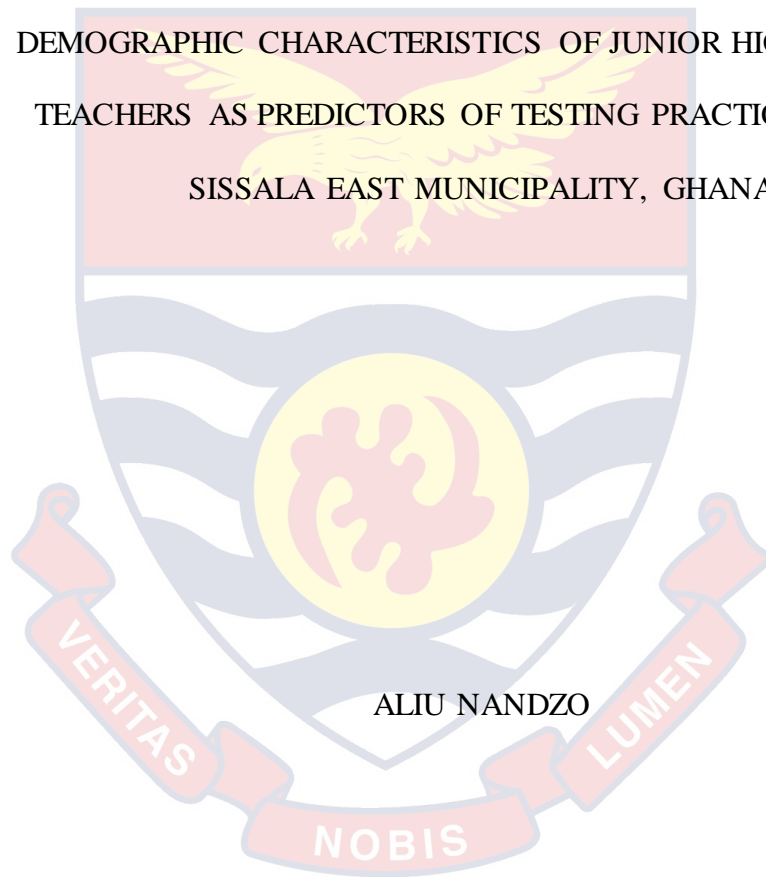


UNIVERSITY OF CAPE COAST

DEMOGRAPHIC CHARACTERISTICS OF JUNIOR HIGH SCHOOL
TEACHERS AS PREDICTORS OF TESTING PRACTICES IN THE
SISSALA EAST MUNICIPALITY, GHANA

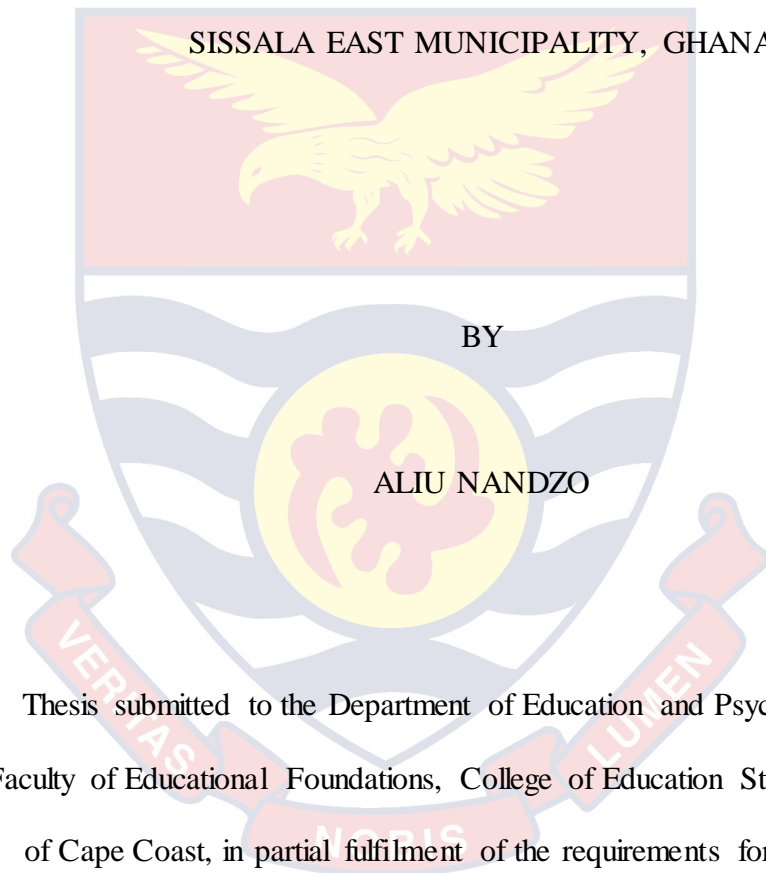


ALIU NANDZO

2021

UNIVERSITY OF CAPE COAST

DEMOGRAPHIC CHARACTERISTICS OF JUNIOR HIGH SCHOOL
TEACHERS AS PREDICTORS OF TESTING PRACTICES IN THE
SISSALA EAST MUNICIPALITY, GHANA



