schools with their respective SHS Form 2 students' enrollment is seen in Table 3.

Table 3: Distribution of Category "A" Schools in Kumasi Metropolis

S/N	School	Status	SHS Form two (2) Enrollment		
			Male	Female	Total
1	Kumasi High SHS	Boys	1356		1356
2	Opoku Ware SHS	Boys	1823		1823
3	Prempeh College	Boys	1984		1984
4	Yaa Asantewaah Girls SHS	Girls		1256	1256
5	St. Louis SHS	Girls		1331	1331
6	T.I Ahmadiyya SHS	Mixed	1189	1041	2230
Total			6352	3628	9,980

Source: EMIS (2016-2019); Ashanti Regional Educational Office (2019)

Sample and Sampling Procedure

According to Creswell (2014), a sample refers to a sub-group of the population that is studied in order to make a generalisation regarding the target population. The multi-stage sampling procedure was used to select participants for the experimental and control group. It must be pointed out that, the 6 category A schools were all used. In view of that, three schools were used in

each of the experimental and control groups. The selection of schools for experimental group was done first, followed by the control group. The various stages adopted in using the multi-stage sampling procedures have been described as follows:

Stage 1

In an attempt to select students from the six category "A" schools, first of all, the stratified sampling technique was used and all single sex boys' school were put in one stratum, all single sex girls' schools were also put in another stratum and finally all mixed schools also in another stratum forming first stratum, second and third stratum respectively in the experimental group. For instance, the first stratum consisted of Prempeh College, Opoku Ware SHS, Kumasi High SHS and all the names of these schools formed the sampling frame for the first stratum. The same procedure was repeated for the single sex girls' schools as well as the mixed school.

Stage 2

After putting the schools in strata, schools within each stratum were written on pieces of papers and the lottery method of the simple random sampling technique was used to select one school from each stratum because each stratum represents a homogeneous group of students, that is, a group with similar characteristics (Ary, Jacobs, Razavieh & Sorensen, 2006). Selection of school in each stratum was done from one stratum to another in order to ensure uniformity. The mixed school (T.I Ahmadiyya SHS) was automatically selected since there was only one mixed school among the category "A" schools. In the single sex boys' stratum, Prempeh College was selected while Yaa Asantewaah Girls SHS was also selected from the single sex girls' schools. Therefore, the

three selected schools (Prempeh College, Yaa Asantewaah Girls SHS and T.I Ahmadiyya SHS) formed the experimental group with a population of 5,470 as at the year 2019 during data collection. The sample distribution of form 2 students in the experimental group can be seen in Table 4

Table 4: Distribution of Selected Schools in The Experimental Group

S/N	School	Status	SHS form 2 enrollments		
			Male	Female	Total
1	Prempeh College	Boys	1,984		1,984
2	Yaa Asantewaah Girls SHS	Girls		1,256	1,256
3	T.I Ahmadiyya SHS	Mixed	1,189	1,041	2,230
Total			3,173	2,372	5,470

Source: Field survey (2019)

The sample size for three school was determined in accordance with Krecjie and Morgan's (1970) table for determining sample size and according to them a sample of 5,470 subjects corresponds to 357 students for the study. Therefore, 357 students were used in the experimental group for the study.

In selecting sample for the control group, the three unselected Category "A" schools were automatically used (Kumasi High SHS, Opoku Ware SHS and St. Louis SHS). The total number of SHS form 2 students in these schools was 4,510. The sample distribution of form 2 students in the control group can be seen in Table 5.

Table 5: Distribution of Selected Schools in The Control Group

S/N	School	Status	SHS form 2 enrollments		
			Male	Female	Total
1	Kumasi High SHS	Boys	1,356		1,356
2	Opoku Ware SHS	Boys	1,331		1,331
3	St. Louis SHS	Girls		1823	1823
Total			2,687	1,823	4,510

Source: Field survey (2019)

In the same way, the Krecjie and Morgan's (1970) table for determining sample size was used to determine the sample size and a population of 4, 510 subjects correspond to 357 students for the study. Therefore, 357 students were also used in the control group.

Stage 3

At this stage, proportionate stratified sampling technique was used to distribute 357 students over the three schools in the experimental group. The same procedure was also used to distribute 357 students over the schools in the control group. The proportionate sample for each school was calculated by dividing the number of SHS two students in that school by the total population of the schools multiplied by the sample size of 357. For instance, the sample size of Yaa Asantewaah Girls SHS which was in the experimental group was calculated by $\frac{1,256}{5,470} \times 357 = 82$ and the sample size of Opoku Ware SHS in the control group was also calculated by $\frac{1,331}{4,510} \times 357 = 105$. The proportionate sample for each of the three schools in the experimental and control group can be seen in the Table 6 and 7 respectively.

Table 6: Distribution of Samples for the Three Schools in the Experimental Group

School	Population (P)	Sample (S)
Prempeh College	1,984	129
Yaa Asantewaah Girls SHS	1,256	82
T.I Ahmadiyya SHS	2,230	146
Total	5,470	357

Source: Field survey (2019)

Table 7: Distribution of Samples for the Three Schools in the Control Group

School	Population (P)	Sample (S)
Kumasi High SHS	1,356	107
Opoku Ware SHS	1,331	105
St. Louis SHS	1,823	144
Total	4,510	357

Source: Field survey (2019)

Moreover, in each school, purposive sampling technique was used to selected an intact class (General Science Class) before the data collection. This is because the use of quasi-experimental design makes use of intact classes. In Prempeh College, T.I Ahmadiyya SHS and St. Louis SHS, three intact sciences classes were used whiles two intact science classes were also used in Yaa Asantewaah Girls SHS, Kumasi High SHS and Opoku Ware SHS. This was as a result of the sample size for each school. Schools where more than 100 students were needed for the study (Prempeh College, T.I Ahmadiyya SHS and St. Louis SHS), three intact classes were used whiles two intact classes were

used in schools with a sample size of close to 100 (Yaa Asantewaah Girls SHS) or a little above 100 (Kumasi High SHS and Opoku Ware SHS). On the whole, the sample size used in the study was 714 students which comprised 357 from the experimental group and another 357 from the control group. In conclusion, the study used multi-stage sampling procedures which included stratified sampling technique, simple random sampling technique, proportionate sampling technique and purposive sampling technique for selecting samples for each of the experimental and control group.

Data Collection Instruments

A standardized WASSCE Core Mathematics and Integrated Science past questions developed by WAEC were adapted for the study. The adaptation of the instrument here means that even though the questions were constructed by WAEC, items were selected based on the common topics that the students were taught within the past questions from the year 2016 to 2019. Moreover, after selecting the questions or items to reflect the topics that students have learnt to ensure content representativeness, the items (Random Sequence) were piloted using Category "A" schools (St. Augustine's College and Holy Child School) in Cape Coast Metropolis and based on the responses, it was sequenced from Hard-To-Easy (HTE) and Easy-To-Hard (ETH) using the item difficulty (See Appendix K and L) of the test for the students in the study area to answer. The item difficulty of the items was judged using guidelines suggested by Allen and Yen (2002). According to Allen and Yen (2002), "item difficulties of about .30 to .70 maximizes information the test provides about differences among examinees" (p. 121). This means that if an item records a difficulty index of above .70, the item is described as "easy", for an item to be moderately

difficulty, the difficulty index should range from .30 to .70 while a difficult item records a difficulty index of below .30. Judging the quality of the items using the item difficulty was prudent since items that needed modification were detected and dealt with accordingly. In each subject, examinees were expected to answer all 50 questions and they had 1hour 15 minutes to do so.

Pilot Testing

The instruments were pilot tested using category "A" schools in the Cape Coast Metropolis. There are 6 category A schools which comprised only single sex boys' and single sex girls' schools in the metropolis namely; Adisadel College, St. Augustine's College, Mfantsipim School, Holy Child School, Wesley Girls Senior High and Mfantsiman Girls Senior High. In selecting participants for the pilot testing, the stratified sampling technique was used and all single sex boys' school were put in one stratum and all single sex girls' schools in another stratum. After putting the schools in strata, the names of schools within each stratum were written on pieces of papers and the lottery method of the simple random sampling technique was used to select one school from each stratum. In the singles sex boys' stratum, St. Augustine's College was selected whiles the Holy Child School was also selected in the single sex girls' schools. The piloting was done using the Category "A" schools since their ability level and performance matches the Category "A" schools in the Kumasi Metropolis. In addition, the Cape Coast Metropolis was selected because of proximity.

Before the pilot testing, a letter was taken from the Head of Department of Education and Psychology, University of Cape Coast. Copies of the letters were sent to the authorities of the schools to seek for permission to conduct the

pilot-testing. The students were then brief and 70 students representing 20% of the sample for the main study were used as recommended by Asamoah-Gyimah and Anane (2019). It must be noted that 35 form two students (intact Science Class) representing 10% were selected each from the single sex schools. In St. Augustine's College as well as Holy Child School, the simple random sampling technique was used to select the 35 form two students (intact Science Class) from each school. Intact Science Classes were used in each of the schools because the use of quasi-experimental design did not allow the researcher to use a sample which comprised students from different classes. In addition, quasi-experimental design advocates for the use of intact classes.

Afterwards, 50 items which were selected based on the topics the students have been taught by their teachers from WASSCE Core Mathematics and Integrated Science past questions developed by WAEC from the year 2016-2019 were given to them to answer. These topics were obtained from the Core Mathematics and Integrated Science teachers in each of the schools. It must be noted that, within the topics, the common topics studied by the students in the two schools were used to sample from the past questions.

After the pilot-testing, the items were then subjected to item analysis. From the results of the analysis, items which needed modification were re-fixed, although these were very few. The difficulty levels of the items were also ascertained (See Appendix K and L) and based on that the items were arranged based on the forms of sequencing we have. Thus, one from Easy-To-Hard (ETH) and another from Hard-To-Easy (THE), while keeping the Random order intact. Arranging the items in order of difficulty was straightforward for both subjects since they both had only one section each.

Validity and Reliability of the Instruments

In ensuring the validity of research instruments, Gall, Borg and Gall (2003) indicate that validating of an instrument is improved through expert judgment. Reliability of research instrument however, refers to the degree to which the instrument would measure consistently a characteristic when applied more than once to the same person(s) under similar conditions (Nitko & Brookhart, 2007). The research instruments adapted for the study were WAEC past questions from the year 2016 to 2019. It is then assumed that standardized items from WAEC are of good psychometric properties including the validity and reliability even though the items were selected to reflect the topics students have learnt in class. This however means that the validity as well as the reliability of the items have been established already. In addition, the researcher gave the adapted items (Random sequence) to his supervisor who is an expert in assessment to scrutinize the items to ensure that the items are of good standard and free from ambiguity. Also, aside the established reliability of the instruments by WAEC, the researcher piloted the instruments with an independent sample to identify issues that might demand modification before distribution of the instruments to the research participants. The reliability of the test items (Random sequence) was calculated manually using KR20 and reliability coefficients of 0.81 and 0.74 for both the mathematics and science instruments respectively. The reliability coefficients of the Hard-To-Easy (HTE) sequence were 0.76 and 0.71 for Core Mathematics and Integrated Science respectively whiles that of the Easy-To-Hard (ETH) sequence were 0.78 and 0.73 for Core Mathematics and Integrated Science respectively. As indicated by Pallant (2010), a reliability coefficient of .060 or higher is

considered moderately appropriate. This indicates that the reliability coefficients of 0.81, 0.76 and 0.78 for Mathematics instruments and 0.74, 0.73 and 0.71 for Science instruments show that the items were appropriate in measuring the construct under consideration.

Ethical Consideration

Seidman (2006) defined research ethics refers to the correct rules of conduct necessary when carrying out research. He added that ethical consideration helps participants to understand the aims, objectives and potential harm that such involvement may have on them. This is supported by the assertion of Mertens (2010), who stated that informed consent arises from the participant's right to freedom. Punch (2008) is also of the view that researchers should be mindful of ethical issues especially in social research because it is concerned with data about people. Consideration of moral issues and respect for participants is essential in social research. Hence, in this research several ethical issues were taken into consideration.

Firstly, the right to privacy, voluntary participation, no harm to participants, anonymity and confidentiality were held in high esteem. It should be emphasised that students have right to privacy and as a result, these rights must be respected at every point in time. In this light, the rights of privacy of respondents in the study was respected and under no circumstances were respondents studied without their consent. In addition, one of the key components regarding ethical issues in research has to do with respondents' voluntary participation. Responding to items in the study of this nature demanded a lot of time and energy which can lead to the disruption of the regular activities of participants. It is for this motive that the objectives and

significance of the study were explained to the participants and therefore, allowed them to exercise their voluntary right in their participation of the study.

Another ethical issue in educational research has to do with the fact that, the exercise should not cause any injury to the participants under study irrespective of whether they volunteer or not to participate in the study. The concept of harm as used in this regard can be physical, psychological or emotional. In pursuance of this, the researcher made sure none of the questions posed physical, psychological or emotional harm to the students.

Furthermore, as part of the ethical issues in research, the ultimate goal is to protect and safeguard the well-being, interest and the identity of the participants. In pursuance of this, I adopted anonymity and confidentiality techniques such as non-disclosure of participants' names in ensuring the protection of participants. The participants were therefore assured that the information they provided in the form of their responses would be kept confidential. In addition, it must also be noted that, unethical behaviour which include plagiarism is not welcomed. This normally originates when a researcher falsifies, distorts data or plagiarises other peoples' works. In view of this, the researcher followed strictly the prescribed standard of scientific behaviour to avoid plagiarism. The researcher therefore gathered information from the right participants and subjected the information gathered in proper analyses before writing the research report. Notably, ideas, works and writings were duly acknowledged by way of providing appropriate references in the in-text referencing and the main referencing as necessary.

Data Collection Procedures

Before the start of the data collection exercise, an ethical clearance was obtained from the Ethical Review Board and the headmasters and headmistresses of the various Senior High Schools selected for the study were officially served with introductory letters from the Head of Department of Education and Psychology of the University of Cape Coast to sought for permission to conduct the study in their schools.

The researcher used eight research assistants and they were trained before the test administration. The training programme included explaining the objectives of the study to them as well as data management. It is worth noting that, each school in the experimental group took the three forms of sequencing thus Easy-to-Hard (ETH), Hard-to-Easy (HTE) or Random (RDM) whiles each school in the control group were given only Random (RDM) sequence to determine overall performance of the students. Upon arrival at the Senior High Schools selected for the study, the authorities of the various schools were informed about the study. Having sought the consent of the management of the Senior High Schools, students were selected using the multi-stage sampling procedures as already indicated. The purpose and focus of the study was then explained to the participants and the tests were given out to them to complete using the required number of minutes allotted for completing the WASSCE multiple choice-test as stipulated by WAEC.

The tests were administered and collected within a period of 8 weeks. The first two weeks were used to administer the Radom sequence of items in Core Mathematics items and Integrated Science items to the control group in order to determine their general performance. Afterwards, the experimental

group were given the various item sequence biweekly and therefore the second two weeks were used to administer the Easy-To-Hard (ETH) Mathematics items and the Hard-To-Easy (HTE) Integrated Science. The third two weeks were also used to administer Hard-To-Easy (HTE) Mathematics items and Easy-To-Hard (ETH) Integrated Science items. Finally, the last two weeks were used to administer Random (RDM) Integrated Science items and Random (RDM) Mathematics items since Random was a kind of item sequence on its own even though it was initially given to the control group to ascertain the general performance. This was done based on the assertion of Taub and Bell (1975) who pointed out that, test items are administered in random sequence to determine academic achievement/performance. The instruments were administered within a period of two weeks because according to Amedahe and Asamoah-Gyimah (2016), test-retest should be done in a period of two weeks in order to control for the effect of students getting exposed to the items.

Data Processing and Analysis

The data analysis phase consisted of editing, coding and statistical computation. Specifically, right after data collection, the items for each subject were scored and labelled serially to ensure easy identification, errors and easy coding. Frequencies were run to check for all errors such as outliers and missing values. The data gathered was then analysed with the aid of Statistical Packages for Social Sciences (SPSS, version 26) after the data had been collated and edited in order to address questions that were answered partially or not answered at all. It must be pointed out that, percentages and frequencies as well as tables were used to analyse the background information of the participants.

In analysing the research question which is in two folds and which sought to find out the effect of sequence of items in multiple-choice test on students' academic performance in Core Mathematics and Integrated Science, the responses to the items in both subjects were subjected to One-Way Repeated Measures Analysis of Variance.

For hypothesis one which sought to find out if there is statistically significant difference in students' academic performance in item sequencing with respect to gender in Core Mathematic and Integrated Science, the analysis was done separately using One-Way Multivariate Analysis of Variance (MANOVA) at $p \leq 0.05$ to determine if any significant differences existed in their performance in item sequence with respect to their gender. In addition, the dependent variable was students' academic performance in the various item sequence namely; Randomly (RDM), Easy-To-Hard (ETH) and Hard-To-Easy (HTE) and the independent variable was gender.

Lastly, the hypothesis two which sought to find out if gender significantly moderates the relationship between items sequencing in Multiple Choice Test (MCT) and students' academic performance in Core Mathematics and Integrated Science was also analysed separately for both subjects using process procedure by Hayes (2018).

Chapter Summary

This chapter discussed the methods and procedures that were used to accomplish the objectives of the study. A review of the research design, population and sample, data collection instruments, data collection and analysis procedures as well as validity and reliability of the instruments have been described. As indicated above, even though quasi-experimental research design

was most appropriate for the study, the design came with some limitations including not having the same validity as in true experiments because of the absence of randomization of the subjects in the experiment and control groups which cannot guarantee equivalence between the groups at baseline. Amidst the pitfalls of the design, since the study focused on the effect of item sequencing in multiple-choice tests on students' academic performance in Core Mathematics and Integrated Science, the design was deemed appropriate since it was difficult to assemble all the testees at one place and randomly assign them into control group and experimental groups.

CHAPTER FOUR

RESULTS AND DISCUSSION

Overview

The purpose of the study was to investigate the effect of item sequencing in multiple-choice test on Senior High School students' academic performance and further investigate the moderating role of gender in the relationship between item sequencing and students' academic performance. In achieving the purpose of the study, the quasi-experimental design was employed. In the data collection stage, WAEC past questions from the year 2016 to 2019 were adapted which were piloted to obtain the various item sequence forms for the students in the study area to answer. A One-Way repeated ANOVA, One-Way Multivariate Analysis of Variance (MANOVA) as well as the process procedure by Hayes (2018) were used to analyse the data. The sample size for the study was 714 which consisted of 357 students for the control group and 357 for the experimental group. The response rate after the instrument administration was 87% corresponding to 311 students for the experimental group and 85% corresponding to 302 for the control group. Therefore 613 students were used for analysis stage.

Results

Demographic Characteristics of the Senior High School Students

This section highlighted students' responses on their demographic characteristics specifically, their gender.

Gender of Students

The gender of students in the experimental and control group is presented in Table 8 and 9 respectively.

Table 8: Distribution of Gender of Students in the Experimental Group

Gender	Frequency	Percentages
Male	146	46.9
Female	165	53.1
Total	311	100

Source: Field survey (2019)

Table 9: Distribution of Gender of Students in the Control Group

Gender	Frequency	Percentages
Male	163	54.0
Female	139	46.0
Total	302	100

Source: Field survey (2019)

From Table 8, it can be observed that, for the experimental group, out of the 311 students, 146 representing 46.9% were male whiles 165 representing 53.1% were female. This means that more females partook in the study than males.

Inferring from Table 9, also shows that, for the control group, out of the 302 students, 163 representing 54.0% were male whiles 139 representing 46.0% were female. This means that more males partook in the study than females.

Results

Research Question

What is the effect of sequence of items in multiple-choice test on students' academic performance in:

i. Core Mathematics?

ii. Integrated Science?

The research question was in two folds and for that matter, the analysis was done separately. The main aim of the research question was to basically find out if the sequence of items in multiple-choice test has any effect on students' academic performance by comparing their performance over the three kinds of item sequence (Easy-To-Hard (ETH), Random (RDM) and Hard-To-Easy (HTE)) in both Core mathematics and integrated science.

Considering the effect of item sequence and students' academic performance in Core Mathematics, the analysis of this question was done using one-way repeated measures Analysis of Variance (ANOVA), which emphasises the use of means to compare the performance of students over the three kinds of item sequence. Before the analysis, students' academic performance over the three kinds of item sequence was determined by scoring their correct responses out of 50 marks as presented in Appendix A. All the conditions were checked and assumptions tested. In checking the first condition, which was the dependent variable should be continuous in nature (interval or ratio variables), it must be established that the dependent variable in this study were scores which were continuous and were obtained from the various kinds of item sequence. The second condition which emphasises that the independent variable should consist of at least two categorical, "related groups" or "matched pairs" was satisfied since the independent variable in this study was item sequence kinds which consist of three categories namely; (Easy-To-Hard (ETH), Random (RDM) and Hard-To-Easy (HTE)).

The assumptions which included a normality test and sphericity were tested. In testing for the sphericity, the Mauchly test of sphericity which was the more formal way for testing sphericity was used to test if variance differences

between the groups are equal. It must be established that if Mauchly test of sphericity is statistically significant (p<.05), we can reject the null hypothesis and accept the alternative hypothesis that the variance of the differences is not equal (i.e., sphericity has been violated). In this regard, the results of the test of sphericity is provided in Appendix E and the results indicated that the Mauchly test of sphericity was not statistically significant since p >.05 which signifies that sphericity has not been violated, $X^2(2) = 4.111, p = .128$. This means that, there are equal variances between the kinds of sequence of items.

The Test of Within-Subjects Effect was conducted to explore the main effect of item sequence by comparing the mean performance of students with respect to item sequence. Firstly, summary of descriptive statistics in Table 10 was explored.

Table 10: Descriptive Statistics of Performance in Item Sequence in Core Mathematics

	Mean	Std. Deviation	N
Maths HTE Scores	18.13	6.559	311
Maths Random Scores	24.85	8.433	311
Maths ETH Scores	30.08	7.701	311

Source: Field survey (2019), N = 311

From Table 10, it is evident that, there are differences in the mean scores of the kinds of item sequence. It can further be discovered that students' academic performance in the ETH (Mean = 30.08, Std. Dev. = 7.70) was better than the RDM (Mean = 24.85, Std. Dev. = 8.43) and HTE (Mean = 18.13, Std. Dev. = 6.55). This is because the ETH recorded the highest mean, followed by RDM and lastly the HTE. This therefore means that ETH item sequence has

greatest effect on students' academic performance since students attained higher scores for that sequence. Summary of results from ANOVA Test of Within Subjects Effect is also provided in Table 11.

Table 11: ANOVA Test of Within Subjects Effect

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Item sequence	Sphericity Assumed	22339.693	2	11169.847	343.523	.000	.526
	Greenhouse-Geisser	22339.693	1.974	11317.484	343.523	.000	.526
	Huynh-Feldt	22339.693	1.986	11245.900	343.523	.000	.526
	Lower-Bound	22339.693	1.000	22339.693	343.523	.000	.526
Error (Item Sequence)	Sphericity Assumed	20159.640	620	32.516			
	Greenhouse-Geisser	20159.640	611.912	32.945			
	Huynh-Feldt	20159.640	615.807	32.737			
	Lower-Bound	20159.640	310.000	65.031			

Source: Field survey (2019), N = 311

It must be established that, because the Mauchly test of sphericity was not violated or in others words was assumed, the sphericity assumed row as indicated in the test was used for the reporting in Table 11. In view of this, it can be observed from the results of the test in Table 11 that, there is a statistically significant difference between the means of performance of students over the three kinds of item sequence (F (2, 620) = 343.523, $p < 0.001$). In order to find where the differences in means of students' academic performance were coming from, a post hoc test was conducted using the Bonferroni correction as shown in Table 12.

Table 12: Pairwise Comparisons

(I) Item sequence	(J) Item sequence	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
HTE	ETH	-6.723*	.453	.000	-7.815	-5.632
	RDM	-11.955*	.436	.000	-13.005	-10.905
ETH	HTE	6.723*	.453	.000	5.632	7.815
	RDM	-5.232*	.481	.000	-6.390	-4.073
RDM	HTE	11.955*	.436	.000	10.905	13.005
	ETH	5.232*	.481	.000	4.073	6.390

Source: Field survey (2019), N = 311

From Table 12, there is a significant difference in mean scores between Mathematics Hard-To-Easy (HTE) and Mathematics Easy-To-Hard (ETH) ($p < 0.001$) and between Mathematics Random (RDM) and Mathematics Easy-To-Hard (ETH) ($p < 0.001$), as well as between Mathematics Hard-To-Easy (HTE) and Mathematics Random (RDM) ($p < 0.001$). In conclusion, the results indicate that item sequence has effect on students' academic performance in Core Mathematics and that ETH item sequence has greatest effect on students' academic performance since students attained higher scores for that sequence.

Taking into account the second fold of the research question which intended to investigate the effect of item sequence on students' academic performance in Integrated Science, the same procedure was used in analysing the science scores. Before the analysis, all the conditions were satisfied and assumptions tested. The normality and sphericity assumptions were also tested. The summary of sphericity assumptions which was tested using Mauchly test of sphericity is provided in Appendix F.

From Appendix F, it can be seen that the Mauchly test of sphericity was not statistically significant since $p>0.05$. This indicated that sphericity has not been violated, $X^2(2) = 3.962, p = .0570$. This means that, there are equal variances between the kinds of sequence of items. In furtherance, the Test of Within-Subjects Effect was conducted to discover the main effect of item sequence by comparing the mean performance of students with respect to item sequence. Before the Test of Within-Subjects Effect Table, the descriptive statistics was explored and the summary of the results is provided in Table 13.

Table 13: Descriptive Statistics of Performance in Item Sequence in Integrated Science

	Mean	Std. Deviation	N
SCI HTE Scores	15.07	3.649	311
SCI RDM Scores	23.05	9.001	311
SCI ETH Scores	27.10	8.475	311

Source: Field survey (2019), N = 311

From Table 13, it is clearly indicated that there were differences in the mean scores of the kinds of item sequence. It was again discovered that students' academic performance in the ETH (Mean = 27.10, Std. Dev. = 8.47) was better than the RDM (Mean = 23.05, Std. Dev. = 9.00) and HTE (Mean = 15.07, Std. Dev. = 3.64). This is because the ETH recorded the highest mean, followed by RDM and lastly the HTE. This therefore means that ETH item sequence has greatest effect on students' academic performance since students attained higher scores for that sequence. The ANOVA Test of Within Subjects Effect Table was then explored and the summary of the results is provided in Table 14.

Table 14: ANOVA Test of Within Subjects Effect

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Item Sequence	Sphericity Assumed	23325.764	2	11662.882	241.439	.000	.438
	Greenhouse-Geisser	23325.764	1.658	14064.689	241.439	.000	.438
	Huynh-Feldt	23325.764	1.666	13998.744	241.439	.000	.438
	Lower-bound	23325.764	1.000	23325.764	241.439	.000	.438
Error (Item Sequence)	Sphericity Assumed	29949.569	620	48.306			
	Greenhouse-Geisser	29949.569	514.123	58.254			
	Huynh-Feldt	29949.569	516.545	57.981			
	Lower-bound	29949.569	310.000	96.612			

Source: Field survey (2019), N = 311

From Appendix F, because the Mauchly test of sphericity was not violated, the sphericity assumed row as indicated in Table 14 was used for the reporting. In this regard, it can be observed from Table 14 that, there is a statistically significant difference between the mean performance of students over the three kinds of item sequence (F (2, 620) = 241.439, $p < 0.001$). In exploring where the differences in means of students' academic performance lies, a post hoc test was conducted using the Bonferroni correction. The summary of the results from the post hoc test is presented in Table 15.

Table 15: Pairwise Comparisons

(I) Item sequence	(J) Item sequence	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
HTE	ETH	-7.984*	.504	.000	-9.197	-6.771
	RDM	-12.035*	.477	.000	-13.182	-10.888
ETH	THE	7.984*	.504	.000	6.771	9.197
	RDM	-4.051*	.671	.000	-5.667	-2.435
RDM	THE	12.035*	.477	.000	10.888	13.182
	ETH	4.051*	.671	.000	2.435	5.667

Source: Field survey (2019), N = 311

The results of the post hoc test shown in Table 15 indicated that there was a significant difference in mean scores between Integrated Science Hard-To-Easy (HTE) and Integrated Science Easy-To-Hard (ETH) ($p < 0.001$) and between Integrated Science Random (RDM) and Integrated Science Easy-To-Hard (ETH) ($p < 0.001$), as well as between Integrated Science Hard-To-Easy (HTE) and Integrated Science Random (RDM) ($p < 0.001$). In conclusion, the

results indicate that item sequence has effect on students' academic performance in Integrated Science and that ETH item sequence has greatest effect on students' academic performance since students attained higher scores for that sequence.

From the results so far, it can be observed that there is a significant effect of item sequence on students' academic performance in Core Mathematics since a significant difference was found in the means of the three kinds of sequence, with ETH sequence leading to better students' academic performance. Similarly, for Integrated Science, the results indicated that there is a significant effect of item sequence on students' academic performance since a significant difference was observed in the means of the three kinds of sequence where ETH recorded a better performance.

Research Hypothesis One

There is no statistically significant difference in students' academic performance in item sequencing in Core Mathematic and Integrated Science with respect to gender.

This research hypothesis is also in two folds and therefore the analysis was done for each of them. The main goal of this hypothesis was to find out if a statistically significant difference exists in academic performance of student over the three kinds of item sequence in Core Mathematic and Integrated Science with respect to gender.

Considering one-fold of the hypothesis which was to find out if a statistically significant difference exists in academic performance of students in item sequencing in Core Mathematic with respect to gender, the analysis was done using the One-Way Multivariate Analysis of Variance (MANOVA).

Before the main analysis, all the conditions as well as the assumptions were adhered to and tested respectively. In checking the first condition which highlights that, there should be two or more dependent variables which are measured in interval or ratio scale (i.e., they are continuous), it must be emphasised that, the dependent variables in this hypothesis are the three kinds of item sequence namely the Hard-To-Easy (HTE), Easy-To-Hard (ETH) and Random (RDM) which are all continuous in nature because scores were obtain for each of them. The second condition which also says that the independent variable must consist of two or more categorical independent groups was also met since the independent variable in the second hypothesis are the males and females (gender).

With regards to the assumptions, the multivariate normality was tested using Shapiro Wilk where the dependent variables were approximately normally distributed on the independent variables. The test indicated that, the ETH and RDM were not significant (EHT (Male, $p = .000$, Female, $p = .000$), (RDM (Male, $p = .000$, Female, $p = .005$) but the HTE was significant (HTE (Male, $p = .232$, Female, $p = .377$) (see Appendix G). Even though the ETH and RDM were not significant, the F test is robust to non-normality and therefore MANOVA was still conducted. Another important assumption which was tested for was the homogeneity of variance-covariance matrices. In testing for this assumption, the Box's M test is used. The summary of the results is provided in Appendix H. It is evident from Appendix H that, $p > .05$ was obtained which suggest that this assumption was not violated. This means that the variance-covariance matrices of the kinds of Mathematics item sequence (dependent variable) are equal. Once all the assumptions and the conditions

were met, further tests were conducted. The next important test was the multivariate test which indicated whether significant differences existed among the groups (male and female) on a linear combination of the dependent variable (HTE, ETH and RDM). The summary of the results is presented in Table 16.