BY

ALIU NANDZO

This 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 fulfillment of the requirements for the award of Master of Philosophy degree in Measurement and Evaluation

JANUARY 2021

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result 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:

Supervisor's Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision 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 teachers' demographic characteristics of Junior High School teachers as predictors of testing practices in the Sissala East Municipality. The study was conducted by employing the descriptive cross-sectional survey design. The target population for the study was basic school teachers totalling, 135 in the Sissala East Municipality. The accessible population was teachers who teach English Language, Mathematics, Integrated Science and Social Studies in the 78 Junior High schools (JHS). Through a multistage sampling procedure, a sample of 55 schools, comprising 220 teachers were used as respondents. The study adapted a Questionnaire which had three hypothetical dimensions for the data collection. The data collected were analysed using, frequencies, percentages, mean, standard deviation, multivariate multiple regression and the hierarchical multiple regression. It was found that demographic characteristics of teachers: gender, years of teachers, academic qualification and subject teachers teach, individually, did not significantly predict teachers' testing practices. The study also, revealed that teachers' demographic characteristics, combined together, did not contribute significantly to the way teachers practise testing. It was concluded that teachers' demographic characteristics do not contribute to adherence the basic principles of testing. The Directorate of the Ghana Education Service (GES) in the Sissala East Municipality is encouraged, as part of their training and development programmes, to continue to organise special seminars and workshops, which would sensitise teachers in the municipality on the essence of developing quality classroom assessments.

KEY WORDS

Assessment

Test

Testing practice



ACKNOWLEDGEMENTS

I would like to express my utmost gratitude to Dr. Kenneth Asamoah-Gyimah of the Department of Education and Psychology, who, as my principal supervisor made profound contributions to the work, and offered expert guidance, advice and encouragement towards the success of this research work. Sir, I am really grateful.

I wish to also recognise and acknowledge my sincere indebtedness to Dr. Eric Anane of Institute of Education, who, as my co-supervisor made great contribution, supported and provided expert and professional guidance in line with the supervision of this research work. Sir, I say a very big thank you.

I am grateful to Dr. Ruth Keziah Anan-Brew, Mr. Francis Ankomah and Mr. Isaac Amoako for the assistance and motivational insight they gave me throughout this research work. I also acknowledge the Directorate and the various heads of schools of the Sissala East Municipality, for their warm reception and support during the data collection. To the teachers who took some time to respond to the questionnaires, I appreciate the co-operation you gave me.

I express my appreciation and gratitude to Miss Kafui Foli and Mr. Godwin K. Ledi, for your encouragement and support throughout my work. Finally, I wish to acknowledge and thank Mrs. Zuleyhatu Nandzo, my wife, for all the sacrifices you made for me throughout this research work. I say Almighty Allah bless you bountifully.

DEDICATION

To the Nandzo' family of Fuojan – Bujan.



TABLE OF CONTENTS

	Page
DECLARATION	ii
ABSTRACT	iii
KEY WORDS	iv
ACKNOWLEDGEMENTS	v
DEDICATION	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER ONE: INTRODUCTION	
Background to the Study	2
Statement of the Problem	12
Purpose of the Study	13
Research Hypotheses	14
Significance of the Study	15
Delimitations	16
Limitations	16
Assumptions of the Study	17
Definition of terms	18
Organisation of the Study	18
CHAPTER TWO: LITERATURE REVIEW	
Introduction	19
Theoretical Review	19
Classroom Assessment Environment Model (Brookhart, 1997)	19