Table 16: Multivariate Test for Core Mathematics

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.947	1810.925[b]	3.000	307.000	.000	.947
	Wilks' Lambda	.053	1810.925[b]	3.000	307.000	.000	.947
	Hotelling's Trace	17.696	1810.925[b]	3.000	307.000	.000	.947
	Roy's Largest Root	17.696	1810.925[b]	3.000	307.000	.000	.947
Gender of Students	Pillai's Trace	.003	.297[b]	3.000	307.000	.828	.003
	Wilks' Lambda	.997	.297[b]	3.000	307.000	.828	.003
	Hotelling's Trace	.003	.297[b]	3.000	307.000	.828	.003
	Roy's Largest Root	.003	.297[b]	3.000	307.000	.828	.003

Source: Field Survey (2019), N = 613

From Table 16, It can be observed from the Wilks' Lambda row in the second section that p >.05. This means that item sequence (HTE, ETH and RDM) in Core Mathematics was not significantly dependent on the gender of students with η^2 = 0.003 indicating that the effect size is very small. This explains the amount of variance the predictor explains in the outcome variable. This further means that gender only explains 0.3% in students item sequence performance. Because the multivariate test was not significant, further investigation to find out where males and females did not differ significantly on all the dependent measures was not necessary. In conclusion, the results indicate that students' academic performance in the various item sequence (HTE, ETH and RDM) in Core Mathematics did not significantly differ with respect to their gender.

Taking into consideration the second fold of the hypothesis one which aimed at finding out if statistically significant difference exists in academic performance of students in item sequencing in Integrated Science with respect to gender, the analysis was done through the use of One-Way Multivariate Analysis of Variance (MANOVA). It must be emphasised that, the procedures used for analysing the Core Mathematics items was also used for analysing scores for the kinds of item sequence in the Integrated Science. Specifically, all the conditions as well as the assumptions were adhered to and tested respectively including normality test (See Appendix I) just like it was done for Core Mathematics items. In testing for the homogeneity of variance-covariance matrices which is the one of the important assumptions of MANOVA. The Box's M test was used and the summary of the results is provided in Appendix J.

The results in Appendix J produced $p > .05$ which indicates that the homogeneity of variance-covariance matrices assumption was not violated. This means that the variance-covariance matrices of the kinds of Integrated Science item sequence (dependent variable) are equal. This further indicates that the variance and covariance of the various item sequence were equal. Once the homogeneity of variance-covariance matrices assumption was not violated, a multivariate test which indicated whether significant differences existed among the groups (male and female) on a linear combination of the dependent variable (HTE, ETH and RDM) was performed and the summary of the results is presented in Table 17.

Table 17: Multivariate Test for Integrated Science

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.962	2569.980[b]	3.000	307.000	.000	.962
	Wilks' Lambda	.038	2569.980[b]	3.000	307.000	.000	.962
	Hotelling's Trace	25.114	2569.980[b]	3.000	307.000	.000	.962
	Roy's Largest Root	25.114	2569.980[b]	3.000	307.000	.000	.962
GENDER	Pillai's Trace	.017	1.811[b]	3.000	307.000	.045	.057
	Wilks' Lambda	.983	1.811[b]	3.000	307.000	.045	.057
	Hotelling's Trace	.018	1.811[b]	3.000	307.000	.045	.057
	Roy's Largest Root	.018	1.811[b]	3.000	307.000	.045	.057

Source: Field survey (2019), N = 311

From Table 17, it can be observed from the Wilks' Lambda row in the second section that, the multivariate test is significant thus; $p < .05$. This means that item sequence (HTE, ETH and RDM) in Integrated Science was significantly dependent on the gender of students. In other words, students' academic performance over the three kinds of item sequence in Integrated Science differed with respect to being male or female with $\eta^2 = 0.057$ indicating that the effect size is small. This explains the amount of variance the predictor explains in the outcome variable. This further means that gender only explains 5.7% in students item sequence performance. The multivariate test was significant and that indicated that further investigation needed to be done to find out where males and females significantly differed on all the dependent measures. This was done using the Test of Between-Subjects Effect. Before exploring the Test of Between-Subjects Effect, the descriptive statistics was firstly explored and the summary of the results is provided in Table 18.

Table 18: Descriptive Statistics of Item Sequence with Respect to Gender

	Gender of Students	Mean	Std. Deviation	N
SCI HTE	Male	15.64	3.478	146
Scores	Female	17.45	3.763	165
	Total	15.07	3.649	311
SCI RDM	Male	22.70	8.975	146
Scores	Female	25.36	9.039	165
	Total	23.05	9.001	311
SCI ETH	Male	27.25	8.342	146
Scores	Female	29.85	8.546	165
	Total	27.10	8.475	311

Source: Field survey (2019), N = 311

It is clearly indicated in Table 18 that the means and standard deviations of the dependent variables (Integrated Science HTE (Male: mean = 15.64, Std dev. = 3.478; Female: mean = 17.45, Std dev. = 3.763), Integrated Science RDM (Male: mean = 22.70, Std dev. = 8.975; Female: mean = 25.36, Std dev. = 9.039) and Integrated Science ETH (Male: mean = 27.25, Std dev. = 8.342; Female: mean = 29.85, Std dev. = 8.546)) with respect to gender (Male and Female) significantly differed which suggest that academic performance of students with respect to the sequence of items in Integrated Science significantly depended on whether the student was a male or female. The Test of Between-Subjects Effects was furthered explored and the summary of the results is presented in Table 19.

Table 19: Test of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	SCI HTE Scores	51.010[a]	1	51.010	3.866	.030	.012
	SCI RDM Scores	34.255[b]	1	34.255	.422	.039	.001
	SCI ETH Scores	198.575[c]	1	198.575	2.781	.042	.009
Intercept	SCI HTE Scores	70111.627	1	70111.627	5314.389	.000	.945
	SCI RDM Scores	164348.982	1	164348.982	2024.961	.000	.868
	SCI ETH Scores	226776.942	1	226776.942	3175.639	.000	.911
GENDER	SCI HTE Scores	51.010	1	51.010	3.866	.030	.012
	SCI RDM Scores	34.255	1	34.255	.422	.039	.001
	SCI ETH Scores	198.575	1	198.575	2.781	.042	.009
Error	SCI HTE Scores	4076.572	309	13.193			
	SCI RDM Scores	25078.922	309	81.162			
	SCI ETH Scores	22066.132	309	71.411			
Total	SCI HTE Scores	74734.000	311				
	SCI RDM Scores	190369.000	311				
	SCI ETH Scores	250715.000	311				
Corrected Total	SCI HTE Scores	4127.582	310				
	SCI RDM Scores	25113.177	310				
	SCI ETH Scores	22264.707	310				

Source: Field survey (2019), N = 311

It is evidently shown in the third section of Table 19 that, the sig values for Integrated Science HTE, Integrated Science RDM and Integrated Science ETH with respect to gender were .030, .039 and .042 respectively which means $p < .05$. This indicates that the kinds of item sequence, whether HTE, ETH or RDM in Integrated Science was dependent on the student being male or female. In other words, there is statistically significant difference in the means scores of the kinds of item sequence with respect to gender.

From the results indicated so far, it can be observed that, for Core Mathematics, there was no significant difference in the means scores of the kinds of item sequence with respect to gender. On the other hand, there was a statistically significant difference in the means scores of the kinds of item sequence with respect to gender for Integrated Science with the females performing better than the males in all the three kinds of item sequence.

Research Hypothesis Three

Gender does not significantly moderate the relationship between items sequencing in Multiple Choice Test (MCT) and students' academic performance in Core Mathematics and Integrated Science.

This hypothesis sought to find out if gender significantly moderates the relationship between items sequencing in Multiple Choice Test (MCT) and students' academic performance in Core Mathematics and Integrated Science.

With regards to the first fold of the hypothesis which sought to find out if gender significantly moderates the relationship between items sequencing in Multiple Choice Test (MCT) and students' academic performance in Core Mathematics, a moderation analysis was done to explore the moderating role gender plays in the relationship between items sequencing in Multiple Choice

Test (MCT) and students' academic performance in Core Mathematics. In view of this, the first item sequence that was considered was the scores for the Easy-To-Hard (ETH) sequence.

In the moderation model, the predictor variable was the Easy-To-Hard (ETH) sequence scores, with the scores obtained from the control group as the outcome variable while the moderator was gender. The moderation analysis was conducted using the Process by Hayes (2018). The moderation analysis was done using 5,000 bootstrap samples. Results are shown in Table 20.

Table 20: Moderating Role of Gender in The Relationship Between Easy-To-Hard (Eth) Sequence and Students' Academic Performance

	Coeff	SE	t-value	LLCI	ULCI
Constant	25.7018	6.4073	4.0113	13.0939	38.3096
METH	-.0908	.2075	-.4378	-.4991	.3174
Gender	-.5667	4.0072	-.1414	-8.4518	7.3185
METH *Gender	.0587	.1292	.4546	-.1955	.3129

Source: Field survey (2019), N = 613

Model summary: R^2=.0734; F (3, 609) = .5538, p>. 05

METH *Gender: R^2 *change*= .0007; F (1, 609) = .2067, p>.05

Criterion: Students' academic performance

The result in Table 20 shows that gender is not a significant moderator in the relationship between Easy-To-Hard (ETH) sequence and students' academic performance R^2 = .0054, F (3, 609) = .5538. The interaction effect between gender and Easy-To-Hard (ETH) sequence is not significant, p = .6497 > .05.

The same procedure was used to test the moderating role of gender in relationship between Random (RDM) sequence and students' academic performance. The results are shown in Table 21.

Table 21: Moderating Role of Gender in The Relationship Between Random (RDM) Sequence and Students' Academic Performance

	Coeff	SE	t-value	LLCI	ULCI
Constant	20.2134	4.9154	4.1124	10.5416	29.8851
MRDM	.1103	.1864	.5920	-.2564	.4770
Gender	2.1973	3.0879	.7116	-3.8788	8.2733
MRDM*Gender	-.0396	.1176	-.3369	-.2711	.1918

Source: Field survey (2019), N = 613

Model summary: R^2=.0075; *F* (3, 609) = .7718, *p* > .05

MRDM *Gender*: R^2 change*= .0004; *F* (1, 609) = .1135, *p* > .05

Criterion: Students' academic performance

The result in Table 21 shows that gender is not a significant moderator in the relationship between Random (RDM) sequence and students' academic performance R^2 = .0075, F (3, 609) = .7718. The interaction effect between gender and Random (RDM) sequence is not significant, p = .7364 > .05.

Finally, the moderating role of gender in relationship between Hard-To-Easy (HTE) sequence and students' academic performance was also explored. The results are shown in Table 22.

Table 22: Moderating Role of Gender in The Relationship Between Hard-To-Easy (HTE) Sequence and Students' Academic Performance

	Coeff	SE	t-value	LLCI	ULCI
Constant	-1.0821	12.2076	-.0886	-25.5416	22.9773
MHTE	.7319	.1815	4.0325	.3742	1.0897
Gender	10.5610	6.6582	1.5862	-2.5614	26.6833
MHTE*Gender	-.2010	.0999	-2.0128	-.3978	-.0042
Conditional effect of the focal predictor at the values of the moderator					
Male	.5309	.0875	6.0671	.3585	.7034
Female	.3300	.0481	6.8608	.2352	.4247

Source: Field survey (2019), N = 613

Model summary: R^2=.3033; F (3, 609) = 31.7856, p <. 001

MHTE *Gender*:* R^2 *change*= .0129; F (1, 609) = 4.0515, p < .05

Criterion: Students' academic performance

The result in Table 22 shows that gender is a significant moderator in the relationship between Hard-To-Easy (HTE) sequence and students' academic performance. The results indicate that the relationship between Hard-To-Easy (HTE) and performance in males is higher, b= .5309, t= 6.0671, CI (.3585, .7034) and p < .001 as compared to females, b= .3300, t = 6.8608, CI (.2352, .4247). This further means that in males, the positive relationship that exists between HTE sequence and performance is higher than in females. A diagrammatic representation of the moderating role of gender in the relationship between Hard-To-Easy (HTE) sequence and students' academic performance is shown in Figure 2.

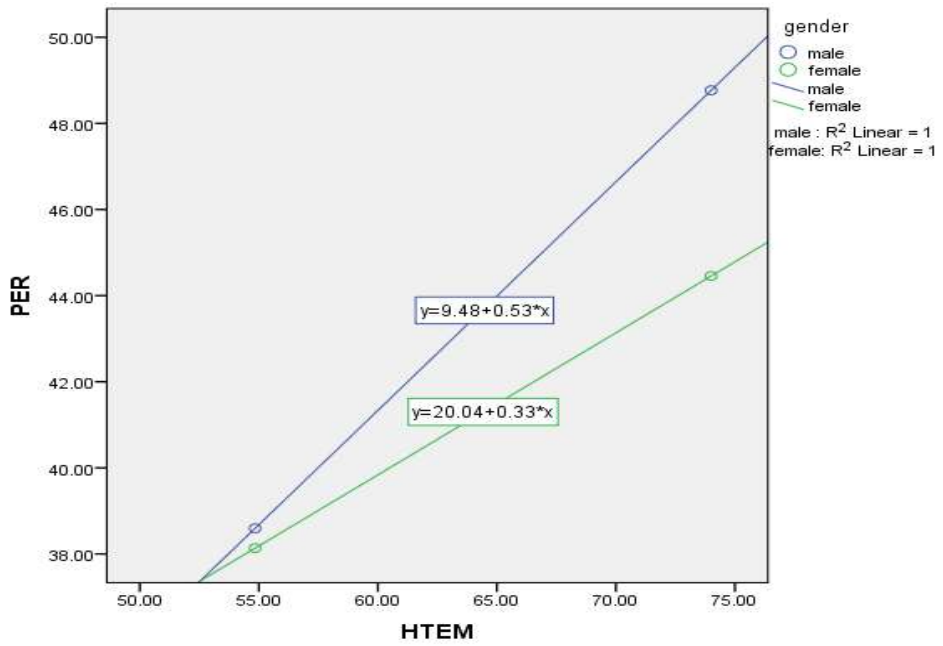


Figure 2: Shows moderating role of gender in the relationship between Hard-To-Easy (HTE) sequence and students' academic performance

Considering the second fold of the hypothesis which sought to find out if gender significantly moderates the relationship between items sequencing in Multiple Choice Test (MCT) and students' academic performance in Integrated Science, the same moderation analysis was done to explore the moderating role gender plays in the relationship between items sequencing in Multiple Choice Test (MCT) and students' academic performance in Integrated Science. In view of this, the first item sequence that was considered was the scores for the Easy-To-Hard (ETH) sequence.

In the moderation model, the predictor variable was the Easy-To-Hard (ETH) sequence scores, with the scores obtained from the control group as the outcome variable while the moderator was gender. The moderation analysis was conducted using the Hayes (2018) approach (PROCESS). Moderation analysis was done using 5,000 bootstrap samples. Results are shown in Table 23.

Table 23: Moderating Role of Gender in The Relationship Between Easy-To-Hard (ETH) Sequence and Students' Academic Performance

	Coeff	SE	t-value	LLCI	ULCI
Constant	32.5003	7.7028	4.2193	-17.3191	47.6814
SETH	.3705	.1667	2.2217	.0418	.6991
Gender	4.6950	4.1734	1.1250	-3.5303	12.9202
SETH*Gender	-.2011	.0949	-2.1204	-.3881	-.0142
Conditional effect of the focal predictor at the values of the moderator					
Male	.1693	.0792	2.1386	.0133	.3254
Female	-.0318	.0522	-.6088	-.1347	.0711

Source: Field survey (2019), N = 613

Model summary: R^2=.0578; F (3, 609) = 4.4768, p <. 05

SETH*Gender R^2 *change*= .0193; F (1, 609) = 4.4960, p < .05

Criterion: Students' academic performance

The result in Table 23, shows that gender is a significant moderator in the relationship between Easy-To-Hard (ETH) sequence and students' academic performance. The results further indicate that, in males, there is positive relationship between ETH sequence and performance and such relationship is higher, b = .1693, t = 2.1386, CI (.0133, .3254) and p < .001 as compared to females which exhibits a negative relationship between ETH sequence and performance, b= -.0318, t = -.6088, CI (-.1347, .0711). This therefore means that, the relationship between ETH sequence and performance is higher than in females. A diagrammatic representation of the moderating role of gender in the relationship between Easy-To-Hard (ETH) sequence and students' academic performance is shown in Figure 3.

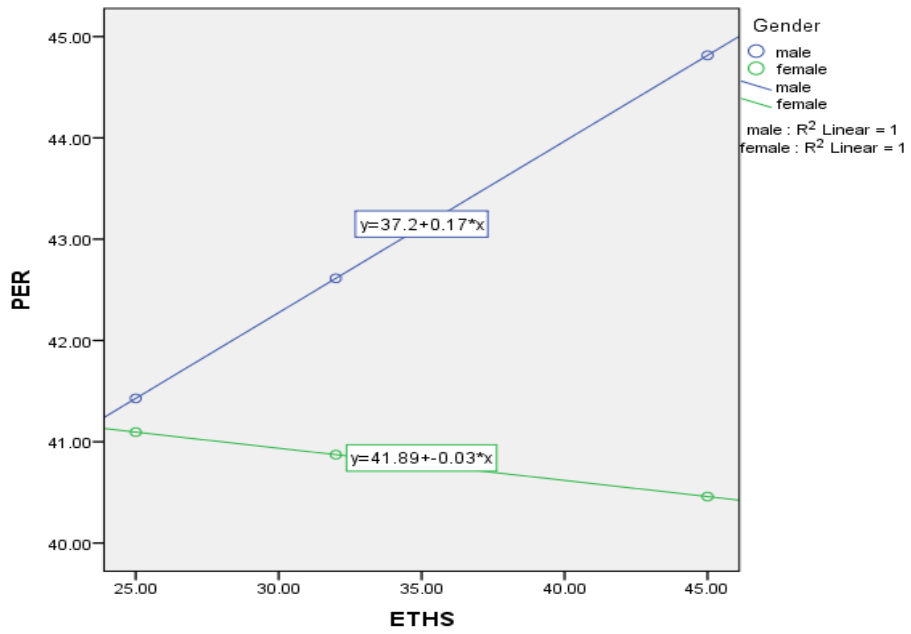


Figure 3: Shows moderating role of gender in the relationship between Easy-To-Hard (ETH) sequence and students' academic performance

In checking for the moderating role of gender in relationship between Random (RDM) sequence and students' academic performance, the same aforementioned procedure was used. The results are shown in Table 24.

Table 24: Moderating Role of Gender in the Relationship Between Random (RDM) Sequence and Students' Academic Performance

	Coeff	SE	t-value	LLCI	ULCI
Constant	29.3079	4.4986	6.5148	20.4559	38.1600
SRDM	-.1893	.1835	-1.0313	-.5504	.1718
Gender	-3.4285	2.8095	-1.2204	-8.9568	2.0997
SRDM *Gender	-.0902	.1137	.7927	-.1337	.3140

Source: Field survey (2019), N = 613
Model summary: R^2=.0106; F (3, 609) = 1.0937, p > .05
SRDM *Gender: R^2 *change*= .0020; F (1, 609) = .6284, p > .05
Criterion: Students' academic performance

The result in Table 24 shows that gender is not a significant moderator in the relationship between Random (RDM) sequence and students' academic performance R^2 = .0106, F (3, 609) = 1.0937. The interaction effect between gender and Random (RDM) sequence is not significant, p = .4285 > .05.

The moderating role of gender in the relationship between the Hard-To-Easy (HTE) sequence and students' academic performance was also explored using the same procedure. The results are shown in Table 25.

Table 25: Moderating Role of Gender in The Relationship Between Hard-To-Easy (HTE) Sequence and Students' Academic Performance

	Coeff	SE	t-value	LLCI	ULCI
Constant	33.9085	7.0718	4.7949	19.9933	47.8238
SHTE	-.6166	.4658	-1.3236	-1.5332	.3001
Gender	-4.5563	4.3628	-1.0444	-13.1410	4.0284
SHTE *Gender	-.2256	.2830	.7973	-.3312	.7825

Source: Field survey (2019), N = 613
Model summary: R^2=.0192; *F* (3, 609) = 2.0018, *p* >.05
SHTE *Gender: R^2 *change*= .0020; *F* (1, 609) = .6357, *p* > .05
Criterion: Students' academic performance

The result in Table 25 shows that gender is not a significant moderator in the relationship between Hard-To-Easy (HTE) sequence and students' academic performance R^2 = .0192, F (3, 609) = 2.0018. The interaction effect between gender and Hard-To-Easy (HTE) sequence is not significant, p = .4259 > .05.

Discussion of Research Findings

The discussion of the research findings was done in line with the following themes:

1. The effect of item sequencing in Multiple-Choice Test (MCT) on students' academic performance.
2. Difference in students' academic performance in item sequencing in Core Mathematic and Integrated Science with respect to gender.
3. The moderating role of gender in the relationship between item sequencing in Multiple-Choice Test (MCT) and students' academic performance

The Effect of Item Sequencing in Multiple-Choice Test (MCT) on Students' Academic Performance

The first research question sought to find out the effect of item sequencing in Multiple-Choice Test (MCT) on students' academic performance in Core Mathematics and Integrated Science. In view of this, the discussion was done separately for each subject. With regards to the effect of item sequencing in Multiple-Choice Test (MCT) on students' academic performance in Core Mathematics, the findings from the results indicated that item sequencing in Multiple-Choice Test (MCT) has effect on the performance of students in the Kumasi Metropolis. This is rightly supported by the assertion of Shepard (1994) who points out that tiny changes with respect to the sequence of items in a Multiple-Choice Test (MCT) can also make large difference in students' academic performance. Similarly, the findings are supported by Opara and Uwah (2017) who reported that item sequencing based on ascending (ETH) and descending order (HTE) positively and significantly affect students' academic

performance. He further reported that, the random (RDM) or no order arrangement was also positive but not significant which means that arranging items in random order had no effect on students' academic performance. The findings of Opara and Uwah's study are line with the findings of this study since all the various kinds of item sequence (ETH, RDM and HTE) recorded different mean values.

Juxtaposing the findings of Cobbinah (2016) with the findings of this study revealed that even though, Easy-To-Hard (ETH) sequence recorded the best scores followed by Random (RDM) and lastly Hard-To-Easy (HTE) in this study, Cobbinah's findings indicated slight differences with the Random (RDM) sequence recording the highest, followed by Easy-To-Hard (ETH) and finally Hard-To-Easy (HTE). This might have happened as a result of differences in the students used for the study. Cobbinah conducted his study using Senior High Schools students in the Cape Coast Metropolis in the" Central region of Ghana whiles this still focused on Kumasi Metropolis in the Ashanti Region of Ghana. The findings of the study are again consistent with the findings of Ollennu and Etsey (2015). This is because, the pattern of the effect of item sequence in their study was the same as the findings of this study. The results of their study indicated especially in Mathematics that when the sequence of item is altered, the difference in performance was generally significant where students' scores in Easy-To-Hard (ETH) recorded the best scores followed by Random (RDM), and lastly, Hard-To-Easy (HTE) sequence just like it was revealed in the findings of this research question.

Again, the findings of the study are consistent with Alakayleh (2017) who reported Random (RDM) sequence recording the highest mean score,

followed by Hard-To-Easy (HTE) and lastly the Easy-To-Hard (ETH). The findings of this study is in agreement with Alakayleh's findings since both findings proved that item sequencing has effect on students' academic performance. In this study, the ETH recorded the highest mean, followed by RDM and finally HTE which differed from Alakayleh's findings where Random (RDM) recorded the highest mean score, followed by Hard-To-Easy (HTE) and lastly the Easy-To-Hard (ETH). This might have also happened as a result of the differences in students as well as geographical setting used in the both studies.

Considering the second fold of the research question that sought to find out the effect of item sequencing in Multiple-Choice Test (MCT) on students' academic performance in Integrated Science, the findings of study show that item sequencing in Multiple-Choice Test (MCT) has effect on students' academic performance. The findings of the study are consistent with Gray (2003) who reported that question order can have a statistically significant effect on student responses, where students' academic performance was better in the accelerating (Easy-To-Hard) than the non-accelerating (Hard-To-Easy). The findings of Gray corroborate with the findings of this study since the study reported that, students' academic performance in the Easy-To-Hard was better than the Hard-To-Easy just like in Gray's findings.

The findings of this study are again in agreement with the findings reported in the study conducted by Ollennu and Etsey (2015). The results of Ollennu and Etsey indicated that there was a significant difference in performance in Integrated Science when item order was altered. Students' academic performance in the Hard-To-Easy (HTE) was the best, followed by

Random (RDM) and, lastly the Easy-To-Hard (ETH). Even though, the findings of Ollennu and Etsey are consistent with the findings of study, there were slight differences regarding the means scores of the various item sequence. The findings of this study revealed that the performance of students in the Easy-To-Hard (ETH) was the best, followed by Random (RDM) and lastly Hard-To-Easy (HTE) as compared to the findings of Ollennu and Etsey. From the above discussion, it can be inferred that, even though the findings relate to Senior High School students in the Kumasi Metropolis, it appears the sequence of items has influence on student performance. This finding is applicable to other students in other study areas in Ghana as it has been confirmed by Cobbinah (2016) and Ollennu and Etsey (2015).

It can be linked to the fact that, one of the factors that affect validity and onward reliability is the factor that relate to the arrangement of items. When items are arranged from the least difficult ones to the most difficult ones, it tends to motivate students to answer more or all questions as compared to when the first few items are very difficult. When students are able to answer more questions regarding a construct, we are more likely to conceptualise how well they have grasped the concepts taught, which makes item sequence a determinant of validity and reliability.

Difference in Students' academic performance in Item Sequencing in Multiple-Choice Test (MCT) with Respect to Gender

The first hypothesis sought to find out if there is a statistically significant difference in students' academic performance in item sequencing in Core Mathematic and Integrated Science with respect to gender.

For Core Mathematics, the findings from the results indicated that item sequence (HTE, ETH and RDM) in Core Mathematics was not significantly dependent on the gender of students. In addition to this finding, the descriptive statistics of the students' academic performance over the three kinds of item sequence also confirms that, the mean scores for the three kinds of item sequence did not significantly differ on students' gender as males and females recorded approximately the same mean scores. However, Cobbinah (2016) found otherwise as he reported that there are significant differences in the achievement of students with respect to gender where the male students generally performed better no matter the order of items than their female counterpart. The findings oppose the findings of this study since there is a statistically significant difference in the achievement of students with respect to gender. The differences in these findings may be as a result of the fact that Cobbinah used students from different school categories as his main focus was on the students being at the same stage of completion as far as the coverage of the form two SSCE syllabus was concerned to the neglect of their school category.

The findings derived from students' academic performance in item sequencing in Integrated Science, indicated that item sequence (HTE, ETH and RDM) was significantly dependent on the gender of students. In addition to this finding, the descriptive statistics of the students' academic performance over the three kinds of item sequencing also confirmed that, the mean scores for the three kinds of item sequence significantly differed with respect to students' gender where the females outperformed the males. This finding is in line with the findings of Research Division of WAEC (1993). It was reported that

performance of students in item sequence differed with respect to gender except Hard-to-Easy in Mathematics and Random Order in English Language. Their findings further reported that irrespective of the arrangement adopted in Economics, Biology and Agricultural Science, the pattern of performance, was susceptible to differences in sex where the Hard-to-Easy sequence in Mathematics disadvantaged the female candidates while the Random-Order in English Language seem to have reverse effect though it is assumed that the occurrence is by chance. In conclusion, from the SHS students that were sampled from the Kumasi Metropolis, there was no differences in the performance of students in item sequencing with respect to gender in Core Mathematics but otherwise for Integrated Science. It is worth noting that, these findings if replicated using category "A" schools may produce similar findings.

The Moderating Role of Gender in the Relationship Between Item Sequencing in Multiple-Choice Test (MCT) and Students' Academic Performance

The second hypothesis sought to find out if gender significantly moderates the relationship between item sequencing in Multiple Choice Test (MCT) and students' academic performance in Core Mathematics and Integrated Science. The findings derived from this hypothesis indicated that while gender moderates the relationship between some kinds of sequence, others do not.

For item sequence in Mathematics, the results show that gender partially moderates the relationship between HTE item sequence and performance but gender does not moderate the relationship between ETH and RDM and performance.

In Integrated Science, the results also indicated that gender only moderated the relationship between ETH item sequence and performance but gender does not moderate the relationship between HTE and RDM and performance. These findings agree with assertion of Linver, Davis-Kean and Eccles (2002) who noted that for females, the relationship between item sequence and academic performance was significant, and positive. Also, for a gender type (especially being a male), there was a significant relationship between item sequence and academic performance, and was positive as well.

From the above discussion, it is clearly indicated that the study was important since limited information can be found on the moderating role of gender in the relationship between item sequence and students' academic performance. From the SHS students used in Kumasi Metropolis, gender moderates the effect item sequencing has on students' academic performance.

Chapter Summary

The chapter has presented the results and discussion of the study. The findings of the study have uncovered the effect that item sequencing has on students' academic performance in both Core Mathematics and Integrated Science with ETH sequence in Core Mathematics and Integrated Science having the greatest effect since it recorded the highest mean score leading to better students' academic performance. The study has further indicated that students' academic performance in item sequencing differed with respect to their gender in Integrated Science but indicated otherwise in Mathematics. The moderating role of gender in the relationship between item sequence and students' academic performance in both subjects was also investigated and the findings revealed that gender significantly moderated the relationship between item sequence and

students' academic performance in some kinds of item sequence but others did not. Specifically for Mathematics MCT, the results showed that gender moderates the relationship between HTE item sequence and performance but gender did not moderate the relationship between ETH and RDM and performance whiles for Integrated Science, the results also indicated that gender only moderated the relationship between ETH item sequence and performance but gender did not moderate the relationship between HTE and RDM and performance.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction

This chapter highlights the summary of the key findings, the conclusions drawn as well as recommendations of the study. The chapter also presents the contributions of the study to knowledge and suggestions for further research.

Overview of the Study

The study sought to investigate the effect of item sequencing in multiple-choice test on students' academic performance and the moderating role of gender in the relationship between item sequencing in multiple-choice test and students' academic performance in the Kumasi Metropolis of the Ashanti Region of Ghana. In pursuing to this purpose, the following research question and hypotheses were formulated to guide the study:

Research Question

What is the effect of sequence of items in multiple-choice test on students' academic performance in:

i. Core Mathematics?

ii. Integrated Science?

Hypotheses

H_0: There is no statistically significant difference in students' academic performance in item sequencing in Core Mathematic and Integrated Science with respect to gender.

H_1: There is a statistically significant difference in students' academic performance in item sequencing in Core Mathematic and Integrated Science with respect to gender.

H_0: Gender does not significantly moderate the relationship between item sequencing in Multiple Choice Test (MCT) and students' academic performance in Core Mathematics.

H_1: Gender significantly moderates the relationship between item sequencing in Multiple Choice Test (MCT) and students' academic performance in Core Mathematics.

The quasi-experimental research design was adopted. Adapted WAEC past questions were used to collect data from a sample of 714 Senior High School Form 2 students who were selected through multi-staged sampling procedures. One-way repeated ANOVA, One-Way Multivariate Analysis of Variance (MANOVA), and the process procedure (Hayes 2018) were used in the analysis.

Summary of Key Findings

The key findings are outlined in consonance with the objectives of the study as follows:

1. The key findings were that item sequencing comprising of Easy-To-Hard (ETH), Hard-To-Easy (HTE) and Random (RDM) significantly affected students' academic performance in both Core Mathematics and Integrated Science. Further analysis showed that, for Core Mathematics, students performed better on the ETH sequence, followed by the RDM and lastly the HTE. Similarly, in Integrated Science, students performed well in ETH sequence, followed by the RDM, and lastly the HTE.
2. In Core Mathematics, there was no significant difference in the academic performance in item sequencing of students with respect to gender. However, for Integrated Science, there was a significant

difference in students' academic performance in item sequencing with respect to gender where the female students outperformed the males.

3. The key findings that emerged was that, for Core Mathematics, gender partially moderated the relationship between items sequence and students' academic performance both in Mathematics and Integrated Science. This is because, in males, the relationship between HTE sequence and performance were higher as compared to their female counterparts. For Integrated Science, gender also moderated the relationship between items sequence and students' academic performance where in males, the relationship between ETH sequence and performance was also higher, as compared to females.

Conclusions

Inferring from the findings of the study, the following conclusions are made:

First and foremost, it can be concluded from the findings that, item sequence has effect on students' academic performance in both Core Mathematics and Integrated Science. This is evident because Senior High School students tend to perform better when items are sequence using Easy-To-Hard sequence in both subjects. This might have happened because arranging items from Easy-To-Hard might have served as a motivating factor to sustain students' interest in answering questions which could increase their performance, and this might have happened among the students in the study area. This strengthens the conclusion that Easy-To-Hard sequence best relates to students' performance.

Secondly, it can also be concluded from the findings that, while gender affect students' academic performance in specific sequence type in Integrated

Science, it is not the case in Core Mathematics. This is evident because, males tend to perform differently than their female counterpart when items are sequenced from ETH, RDM and HTE in Integrated Science. Conversely, in Core Mathematics, Senior High School students' academic performance in ETH, RDM and HTE did not differ among males and female students. It can therefore be concluded that, students' academic performance in item sequencing in Integrated Science is influenced by gender but that is not the case in Core Mathematics.

Finally, the study makes a conclusion that while gender moderates the relationship between some kinds of item sequence, it fails to moderates the relationship between other kinds of item sequence and students' academic performance both in Core Mathematics and Integrated Science. This conclusion is premised on the facts that gender only moderates the relationship between Hard-To-Easy (HTE) sequence and students' academic performance in Core Mathematics but did not moderate the relationship between Easy-To-Hard (ETH) and Random (RDM) in Mathematics. Similarly, in Integrated Science, gender only moderates the relationship between Easy-To-Hard (ETH) sequence and students' academic performance, but not in Random (RDM) and Hard-To-Easy sequence (HTE). Thus, gender does not serve as a wholistic moderator of the relationship between item sequence and students' academic performance although it can be seen as a partial moderator in such relationship.

Recommendations

In view of the aforementioned research findings as well as the conclusions, the following recommendations are made:

1. The study recommends that stakeholders in education, such as Integrated Science and Mathematics teachers, WAEC, other assessment agencies and testing institutions should ensure the use of particular item sequence in order to encourage students' performance in Mathematics and Science. These stakeholders are encouraged to sequence items using ETH sequence so that it will motivate students which will have a cumulative effect on students' performance.
2. The study makes a recommendation that classroom teachers should encourage SHS students and in particularly male students to develop habits that will lead to an increase in performance in Integrated Science irrespective of a particular sequence type. This can be done by encouraging them especially male students to develop proper study habits, cooperative learning, and punctuality in class, amongst others which can predict good performance.
3. The study also makes a recommendation that test constructors and classroom teachers should take into consideration the role of gender and its effect when constructing test items to assess students' performance. Particularly, test constructors should be mindful of the role of gender in constructing HTE test items in Mathematics, and ETH test items in Science, as such sequences tend to favour males as compared to females.

Contribution to Knowledge

One of the reasons why a study is conducted is to contribute to existing knowledge. In view of this, Silverman (as cited in Asamoah, 2018) maintains that the ability of any research to contribute to knowledge could be displayed in four key areas which include developing a concept, thinking through the

methodology, building on an existing study and being able to change directions. In this regard, generally, the findings of this study contribute to knowledge on the variables that relates to the effect of item sequencing on the students' academic performance and also whether or not gender moderates the relationship between item sequencing and students' academic performance.

It is worth mentioning that, the study has revealed that characteristics of items such as item sequence or arrangement in multiple-choice test can have effect on students' academic performance. This has added to body of knowledge especially in the Kumasi Metropolis. The study has also revealed the role gender plays in the relationship between item sequence and its effect on students' academic performance which is a contribution to literature by providing empirical evidence on the aforementioned variables for researchers who wish to conduct a similar study.

In furtherance, the study has contributed to knowledge since it appears no study of this nature has ever been conducted in the Kumasi Metropolis in the Ashanti Region. In exploring literature, it was discovered that, the studies conducted on the aforementioned variables especially in Ghana focused on Central Region and Greater Accra to the neglect of other regions including Ashanti Region and for that matter the Kumasi Metropolis.

Suggestions for Further Study

1. Future researchers should consider conducting a similar study using other school categories such as B or C to confirm the findings of this study since the study used only category A schools in Kumasi Metropolis. In addition, the same study can be replicated in other districts or municipalities in Ghana.

2. Considering the variables that can moderate the effect of item sequence on students' academic performance, other researchers can consider investigating the moderating role of academic programme of students in the effect of item sequence. Furthermore, using private Senior High schools within Kumasi Metropolis as well as other districts or municipality for a similar study will be worthwhile.

The study considered Core Mathematics and Integrated Science. Further investigation can be done using other subject areas in order to obtain a more comprehensive view as well as generalizing the findings.

REFERENCES

Abbasian, G., & Zadsar, H. (2019). Relationship between test item arrangement and testee's performance and test usefulness criteria. *International Journal English Language Teaching, 6*(2), 52-70.

Adane, L. (2013). *Factors affecting low academic achievement of pupils in Kemp Methodist Junior High School in Aburi, Eastern Region*. (Unpublished Master's thesis), University of Ghana, Legon.

Adebule, S. O., & Awodele, B. A. (2015). Comparative effectiveness of conventional testing, liberal marking and confidence scoring on psychometric properties of chemistry multiple-choice test items in Osun state secondary school. *European Journal of Education Studies, 1*(1), 59-59.

Adusei, A. (2017). *A comparative study of the perceived learning strategies Junior and Senior High School students adopt when assessed with different item formats*. (Unpublished Master's thesis), Department of Education and Psychology. University of Cape Coast. Ghana.

Afful, R. (2014). *Perception of what assessment formats measure and learning strategies of students in senior high schools in Ajumako-Enyan-Essiam district, Ghana*. Cape Coast: University of Cape Coast Press.

Airasian, P. W. (2005). *Classroom assessment* (5th ed.). New York: The McGraw-Hill Companies, Inc.

Akinyi, O. 1. (2003). *An evaluation assessment of mathematics textbooks used in Kenyan secondary schools*. (Unpublished Master's dissertation), University of Nairobi, Nairobi, Kenya.

Alakayleh, S. A. (2017). Could different item arrangements affect 10th grade students' academic performance in multiple-choice tests in Maths, Science and English language final exams? *Journal of Education and Practice, 8*(7), 180-187.

Ali, M. S., & Mohsin, M. N. (2013). Test Anxiety Inventory (TAI): Factor analysis and psychometric properties. *Journal of Humanities and Social Science, 8*(1), 73-81.

Allam, S. (2007). *Measurement and Evaluation in teaching process*. Amman: Dar Almasira.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. USA: Waveland Press, Inc.

Amedahe, F. K. (1989). *Testing practices in secondary schools in the Central Region of Ghana*. (Unpublished Master's thesis), University of Cape Coast, Ghana.

Amedahe, F. K. (2000). *Continuous assessment*. (Unpublished paper), University of Cape Coast, Ghana.

Amedahe, F. K. (2004). *Notes on educational research*. (Unpublished document), University of Cape Coast, Ghana.

Amedahe, F. K., & Asamoah-Gyimah, K. (2016). *Introduction to measurement and evaluation*. Cape Coast: CCE Publications.

Amedahe, F. K., & Asamoah-Gyimah, K. (2016). *Introduction to measurement and evaluation* (6th ed.). Hampton press, Cape Coast.

Amedahe, F. K., & Asamoah-Gyimah, K. (2018). *Introduction to research methods in education*. Cape Coast: Hampton Press Ltd.

Amedahe, F. K., & Etsey, Y. K. A. (2003). *Assessment criteria, feedback and evaluation*. (Unpublished paper) presented at the University of Cape Coast, Ghana.

Amedahe, F. K., & Gyimah, K. A. (2003). *Measurement and evaluation.* Cape Coast, Ghana: Centre for Continuing Education.

American Federation of Teachers, National Council on Measurement in Education and National Education Association (AFT, NCME & NEA) (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice, 9*(4), 30-32.

Anamuah-Mensah, J. (2004). *Enhancing the teaching and learning of Ghana's science and technology for nation building*. GAST Annual Conference, Sekondi: Ghana.

Ananthakrishnan, N. (2000). Item analysis-validation and banking of MCQs. *Medical Education Principles and Practice Pondicherry: JIPMER,* 131-71.

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist, 36*, 1086-1093.

Ankomah, Y. A. (2002). The success of private basic schools in Ghana: The case of three schools in Cape Coast. *Journal of Educational Management, 4*, 14.

Ankomah, Y. A. (2011). A comparison of public and private basic school heads. *The African Symposium: An Online Journal of the African Educational Research Network, 11*(1), 41-56.

Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. (2006). *Introduction to the research in education* (7th ed.). New York, NY: Thompson/Wadsworth.

Ary, D., Jacobs, L.C & Razavieh, A. (2002). *Introduction to Research in Education* (6th ed.). Indiana: Indiana University Rotation. Retrieved from: http://wwwelsevier.com/locate.lindif.

Asamoah, D. (2018). *Causes of low academic performance of senior high school students in core Mathematics in the Kumasi Metropolis*. (Unpublished Master's thesis), University of Cape Coast, Ghana.

Asamoah, D. (2019). Development and validation of an instrument to measure test anxiety among students. *American Journal of Humanities and Social Sciences Research (AJHSSR), 3*(6), 203-208.

Asamoah-Gyimah, K., & Anane, E. (2019). *Research methods in education*. Cape Coast: University of Cape Coast Press.

Ashanti Regional Educational Office (2018). *Data on the statistics of SHS students*. Kumasi, Ghana: Ghana Education Service.

Attali, Y., & Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. *Journal of Educational Measurement, 40*(2), 109-128.

Balch, W. R. (1989). Item order affects performance on multiple-choice exams. *Teaching of Psychology, 16*, 75–77.

Balkis, M., & Duru, E. (2017). Gender differences in the relationship between academic procrastination, satisfaction with academic life and academic performance. *Electronic Journal of Research in Educational Psychology*, *15*(1), 105-125.

Baxter, I. E. (1998). *The Effect of question order on students' responses to multiple-choice physics questions*. Retrieved from: http://web.phys. Ksu.edu/…/msthesis.nd. di.

Benson, J. (2004). Developing a strong program of construct validation: A test anxiety example. *Educational measurement: Issues and Practice, 17*, 10-17.

Brame, C. (2013). *Writing good multiple-choice test questions.* Retrieved from https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/.

Bryman, A. (2001). *Quantitative data analysis for social scientists*. London: Routledge.

Burton, R. F. (2005). Multiple-choice and true/false tests: myths and misapprehensions. *Assessment and Evaluation in Higher Education, 30*(1), 65-72.

Cai, L., Choi, K., Hansen, M., & Harrell, L. (2016). Item response theory. *Annual. Rev. Stat. Appl. 3*(1), 297-321.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In E. Schmitt, W. C. Borman, & Associates (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco: Jossey-Bass.

Capraro, R. M., Roe, M. F., Caskey, M. M., Strahan, D., Bishop, P., Weiss, C., & Swanson, K. W. (2012). *Research summary: Assessment*. Association for Middle Level Education, 1-6.

Carlson, J. L., & Ostrosky, A. (1992). Item sequence and student performance on multiple choice exams: Further evidence. *Journal of Economic Education, 23,* 232–35.

Case, S. M., & Swanson, D. B. (2001). *Constructing observed to have effectiveness more than 50% written test questions for the basic and clinical sciences* (3rd ed.). Philadelphia: National Board of Conclusion Medical Examiners (NBME).

Cassady, J. C. (2004). The impact of cognitive test anxiety on text comprehension and recall in the absence of external evaluative pressure. *Applied Cognitive Psychology, 18*, 311 – 325.

Chavous, M. T. (2008). Evaluation (test) anxiety. *Psychology of classroom learning, 1*, 387-389.

Chen, H. (2012). The moderating effect of item order arranged by difficulty on the relationship between test anxiety and test performance. *Journal of Creative Education, 3*(3), 328-333.

Chidomere, R. C. (1989). Test item arrangement and student performance in principles of marketing examination: A replication study. *Journal of Marketing Education, 11* (Fall), 36-40.

Cobbinah, A. (2016). Item sequencing on difficulty level and students' achievement in Mathematics test in central region of Ghana. *African Journal of Interdisciplinary Studies, 9,* 55-62.

Cockcroft, W. H. (1982). *Mathematics counts*. London: HMSO.

Cohen, L., Manion. L., & Morrison, K. (2011). *Research methods in education*. London: Routledge.

Coniam, D. (1993). Does the ordering of questions in a test affect students' academic performance? *Educational Research Journal, 8*, 74-78.

Cook, T. D., & Campbell, D. T. (2001). *Quasi-experimentation: Design &analysis issues in field settings.* Boston, MA: Houghton.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed the methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Crocker, L., & Algina, J. (1986). *Introduction to item response theory*. Orlando, Florida: ERIC.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harpers and Brothers Publishers.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. St Leonards: Allen & Unwin.

Cunningham, G. K. (1986). *Educational and psychological measurement*. New York: Macmillan Publishing Company.

Danso, T. K. O. (2011). *Factors associated with students' academic performance at the basic education certificate examination in the Bia District of Ghana*. (Unpublished Master's thesis), University of Cape Coast, Ghana.

De Champlain, A. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Med Educ 44*: 109–117.

Declos, V. R., Vye, N., Burns, M. S., Bransford, J. D., & Hasselbring, T. S. (1992). Improving the quality of instruction: Roles for dynamic instruction. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (317-331). New York, NY: Spinger-Verlag.

DeVellis, R. F. (2016). *Scale development: Theory and applications, applied social research methods* (series 26). Newbury Park: Sage Publications.

Dimbisso, T. S. (2009). *Understanding female students' academic performance: An exploration of the situation in South Nations nationalities and peoples' regional state-Ethiopia*. A Research Paper (Presented in Partial fulfillment of the Requirements for obtaining the degree of Masters of Arts) in Development Studies, International Institute of Social Science, The Hague, The Netherlands.

Dunn, L., Morgan, C., O'Reilly, M., & Parry, S. (2004). *The student assessment handbook: New directions in traditional and online assessment*. New York: Routledge Falmer.

Dzakadzie, D. (2017). *Educational measurement, evaluation and statistics*. Kumasi-Ghana: Benjoy Ent.

Educational Management Information System (EMIS) (2019). *District performance in core subjects.* Accra: GES.

Embretson, S. E., & Reise, S. (2000). Calibrating Items: Estimation. In: Embretson SE, Reise S, (Eds.). *Item response theory for psychologists.* Mahwah, New Jersey: Lawrence Erlbaum Associates. pp. 187-225.

Etsey, Y. K. (2003). Pre-service teachers' knowledge of continuous assessment techniques in Ghana. *Journal of Educational Development and Practice, 1*(1), 1 - 8.

Ferla, J., Valcke, M., & Cai, Y. H. (2009). Academic self-efficacy and academic self-concept: reconsidering structural relationships. *Learning and Individual Differences, 19*(4), 499–505.

Gall, M. D., Borg, W. R., & Gall, J. P. (2003). *Educational research: An introduction* (6th ed.). White Plains, NY: Longman Publishers.

Galton, F. (1967). Classification of men according to their natural gifts. In J. A. Wiseman. (Ed.), *Intelligence and ability* (pp.34-40), Baltimore: Penguin.

Geiger, M. A., & Simons, K. A. (1994). Multiple-Choice Questions: Effect on exam performance and testing time. *Journal of Education for Business, 70*(2), 87–90.

Gerow, J. R. (1980). Performance on achievement tests as a function of the order of item difficulty. *Teaching of Psychology, 7,* 93-96.

Ghana Statistical Service (2014). *Population and housing report: District analytical report on the Kumasi Metropolitan Assembly*. Accra, Ghana: Ghana Statistical Service.

Goodrum, D., Hackling, M., & Rennie, L. (2000). *The status and quality of teaching and learning of science in Australian schools: A research report*. Canberra: Department of Education, Training and Youth affairs.

Gray, E. K. (2003). *The effect of question order on students' responses to multiple-choice physics questions*. (Unpublished thesis), Kansas State University, USA.

Gregory, R. J. (2016). *Psychological testing: History, principles and applications.* Boston: Allyn and Bacon.

Gronlund, N. E. (1985*). How to construct achievement tests* (3rd ed.). New Jersey: Prentice-Hall, Inc.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Taylor & Francis.

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), 309-333.

Hambleton, R., Swaminathan, H., & Rogers, J. (1991). Ability and item parameter estimation. In: Foster D, editor. *Fundamentals of item response theory*. California: Sage Publications.

Hayes, A. F. (2018). *Introduction to mediation, moderation and conditional process analysis: A regression-based approach* (2nd ed.). New York: The Guilford Press.

Hogan, T. P., & Murphy, G. (2007). Recommendations for preparing and scoring constructed response items: What the experts say. *Applied Measurement in Education, 20,* 427-441.

Hohensinn, C., & Kubinger, K. D. (2009). On varying item difficulty by changing the response format for a mathematical competence test. *Austrian Journal of Statistics, 38*, 231-239.

Ioannou-Georgiou, S. (2003). *Assessing your learners* (Resource Books for teachers). Oxford: OUP Print.

Jordan, A. M. (1953). *Measurement in education*. N.Y.: McGraw-Hill Book Co.

Kanfer, R. (2009). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Kassim, M. A., Hanafi, R. M., & Hancock, D. R. (2008). Test anxiety and its consequences on academic performance among university students. *Advance in Psychology Research, 53*, 75-95.

Kinyua, A., & Okunya, K. (2014). Validity and reliability of teacher-made tests: Case study of year 11 physics in Nyahururu District of Kenya. *African Educational Research Journal, 2*(2), 61-71.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*, 13-14.

Kubiszyn, T., & Borich, G. (2016). *Educational testing and measurement: Classroom application and practice*. New Jersey: Scott, Foresman and Company.

Leary, L. F., & Dorans, N. J (1982). *The effects of item rearrangement on test performance: A review of the literature* (ETS Research Report RR-82-30). Princeton, NJ: Educational Testing Service.

Leary, L. F., & Dorans, N. J. (1985). Implications for altering the context in which test items appear: A historical perspective on an immediate concern. *Review of Educational Research, 55*, 387-413.

Linn, R. E., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

Linn, R. L. & Gronlund, N. E. (2001). *Measurement and assessment in teaching* (7th ed). Englewood Cliffs, NJ: Prentice-Hall (1995).

Linver, M. R., Davis-Kean, P., & Eccles, J. E. (2002). *Influences of Gender on Academic Achievement*. A paper presented at the biennial meetings of the Society for Research on Adolescence, New Orleans, LA.

Livingston, S. A. (2006). Item analysis. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development.* New Jersey: Lawrence Erlbaum Associates.

Mahmud, J., Sutikno, M., & Naga, D. S. (2016). Variance difference between maximum likelihood estimation method and expected: A posteriori estimation method viewed from number of test items. *Educational Research and Reviews, 11*(16):1579-89.

Markman, U., Balik, C., Braunstein-Bercovitz, H., & Ehrenfeld, M. (2010). The effects of nursing students' health beliefs on their willingness to seek treatment for test anxiety. *Journal of Nursing Education, 50*, 248-251.

Martinez, M., E. (1999). Cognition and the question of test item format. *Educational Psychologist, 34* (4), 207-218.

Matlock-Hetzel, S. (1997). *Basic concepts in item and test analysis*. A paper presented at the annual meeting of the Southwest Educational Research Association, Austin. Retrieved from: http://ericae.net/ft/tamu/Espy.htm

McAlpine, M. (2002). *Principles of assessment*. Glasgow: CAA.

McMillan, H. J. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement Issues and Practice 20*(1):20-32

Mefor, C. (2014). *Nigeria: Identifying problems of poor performance in mathematics and way out*. Lagos, Nigeria: NERDC Press.

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology*. New York: Harcourt Brace College Publishers.

Mensah, J. K., Okyere M., & Kuranchie, A. (2013). Students' attitude towards mathematics and performance: Does the teacher attitude matter? *Journal of Education and Practice, 4*(3), 132-139.

Mereku, D. K. (1999). *School mathematics in Ghana: 1960-2000*. A paper delivered at the 6th National Biennial Delegates Conference of the Mathematical Association of Ghana (MAG), at St. Paul's Secondary School, Denu, Ghana, 10th to 13th August, 1999.

Mertens, D. M. (2010). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (3rd ed.). Thousand Oaks, CA: Sage.

Messick, S. (1989). Validity. In R. L. Lim (Ed). *Educational measurement* (3rd ed.). New York: Macmillan.

Miller, M. D., Linn, R. L., & Gronlund, N. F. (2009). *Measurement and assessment in teaching.* Upper Saddle River, NJ: Prentice Hall, Inc.

Moss, P. (2015). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*, 229-258.

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance, 10*, 71–83.

Munz, D. C., & Jacobs, P. D. (1971). An evaluation of perceived item-difficulty sequencing in academic testing. *British Journal of Educational Psychology, 41*, 195-205

Musa, A., Shaheen, S., Elmardi, A., & Ahmed, A. (2018). Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine, Khartoum University. *Khartoum Medical Journal, 11*(02), 1477-1486.

Nenty, H. J. (2005). Primary teachers' perceptions of classroom assessment practices as means of providing quality primary and basic education by Botswana and Nigeria. *Educational Research and Review, 2,* 74-81.

Nitko, A. (2004). *Educational assessment of students* (4th ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Nitko, A., & Brookhart, S. (2007). *Educational assessment of students*. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Notar, C. E., Zuelke, D. C., Wilson, J. D., & Yunker, B. D. (2004). The table of specification: Insuring accountability in teacher made tests. *Journal of Instructional Psychology, 31*, 115-129.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Ohuche, R. O., & Akeju, S. A. (1976). *Testing and evaluation in education*. African educational resources: Lagos.

Ollennu, S. N. (2011). *The impact of item position in multiple-choice test on student performance at the Basic Education Certificate Examination (BECE) Level.* (Unpublished Master's thesis), Department of Education and Psychology. University of Cape Coast. Ghana.

Ollennu, S. N., & Etsey, Y. K. (2015). The impact of item position in multiple-choice test on student performance at the Basic Education Certificate Examination (BECE) Level. *Universal Journal of Educational Research 3*(10): 718-723.

Oosterhof, A (2001). *Classroom applications of educational measurement*.

Opara, I. M., & Uwah, I. V. (2017). Effect of test items arrangement on performance in Mathematics among Junior Secondary School students in Obio/Akpor local government area of Rivers State Nigeria. *British Journal of Education, 5*(8), 1-9.

Ormrod, J. E. (2008). *Educational psychology: Developing learner* (6th ed.). New Jersey: Pearson Educational Inc.

Pallant, J. (2010). *SPSS survival manual* (4th ed.). New York, NY: McGraw Hill.

Paniagua, M. A., & Swygert, K. A. (2016). *Constructing written test questions for the basic and clinical sciences*. Philadelphia, PA: National Board of Medical Examiners.

Parr, A. M., & Bauer, W. (2006). *Teacher made test validity: A comparison of test scores and student study habits from Friday to Monday in a high school biology class in Monroe county, Ohio.* Master Thesis. Retrieved from http://etd.ohiolink.edu/sendpdf.cgi/parr%20anita.pdf Marietta 42864088.

Pecku, N. K. (2000). *Formal assessment in the classroom: The Ghana Education Service termly assessment plan*. Paper presented to the Quality Improvement in Primary Schools QUIPS) Project. Funded by the USAID.

Peters, D. L., & Messier, V. (1970). The effects of question sequence upon objective test performance. *Alberta Journal of Educational Research, 16* (4), 253-265.

Pettijohn, T. F., & Sacco, M. F. (2007). Multiple-Choice exam question order influences on student performance, completion time, and perceptions. *Journal of Instructional Psychology, 34* (3), 142-149.

Poehner, M. E. (2007). Beyond the test: L2 Dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal, 91*, 323–340.

Punch, K. F. (2008). *Introduction to research methods in education*. Thousand Oaks, CA: Sage Publications Ltd.

Putwain, D. W. (2008). Test anxiety and GCSE performance: The effect of gender and socio-economic background. *Educational Psychology in Practice, 24*, 319 – 334.

Rafiq, R., Ghazal, S., & Farooqi, Y. N. (2007). Test anxiety in students: Semester's vs. annual system. *Journal of Behavioural Science, 17*(2), 79-95.

Reiner, C. M., Bothell, T. W., Sudweeks, R. R., & Wood, S. B. (2002). *How to prepare effective essay questions: Guidelines for University Faculty*. Brigham: Brigham Young University Testing Services.

Rennie, L. J., Goodrum, D., & Hackling, M. (2001). science teaching and learning in Australian schools: Results of a National Study. *Research in Science Education, 31*, 455-498.

Republic of Ghana (1993). *Constitution of the republic of Ghana*. Retrieved from http://www.refworld.org/docid/3ae6b5850.html.

Roe, R. A. (2014). Work performance: A multiple regulation perspective. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology, 14*, 231-335.

Sa'ad, U. T., Adamu, A., & Sadiq, A. M. (2014). The causes of poor performance in mathematics among public senior secondary school students in Azare Metropolis of Bauchi State, Nigeria. *IOSR Journal of Research & Method in Education, 4*(6), 32-40.

Salvia, J., & Ysseldyke, J. E. (1978). *Assessment in special and remedial education*. Boston: Houton Mifflin Company.

Sarantakos, S. (2005). *Social research*. New York: Palgrave Macmilan.

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Columbia: Teachers College Press.

Shadreck, M., & Mambanda, I. (2012). Science teacher quality and effectiveness: Gweru Urban junior secondary school students' point of view. *Asian Social Science*, *8*(8), 121-135.

Shepard, L. A. (1994). The challenges of assessing young children appropriately. In Katheleen M. Cauley (12th ed.), *Educational Psychology*. Shefield: Dubuque Inc.

Sjøberg, S. (2002). *Three contributions to science education: Acta didactical*. Dept. of Teacher Education and School Development, University of Oslo.

Skinner, B. F. (1999). When the going gets tough, the tough gets going: Effects of item difficulty on multiple-choice test performance. *North American Journal of Psychology, 1*(1), 79 - 82.

Soureshjani, H. K. (2011). Item sequence on test performance: Easy item first? *Journal of Language Testing in Asia 1*(3), 46-59.

Steinberg, L., & Thissen, D. (2013). *Item response theory*. The Oxford handbook of research strategies for clinical psychology.

Stout, D. E., & Wygal, D. E. (1992). Additional empirical evidence on the relationship between exam question sequencing and accounting student performance. *Advances in Accounting, 8,* 133-52.

Tamakloe, E. K., Atta, E. T., & Amedahe, F. K. (2005). *Principles and methods of teaching.* Accra, Ghana: Ghana Universities Press.

Taub, A. J., & Bell, E. B. (1975). A bias in scores on multiple-form exams. *Journal of Economic Education, 7* (Fall), 58-59.

Tavakol, M., & Dennick, R. (2012). Post-examination interpretation of objective test data: Monitoring and improving the quality of high-stakes examinations. *Medical Teacher, 34*(3), 161-175.

Tei-Firstman, R. I. (2011) *Test item arrangement on student test scores*. M.Ed. (Unpublished thesis), University of Port Harcourt, Nigeria.

Tetteh, M. (2011). *Single-parenting and its effects on children's academic performance in Ashaiman*. (Unpublished Master's thesis), University of Ghana, Ghana.

Thissen, D., Wainer, H., & Wang, X. (1994). Are tests comprising both multiple-choice and free-response items necessarily less unidimensional than multiple-choice tests? An analysis of two tests. *Journal of Educational Measurement, 31*(2), 113-123.

Thorndike, E. L. (1903). *An introduction to the theory of mental and social measurements*. New York: The Science Press.

Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and Evaluation in Psychology and Education* (8th ed.). Boston: Pearson Education, Inc.

Togo, D. F. (2002). Topical sequencing of questions and advance organizers impacting on students' examination performance. *Accounting Education, 11*(3), 203–216.

Tshabalala, T., & Ncube, A. C. (2013). *Causes of poor performance of ordinary level pupils in mathematics in rural secondary schools in Nkayi District: Learner's attritions*. Onitsha, Nigeria: African Fab Publishers.

Umameh, M. A. (2011). *A survey of factors responsible for students' poor performance in mathematics in senior secondary school certificate examination (SSCE) in Idah Local government area of Kogi State, Nigeria*. Lagos, Nigeria: Text and Leisure Publishers.

UNDP (2005). *Human development report*. Oxford, Oxford University Press and UNDP. Retrieved from: http://www.undp.org. Upper Saddle River, New Jersey, Columbus, Ohio: Merill Prentice Hall Inc.

Vandee Schee, A. B. (2009). Test item order, academic achievement and students' academic performance on principles of marketing examinations. *Journal for Advancement of Marketing Education, 14*, 23-29.

Watering, G. V., Gijbels, D., Dochy, D., & Rijt, J. V. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education 56*(6): 645-658.

West African Examination Council (WAEC) (1993). *The effects of item position on performance in multiple choice tests*. Research Report, Research Division, WAEC, Lagos.

West African Examination Council (WAEC) (1995). *The effects of item position on performance in multiple choice tests*. Research Report, Research Division, WAEC, Lagos.

West African Examination Council (WAEC) (2019). *Statistics of WASSCE results in the core subjects from 2016-2019*. Retrieved from: https://www.waecgh.org/article/57/statistics-of-results-for-the-wassce-for-school-candidates-2016-2019/.

Wisconsin Education Association Council (1996). *Performance assessment. Madison, WI: Education issues series*. Retrieved from: http://www.educationforatoz.com.

World Bank (2004). *Books, buildings and learning outcomes: An impact evaluation of world Bank support to basic education in Ghana*. Washington, DC: World Bank.

Worthen, B., Borg, W., & White, K. (1993). *Measurement and evaluation in schools*. New York: Longman.

Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicol-Reflex Crit. 29*(1), 18-30.

APPENDICES

APPENDIX A

UNIVERSITY OF CAPE COAST

DEPARTMENT OF EDUCATION AND PSYCHOLOGY

ADAPTED CORE MATHEMATICS WAEC QUESTIONS FOR SHS FORM TWO STUDENTS

Dear Student,

I am a student of the University of Cape Coast conducting research on the effect of item sequencing in multiple-choice test on Senior High School students' academic performance in the Kumasi metropolis: the moderating role of gender. I humbly seek for your permission to participate in this study. The information you will provide shall be used for academic purpose only without revealing your identity. Please try your best to respond to all the questions. Any form of malpractice is discouraged.

Kindly indicate your choice by ticking (√) or writing your responses in the spaces provided

SECTION A

BACKGROUND INFORMATION

1. Gender:
   Male [ ] Female [ ]

2. Programme of Study:
   Business [ ]
   General Arts [ ]
   Home Economics [ ]
   Science [ ]
   Visual Arts [ ]
   Others (Specify) ……

SECTION B

INSRUCTION **TIME: 1 hour 30 minutes**

Answer all the question

Think carefully before you circle the answer. Do all rough work on this question paper.

Now, answer the following questions.

1. A circular pond of radius 4 *m* has a path of width 2.5 *m* round it. Find it correct to two decimal places, the area of the path. [Take $\pi = \frac{22}{7}$]
   a. 7.83 m^2
   b. 32.29 m^2
   c. 50.29 m^2
   d. 82.50 m^2
2. The roots of a quadratic equation are $-\frac{1}{2}$ and $-\frac{2}{3}$. *Find the equation*
   a. $6x^2$-$x + 2 = 0$
   b. $6x^2$-$x - 2 = 0$
   c. $6x^2$+$x - 2 = 0$
   d. $6x^2$-$x + 2 = 0$
   e. $6x^2 + x + 2 = 0$
3. Calculate the variance of 2, 3, 3, 4, 5, 5, 5, 7, 7 and 9.
   a. 2.2
   b. 3.4
   c. 4.0
   d. 4.2

4. Given that $t = 2^{-x}$, find 2^{x+1} *in terms of* t.

   a. $\frac{2}{t}$

   b. $\frac{t}{2}$

   c. $\frac{1}{2t}$

   d. 2t

5. The diagonal of a square is 60 *cm.* Calculate its perimeter.

   a. $20\sqrt{2}$

   b. $40\sqrt{2}$

   c. $90\sqrt{2}$

   d. $120\sqrt{2}$

6. Express, correct to three significant figures, 0.00359

   a. 0.00359

   b. 0.00360

   c. 0.004

   d. 0.359

7. In a certain community, 1 out of every 8 persons is a graduate. If Yomi and Etteh are members of the community, what is the probability that they are both graduates?

   a. $\frac{1}{64}$

   b. $\frac{1}{16}$

   c. $\frac{1}{8}$

   d. $\frac{1}{4}$

8. Three exterior angles of a polygon are 30^0, 40^0 and 60^0. If the remaining exterior angles are 46^0 each, name the polygon
    a. Decagon
    b. Nonagon
    c. Octagon
    d. Hexagon
9. The average age of a group of 25 girls is 10 years. If one girl, aged 12 years and 4 months joins the group, find correct to one decimal place, the new average age of the group.
    a. 10.1 years
    b. 9.3 years
    c. 8.7 years
    d. 8.3 years
10. Which of the following is true about parallelograms?
    a. Opposite angles are supplementary
    b. Opposite angles are complementary
    c. Opposite angles are equal
    d. Opposite angles are reflex angles
11. A rectangle board had length 15 cm and width x cm. if the sides are doubled, find its new area?
    a. $15x\ cm^2$
    b. $30\ x\ cm^2$
    c. $45x\ cm^2$
    d. $60x\ cm^2$

12. A sum of ₦18,100.00 was shared among 5 boys and 4 girls with each boy taking ₦20.00 more than each girl. Find a boy's share.
    a. ₦1,820.00
    b. ₦2,000.00
    c. ₦2,020.00
    d. ₦2,040.00
13. Bala sold an article for ₦ 6,900.00 and made a profit of 15%. Calculate his percentage profit if he had sold it for ₦ 6,600.00
    a. 13%
    b. 12%
    c. 10%
    d. 5%
14. The interior angles of a polygon are $3x^0$, $2x^0$, $4x^0$, $3x^0$ and $6x^0$. Find the size of the **smallest** angle of the polygon.
    a. 30^0
    b. 40^0
    c. 60^0
    d. 80^0
15. A bearing of 320^0 expressed as a compass bearing is
    a. $N\ 50^0\ W$
    b. $N\ 40^0\ W$
    c. $N\ 50^0\ E$
    d. $N\ 40^0\ E$

16. Two bottles are drawn **with replacement** from a crate containing 8 coke, 12 fanta and 4 sprite bottles. What is the probability that the first is coke and the second is not coke?

    a. $\frac{1}{12}$

    b. $\frac{1}{6}$

    c. $\frac{2}{9}$

    d. $\frac{3}{8}$

17. Make x the subject of the relation d=$\sqrt{\frac{6}{x}-\frac{y}{2}}$

    a. $x=\frac{6}{d^2}+\frac{12}{y}$

    b. $x=\frac{12}{2d^2-y}$

    c. $x=\frac{12}{y}-2d^2$

    d. $x=\frac{12}{2d^2+y}$

18. In what number base was the addition 1 +*nn* =100, *n* > 0, done?

    a. *n* -1

    b. *n*

    c. *n* + 1

    d. *n* + 2

19. Find the median age.

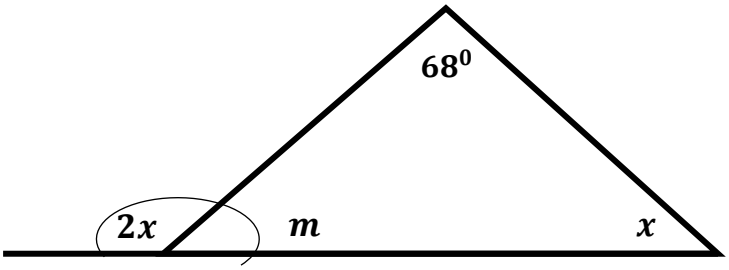
    a. 13mik

    b. 14

    c. 15

    d. 16

20.



Find the value of *m* in the diagram

a. 72^0

b. 68^0

c. 44^0

d. 34^0

21. If the simple interest on a certain amount of money saved in a bank for 5 years at $2\frac{1}{2}$ % per annum is ₦500.00, calculate the total amount due after 6 years at the same rate.

a. ₦ 2,500.00

b. ₦ 2,600.00

c. ₦ 4,500.00

d. ₦ 4,600.00

22. If 20(Mod 9) is equivalent to *y* (mod 6). Find *y*

a. 1

b. 2

c. 3

d. 4

23. If log x = 0.3030, log y = 0.4771 and log z = 0.8451, evaluate $\frac{\log x - \log Z^{\frac{1}{3}}}{\log y^3}$

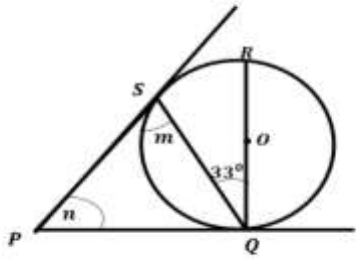
a. 1.4313

b. 0.8466

c. 0.5466

d. 0.0149

24.



In the diagram, PQ and PS are tangents to the circle centre *O*. If <PSQ=m, <SPQ=n and <SQR=33^0, find the value of (m + n)

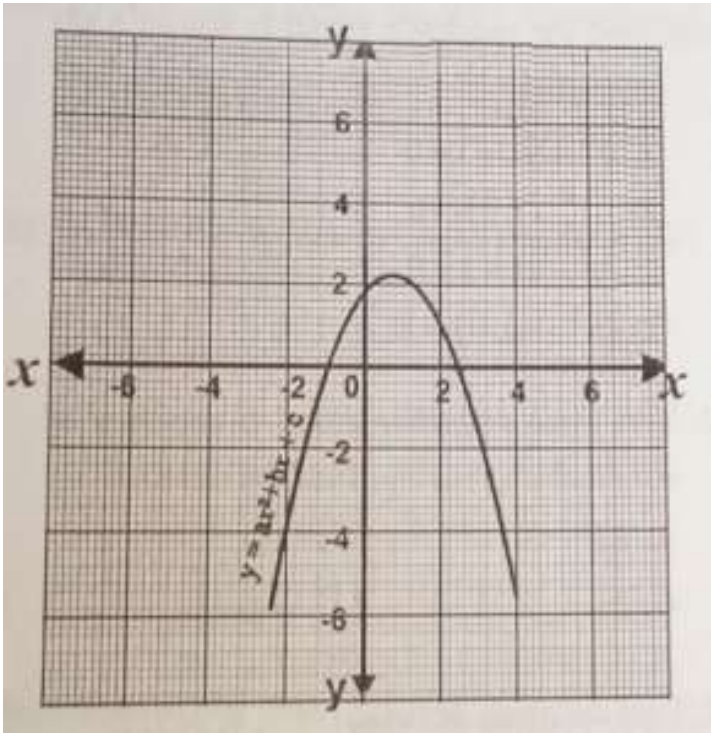
a. 103º

b. 123º

c. 133º

d. 143º

25.



Find the equation of the graph in the diagram

a. $y = 2-x-x^2$

b. $y = 2 – x + x^2$

c. $y = 2 + x – x^2$

d. $y = 2 + x + x^2$

26. Solve: $-\frac{1}{4} < \frac{3}{4}(3x - 2) < \frac{1}{2}$

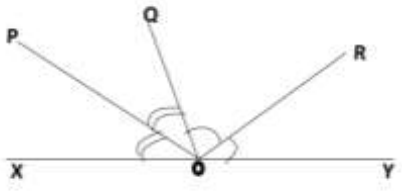
a. $\frac{5}{9} < x < \frac{8}{9}$

b. $-\frac{8}{9} < x < \frac{7}{9}$

c. $-\frac{8}{9} < x < \frac{5}{9}$

d. $-\frac{7}{9} < x < \frac{8}{9}$

27.



In the diagram, *XY* is a straight line, $<POX = <POQ$ and $<QOR$.

Find the value of $<POQ + <ROY$

a. 60^0

b. 90^0

c. 100^0

d. 120^0

28. Find the equation of a straight line passing through the point (1, -5) and having gradient of $\frac{3}{4}$.

a. $3x - 4y - 23 = 0$

b. $3x - 4y + 23 = 0$

c. $3x + 4y + 23 = 0$

d. $3x + 4y - 23 = 0$

29.

$\oplus$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Fig. 1

$\otimes$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Fig. 2

Fig. 1 and **Fig. 2** are the additional and multiplication tables respectively in modulo 5. Use these tables to solve the equation (n ⊗ 4) ⊕ 3 = 0 (mod 5)

a. 1

b. 2

c. 3

d. 4

30. Which of these values would make $\frac{3p-1}{p^2-p}$ $undefined$?

a. -1

b. $-\frac{1}{3}$

c. $\frac{1}{3}$

d. 1

31. If P (2, 3) and Q (2, 5) are points on a graph, calculate the length PQ.

a. 6 units

b. 5 units

c. 4 units

d. 2 units

32. In a class of 45 students, 23 offer Chemistry and 25 offer Biology. If each student offers at least one of the two subjects, calculate the probability that a student selected at random from the class offers Chemistry **only.**

    a. $\frac{2}{9}$

    b. $\frac{4}{9}$

    c. $\frac{5}{9}$

    d. $\frac{7}{9}$

33. Solve $4x^2 - 16x + 15 = 0$

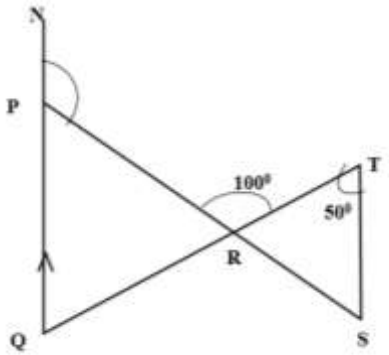
    a. $x = -1\frac{1}{2}$ of $-2\frac{1}{2}$

    b. $x = 1\frac{1}{2}$ or $-1\frac{1}{2}$

    c. $x = 1\frac{1}{2}$ or $2\frac{1}{2}$

    d. $x = 1\frac{1}{2}$ or $-2\frac{1}{2}$

34.



In diagram, *NQ//TS*, <RTS= 50^0 and <PRT=100^0. Find the value of <NPR.

a. 110^0

b. 130^0

c. 140^0

d. 150^0

35. Simplify: $(\frac{10\sqrt{3}}{\sqrt{5}} - \sqrt{15})^2$

a. 75.00

b. 15.00

c. 8.66

d. 3.87

36. Solve the equation: $\frac{1}{5x} + \frac{1}{x} = 3$

a. $\frac{1}{5}$

b. $\frac{2}{5}$

c. $\frac{3}{5}$

d. $\frac{4}{5}$

37. Simplify: $\sqrt{2}(\sqrt{6} + 2\sqrt{2}) - 2\sqrt{3}$

a. 4

b. $\sqrt{3} + 4$

c. $4\sqrt{2}$

d. $4\sqrt{3} + 4$

38. Given that cos 30^0 = sin 60^0 = $\frac{\sqrt{3}}{2}$ and sin 30^0 = cos 60^0 $\frac{1}{2}$, evaluate $\frac{tan60^0-1}{1-tan30^0}$

a. $\sqrt{3}-2$

b. $2-\sqrt{3}$

c. $\sqrt{3}$

d. -2

39. If x varies inversely as y and y varies directly as z, what is the relationship between x and z?

a. $x\ \alpha\ z$

b. $x\ \alpha\ \frac{1}{z}$

c. $x\ \alpha\ z^2$

d. $x\ \alpha\ \frac{1}{z^2}$

40. If x= $\frac{2}{3}$ and y = -6, evaluate $xy-\frac{y}{x}$

a. 0

b. 5

c. 8

d. 9

41. Evaluate $\frac{\log_3 9-\log_2 8}{\log_3 9}$

a. $-\frac{1}{2}$

b. $\frac{1}{3}$

c. $\frac{1}{2}$

d. $-\frac{1}{3}$

42. Evaluate: $\frac{3\frac{1}{4}\times 1\frac{3}{5}}{11\frac{1}{3}-5\frac{1}{3}}$

a. $\frac{14}{15}$

b. $\frac{13}{15}$

c. $\frac{4}{5}$

d. $\frac{11}{15}$

Age (Years)		13	14	15	16	17
Frequency		10	24	8	5	3

The table shows the ages of students in a club. Use it to answer questions 43 and 44.

43. How many students are in the club?

a. 50

b. 55

c. 60

d. 65

44. Find the median age.

a. 13

b. 14

c. 15

d. 16

45. What sum of money will amount to D10,400.00 in 5 years at 6% simple interest?

    a. D 8,000.00

    b. D 10,000.00

    c. D 12,000.00

    d. D 16,000.00

46. The marks of eight students in a test are: 10, 4, 5, 3, 14, 13, 16, and 7. Find the range.

    a. 16

    b. 14

    c. 13

    d. 1

47. IF 20(Mod 9) is equivalent to *y* (mod 6). Find *y*

    a. 1

    b. 2

    c. 3

    d. 4

48. A fair die is thrown once. What is the probability of obtaining 3?

    a. $\frac{1}{6}$

    b. $\frac{1}{3}$

    c. $\frac{1}{2}$

    d. 1

A straight line passes through the points P (1,2) and Q (5,8). Use this information to answer questions 24 and 25.

49. Calculate the gradient of the line PQ.

a. $\frac{3}{5}$

b. $\frac{2}{5}$

c. $\frac{3}{2}$

d. $\frac{5}{3}$

50. Calculate the length PQ.

a. $4\sqrt{11}$

b. $4\sqrt{10}$

c. $2\sqrt{17}$

d. $2\sqrt{13}$

APPENDIX B

UNIVERSITY OF CAPE COAST

DEPARTMENT OF EDUCATION AND PSYCHOLOGY

ADAPTED INTEGRATED SCIENCE WAEC QUESTIONS FOR SHS FORM TWO STUDENTS

Dear Student,

I am a student of the University of Cape Coast conducting research on the effect of item sequencing in multiple-choice test on Senior High School students' academic performance in the Kumasi metropolis: the moderating role of gender. I humbly seek for your permission for you to participate in this study. The information you will provide shall be used for academic purpose only without revealing your identity. Please try your best to respond to all the questions. Any form of malpractice is discouraged.

Kindly indicate your choice by ticking (√) or writing your responses in the spaces provided

SECTION A

BACKGROUND INFORMATION

3. Gender:
   Male [ ] Female [ ]

4. Programme of Study:
   Business [ ]
   General Arts [ ]
   Home Economics [ ]
   Science [ ]
   Visual Arts [ ]
   Others (Specify) ……

SECTION B

INSTRUCTION **TIME: 1 Hour**

Each question is followed by four options lettered A to D. Find out the correct option for each question and circle the correct answer on the question paper.

Do all rough work on this question.

Now answer the following questions.

1. In blood transfusion, a person with blood group O can receive blood from a person with blood group
    a. A
    b. B
    c. AB
    d. O
2. Which of the following metals will release hydrogen from dilute acids?
    a. Ag
    b. Cu
    c. Hg
    d. Zn
3. The structure in the human body which stores glycogen is
    a. Ileum
    b. Liver
    c. Lungs
    d. Pancreas

4. As advantage of the use of hay over silage is that hay
    a. Ensure the availability of feed all year round
    b. Has laxative effect.
    c. Has higher proportion of leaves, water and nutrients.
    d. Is more palatable to livestock.
5. Which of the following functions is/are associated with the endoplasmic reticulum of a living cell?
        I. Supports the ribosomes
        II. Provide large surface area for chemical reaction
        III. Provides mechanical support to the cytoplasm
    a. I only
    b. I and II only
    c. II and III only
    d. I, II and III
6. Which of the following indicators could be used to determine a pH of 7.4 of a solution?
    a. Litmus solution
    b. Methyl orange
    c. Phenolphthalein
    d. Universal indicator
7. In cattle, anaemia, damaged skin and heart water disease are caused by
    a. Liver-fluke
    b. Roundworm
    c. Tapeworm
    d. Tick

8. Aluminium is widely used in the manufacture of aircraft bodies because it is

    a. Heavy and has high tensile strength
    b. Light and has high thermal conductivity
    c. Heavy and has high electrical conductivity
    d. Light and resistant to corrosion

9. The application of lime to soil is to

    a. Improve the toxicity of dissolved copper.
    b. Improve the soil structure
    c. Increase the availability of calcium ions.
    d. Decrease the activities of soil living organisms.

10. Introduction of a trivalent element into a tetravalent element

    i. Is known as dropping
    ii. Forms an n-type semi-conductor
    iii. Forms a p-type semi-conductor

    Which of the statement above are correct?

    a. I and II only.
    b. I and III only.
    c. II and III only
    d. I, II and III.

11. Which of the following properties is associated with ionic compounds? They

    a. Are soluble in organic solvents
    b. Conduct electricity in the molten state
    c. Have low melting and boiling points
    d. Consist of molecules

12. In fruit flies, the allele for grey body colour is dominant and black colour is necessary. Two grey-bodies flies mated and they produced 75 grey-bodies flies and 25 black-bodies flies. The cross described is most likely to be

    a. GG x GG
    b. GG x gg
    c. Gg x GG
    d. Gg x Gg

13. Ectopic pregnancy in humans may result in

    I. Rupture of the fallopian tube
    II. Bleeding
    III. Weakness of the abdominal muscles

    Which of the statements above are correct?

    a. I and II only
    b. I and III only
    c. II and III only
    d. I, II and III

14. The rate of evaporation of a liquid could be increased by

    e. Decreasing the surface area
    f. Reducing the temperature
    g. Maintaining the temperature
    h. Removing the vapor as it forms

15. which of the following statement about breathing in humans is correct?
    a. The same amount of nitrogen, more carbon dioxide and less oxygen is breathed out than breathed in.
    b. The same amount of nitrogen, less carbon dioxide and more oxygen is breathed out than breathed in.
    c. More nitrogen, more carbon dioxide and less oxygen is breathed out than breathed in.
    d. Same amount of nitrogen, more carbon dioxide and more oxygen is breathed in that in breathed out.
16. Which of the following uses of substances is a are correct for esters?

    They are used as

        i. Solvent for quick-drying paints
        ii. Flavours in food preparations and cosmetics
        iii. Drying agents in making textiles.

    a. I only
    b. II only
    c. I and II only
    d. II and III only
17. The aim of the ring experiment in a flowering plant is to demonstrate that
    a. Water is transported by the xylem
    b. Food substances are transported by the phloem
    c. Food substances are formed in the stem
    d. Water is excreted from the stem

18. One similarity of arteries and capillaries is that both have

    a. Permeable walls

    b. Muscular walls

    c. No valves

    d. Large lumens

19. Which of the following animals is warm-blooded?

    a. Agama Lizard

    b. Wall gecko

    c. Cockroach

    d. Mouse

20. Which of the following statements explain why a boat floats in water?

    The boat floats because

    I. The up thrust on it equals the weight of water it displaces

    II. Its average density is less than that of the water

    III. Up thrust us equal to the volume of water displaced.

        a. I and II only

        b. I and III only

        c. II and III only

        d. I, II and III

21. The nuclide 226/88 Ra undergoes alpha decay, the mass number and the atomic number of the daughter nuclide are respectively

    a. 226, 89

    b. 226, 88

    c. 230, 90

    d. 222, 86

22. Which of the following factors affect the loudness of sound?

    a. Amplitude
    b. Frequency
    c. Velocity
    d. Wavelength

23. The type of soil erosion caused by eater run-off from exposed land on a gentle slope is referred to as

    a. Gulley erosion
    b. Rill erosion
    c. Sheet erosion
    d. Splash erosion

24. One disadvantage of mulching is that it

    a. May be a source of new weed species
    b. Acidifies the soil
    c. Improves soil texture
    d. Increase soil temperature

25. The gestation period of cow is approximately

    a. Seven months
    b. Eight months
    c. Ten months
    d. Eleven months

26. In testing for starch in a green leaf, the leaf is boiled in ethanol so as to

    a. Kill the leaf cells
    b. Dissolve the chlorophyll
    c. Enable iodine solution to penetrate the tissues
    d. Stop all chemical reaction in the cells.

27. Which of the following physical properties are associated with sandy soil?

    i. It is well aerated

    ii. It has a poor water holding capacity.

    iii. It is poorly drained

28. The concept of natural selection involves the following features except

    a. Survival of the fittest

    b. Competition among organism

    c. Organisms with the best adapted features

    d. The use of the concept of classify organisms.

29. The contraction of the biceps muscle in humans is likely to result in

    a. Straightening of the forearm

    b. Bending of the forearm

    c. Contraction of the triceps

    d. Movement of the scapula towards the stemum

30. The IUPAC name for $Na_2 SO_4$ is

    a. Sodium (II) sulphate

    b. Disodium sulphate

    c. Sodium tetraoxosulphate (IV)

    d. Sodium tetraoxosulphate (VI)

31. The commonest method for propagating crops in Ghana is by

    a. Budding

    b. Root cutting

    c. Seeds

    d. Stem cutting

32. The area occupied by a box placed on a table is 2.0 m. If the box exerts a pressure of 50 Pa on the table, calculate the mass of the box. [g=10ms^{-2}]

    a. 2.5 kg

    b. 10.0 kg

    c. 100.0 kg

    d. 1000.0 kg

33. Release of dust into the atmosphere

    i. Cut down photosynthesis

    ii. Causes respiratory diseases

    iii. Reduce visibility

    Which of the statement above is/are correct?

    a. I only

    b. I and II only

    c. I and III only

    d. I, II and III

34. The centre of gravity of a body is

    a. Always at its centre

    b. a point where its total mass is always concentrated

    c. always at its end

    d. a point where it can be supported to be in equilibrium

35. The chemical which is **most** likely to have **highly inflammable** hazard symbol in its container is

    a. Ethanol

    b. Mercury

    c. Ammonium hydroxide

    d. Sodium hydroxide

36. The atomic number of an atom is 15 and its mass number is 32.

    Determine the number of neutrons in the atom.

    a. 15

    b. 17

    c. 32

    d. 47

37. An energy saving electrical bulb is connected to a 240V power source.

    If the current passing through the bulb is 0.075 A, determine the power rating of the bulb

    a. 18W

    b. 32W

    c. 180W

    d. 320W

38. Which of the following reactions can produce salt and water **only?**

    a. $NaOH_{(aq)} + CH_2COOH_{(aq)}$

    b. $NaOH_{(aq)} + NH_{3(aq)}$

    c. $NH_{3(aq)} + H_2O$

    d. $NaCl_{(aq)} + AgNO_{3(aq)}$

39. A difference between an atom and an ion is that an atom

    a. Is electrically charged whilst an ion is neutral

    b. Has protons whilst an ion has protons

    c. Can move about in solution whilst ions cannot move in solution

    d. Cannot conduct electricity whilst an ion conducts electricity.

40. Rainbow formation is a practical illustration of the light phenomenon known as…..

    a. Dispersion
    b. Interference
    c. Mirage
    d. Reflection

41. Which of the following gases are major atmospheric pollution?

    i. Nitrogen dioxide
    ii. Oxygen
    iii. Chlorofluorocarbon

    a. I and II only
    b. I and III only
    c. II and III only
    d. I, II and III only

42. An effort of 120 N was used to overcome a load of 80 N. What is the mechanical advantage of the system

    a. 0.67
    b. 1.50
    c. 40.00
    d. 200.00

43. Dehorning is important in animal production because it

    a. Control endoparasite
    b. Makes it easy to handle animals
    c. Reduce the food intake of the animal
    d. Makes easy identification of animal

44. The arrangement of electrons in shells of an atom is called

    a. Atomic structure

    b. Periodic arrangement

    c. Electron configuration

    d. Electronic structure

45. Which of the following character tics is associated with insect-pollinated flowers?

    a. They do not secrete nectar

    b. The surface of the stigmas are sticky

    c. Their flowers are inconspicuous

    d. Their stamens have long filaments.

46. Acids react with many reactive metals to give off

    a. Hydrogen gas

    b. Oxygen gas

    c. Ammonia gas

    d. Carbon dioxide gas

47. Given that the Avogadro constant is 6.02×10^{23} mol^{-1}, determine the number of atoms in 0.01 mol of a substance.

    a. 6.02×10^{20} atoms

    b. 6.02×10^{21} atoms

    c. 6.02×10^{23} atoms

    d. 6.02×10^{24} atoms

48. What is the unit for momentum?

    a. Kg m^3

    b. Kg ms^{-1}

    c. Kg ms^{-2}

    d. $m\ s^{-1}$

49. The dimensions of a poultry house are 20m x 16m. If each bird occupies $0.25m^2$, determine the maximum number of birds the house can accommodate.

    a. 128

    b. 320

    c. 1280

    d. 1600

50. Which of the following method is a chemical means of birth control?

    a. Condom

    b. Diaphragm

    c. Pill

    d. Vasectomy

APPENDIX C

INTRODUCTORY LETTER

UNIVERSITY OF CAPE COAST

COLLEGE OF EDUCATION STUDIES
FACULTY OF EDUCATIONAL FOUNDATIONS
DEPARTMENT OF EDUCATION AND PSYCHOLOGY

Telephone: 0332091697
Email:dep@ucc.edu.gh

UNIVERSITY POST OFFICE
CAPE COAST, GHANA

4th October, 2019

Our Ref:

Your Ref:

TO WHOM IT MAY CONCERN

Dear Sir/Madam,

THESIS WORK
LETTER OF INTRODUCTION: MR. JOHN BAFFOE

We introduce to you Mr. Baffoe, a student from the University of Cape Coast, Department of Education and Psychology. He is pursuing Master of Philosophy Degree in Measurement and Evaluation. He is currently at the thesis stage.

Mr. Baffoe is researching on the topic:

"EFFECT OF ITEM SEQUENCING IN MULTIPLE-CHOICE TEST ON SENIOR HIGH SCHOOL STUDENTS' ACADEMIC PERFORMANCE IN THE KUMASI METROPOLIS: THE MODERATING ROLE OF GENDER"

He has opted to collect or gather data at your institution/establishment for his Thesis work. We would be most grateful if you could provide him the opportunity and assistance for the study. Any information provided would be treated strictly as confidential.

We sincerely appreciate your co-operation and assistance in this direction.

Thank you.

Yours faithfully,

Gloria Sagoe
Chief Administrative Assistant
For: **Head**

APPENDIX D

ETHICAL CLEARANCE FORM

UNIVERSITY OF CAPE COAST
COLLEGE OF EDUCATION STUDIES
ETHICAL REVIEW BOARD

UNIVERSITY POST OFFICE
CAPE COAST, GHANA

Our Ref: CES-ERB/ucc.edu/V4/20-11

Date: 12.03.2020

Your Ref: .....

Dear Sir/Madam,

ETHICAL REQUIREMENTS CLEARANCE FOR RESEARCH STUDY

The bearer, John Baffoe....., Reg. No EF/MEP/18/0010 is an M.Phil. / ~~Ph.D~~ student in the Department of Education..... and..... Psychology..... in the College of Education Studies, University of Cape Coast, Cape Coast, Ghana. He / ~~She~~ wishes to undertake a research study on the topic:

Item sequencing in multiple-choice tests and students' performance in the West African Secondary School Certificate Examinations: The moderating role of Gender

The Ethical Review Board (ERB) of the College of Education Studies (CES) has assessed his/~~her~~ proposal and confirm that the proposal satisfies the College's ethical requirements for the conduct of the study.

In view of the above, the researcher has been cleared and given approval to commence his/~~her~~ study. The ERB would be grateful if you would give him/~~her~~ the necessary assistance to facilitate the conduct of the said research.

Thank you.
Yours faithfully,

Prof. Linda Dzama Forde
(Secretary, CES-ERB)

Chairman, CES-ERB
Prof. J. A. Omotosho
jomotosho@ucc.edu.gh
0243784739

Vice-Chairman, CES-ERB
Prof. K. Edjah
kedjah@ucc.edu.gh
0244742357

Secretary, CES-ERB
Prof. Linda Dzama Forde
lforde@ucc.edu.gh
0244786680

APPENDIX E

MAUCHLY TEST OF SPHERICITY FOR CORE MATHEMATICS

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	Df	Sig.	Epsilon Greenhouse-Geisser	Epsilon Huynh-Feldt	Epsilon Lower-bound
Item sequence	.987	4.111	2	.128	.987	.993	.500

APPENDIX F

MAUCHLY TEST OF SPHERICITY FOR INTEGRATED SCIENCE

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	Df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Item Sequence	.794	3.962	2	.0570	.829	.833	.500

APPENDIX G

NORMALITY TEST FOR ITEM SEQUENCE IN CORE MATHEMATICS AND STUDENTS' PERFORMNACE

Tests of Normality

	GENDER OF STUDENTS	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	df	Sig.
MATHS HTE SCORES	MALE	.092	146	.004	.962	146	.000
	FEMALE	.114	165	.000	.966	165	.000
MATHS RANDOM SCORES	MALE	.142	146	.000	.957	146	.000
	FEMALE	.092	165	.002	.975	165	.005
MATHS ETH SCORES	MALE	.083	146	.015	.988	146	.232
	FEMALE	.064	165	.091	.991	165	.377

APPENDIX H

BOX'S M TEST FOR CORE MATHEMATICS

Box's Test of Equality of Covariance Matrices

Box's M	7.086
F	1.168
df1	6
df2	664683.686
Sig.	.320

APPENDIX I

NORMALITY TEST FOR ITEM SEQUENCE IN INTEGRATED SCIENCE AND STUDENTS' PERFORMNACE

Tests of Normality

	GENDER OF STUDENTS	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	Df	Sig.
SCI HTE SCORES	MALE	.077	146	.033	.990	146	.375
	FEMALE	.082	165	.009	.979	165	.015
SCI RDM SCORES	MALE	.077	146	.034	.978	146	.020
	FEMALE	.106	165	.000	.958	165	.000
SCI ETH SCORES	MALE	.099	146	.001	.975	146	.008
	FEMALE	.135	165	.000	.970	165	.001

APPENDIX J

BOX'S M TEST FOR INTEGRATED SCIENCE

Box's Test of Equality of Covariance Matrices

Box's M	5.244
F	.865
df1	6
df2	664683.686
Sig.	.520

APPENDIX K

ITEM DIFFICULTY INDICES OF CORE MATHEMATICS ITEMS

Question Number	Difficulty Index	Question Number	Difficulty Index
1	0.30	26	0.72
2	0.82	27	0.72
3	0.72	28	0.36
4	0.58	29	0.74
5	0.72	30	0.37
6	0.67	31	0.61
7	0.68	32	0.34
8	0.32	33	0.74
9	0.74	34	0.36
10	0.66	35	0.71
11	0.32	36	0.72
12	0.45	37	0.39
13	0.58	38	0.31
14	0.30	39	0.30
15	0.32	40	0.34
16	0.74	41	0.74
17	0.47	42	0.31
18	0.36	43	0.57
19	0.34	44	0.32
20	0.47	45	0.47
21	0.45	46	0.36
22	0.45	47	0.62
23	0.45	48	0.72
24	0.59	49	0.42
25	0.32	50	0.59

APPENDIX L

ITEM DIFFICULTY INDICES OF INTEGRATED SCIENCE ITEMS

Question Number	Difficulty Index	Question Number	Difficulty Index
1	0.35	26	0.73
2	0.56	27	0.45
3	0.72	28	0.70
4	0.69	29	0.66
5	0.55	30	0.68
6	0.35	31	0.39
7	0.65	32	0.61
8	0.66	33	0.69
9	0.61	34	0.71
10	0.69	35	0.69
11	0.72	36	0.74
12	0.50	37	0.68
13	0.66	38	0.74
14	0.71	39	0.67
15	0.70	40	0.68
16	0.62	41	0.55
17	0.49	42	0.64
18	0.69	43	0.65
19	0.68	44	0.61
20	0.63	45	0.51
21	0.70	46	0.66
22	0.74	47	0.69
23	0.71	48	0.71
24	0.33	49	0.70
25	0.68	50	0.55