Classical True Score Theory	21
Assumptions of the Classical True Score Theory	23
The link between Classical Test Theory and this study	24
Conceptual Review	25
Brief History of the Development of Testing	25
Construction of Classroom Achievement Tests	28
Validity of Test Items	35
Validity Evidence	37
The Concept of Content Validity	37
Criterion-Related Validity	39
Construct Validity	40
Principles of Constructing Classroom or Teacher-Made Tests	41
Administration of Classroom Achievement Tests	44
Guidelines in Administering Achievement Tests	47
Scoring of Classroom Achievement Tests	48
Empirical Review	52
Teachers Assessment Practices	52
Gender and Classroom Assessment Practices	58
Years of Teaching and Assessment Practices	61
Teacher Qualification and Assessment Practices	66
Subject Teachers Teach and Assessment Practices	68
Conceptual Framework	70
Summary of Literature Review	71
CHAPTER THREE: RESEARCH METHODS	
Research Design	73

Study Area	75
Population	76
Sample and Sampling Procedure	77
Data Collection Instrument	79
Validity and Reliability of the Instrument	82
Pilot-Testing	83
Data Collection Procedure	84
Data Processing and Analysis	86
CHAPTER FOUR: RESULTS AND DISCUSSION	
Demographic Characteristics of Respondents	87
Distribution of Respondents Based on Gender	88
Distribution of Respondents Based on Years of Teaching	89
Distribution of Respondents Based on Educational Qualification	89
Distribution of Respondents Based on Subjects Teachers are Teaching	90
Testing Practices of Teachers	90
Test construction Practices of Teachers	90
Test Administration Practices of Teachers	93
Test Scoring Practices of Teachers	96
Hypotheses Testing	98
Discussions	120
Gender of Teacher and Testing Practices	120
Years of Teaching and Testing Practices	122
Qualification of Teacher and Testing Practices	124
Subject Teachers Teach and Testing Practices	126

Gender, years of teaching, qualification, and subject teacher teach and testing practices	128
Chapter Summary	129
CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS	
Summary	130
Overview of the study	130
Key Findings	131
Conclusions	131
Recommendations	132
Suggestions for Future Research	133
REFERENCES	134
APPENDICES	147
APPENDIX A: Introductory Letter	147
APPENDIX B: Ethical Clearance	148
APPENDIX C: Questionnaire	149
APPENDIX D: Test for Normality	154
APPENDIX E: Test for Multivariate Normality	156
APPENDIX F: Results Output for Hypothesis One	158
APPENDIX G: Results Output for Hypothesis Two	160
APPENDIX H: Results Output for Hypothesis Three	163
APPENDIX I: Results Output for Hypothesis Four	166

LIST OF TABLES

Table	Page
1 Distribution of Respondents by Demographic Characteristics (n =144)	88
2 Test Construction Practices of Teachers	91
3 Test Administration Practices of Teachers	94
4 Test Scoring Practices of Teachers	97
5 Test for Normality	99
6 Parameter Estimates for Test Practices	103
7 Parameter Estimates for Test Practices	106
8 Parameter Estimates for Test Practices	110
9 Parameter Estimates for Testing Practices	113
10 Hierarchical Regression of Predictor variable and Testing Practices (Construction)	115
11 Regression Coefficients for Predictor Variables (construction)	116
12 Hierarchical Regression of Predictor Variables and Testing Practices (Administration)	117
13 Regression Coefficients for Predictor Variable (administration)	118
14 Hierarchical Regression of Predictor Variables and Testing Practices (Scoring)	119
15 Regression Coefficients for Predictor Variable (scoring)	119

LIST OF FIGURES

Figure		Page
1	Conceptual framework for teachers' demographic characteristics on teachers' testing practices	71
2	Test for Normality for Construction	100
3	Test of Normality for Administration	100
4	Test of Normality for Scoring	101



CHAPTER ONE

INTRODUCTION

In schools, the responsibility of carrying out assessments, rests, largely, on the shoulders of the classroom teacher. Assessment plays a critical role towards the teaching and learning process. Hence, every teacher is required to take at least a semester course in educational assessment, as part of his/her pre-service training. However, over several decades now, researchers globally, continue to bemoan the assessment practices of teachers. For instance, it has been found that teachers, as part of their practice of assessment, either construct poor items or often rely on past questions to assess their students (Ololube, 2008; Onyechere, 2000; Frey & Schmitt, 2010).

In Ghana, some previous studies have found poor assessment practices (e.g., non-adherence to testing principles; test construction, test administration and test scoring etc.) among teachers (Quagrain, 1992; Oduro-Okyireh, 2008; Anhwere, 2009; Quansah, Amoako & Ankomah, 2018; Quansah, Amoako & Ankomah, 2019).

Some researchers have found that, besides teachers' knowledge in educational assessment, their demographic characteristics such as gender, years of teaching, educational qualification, and subject area, etc. have a relationship with the way teachers practice assessment (Adeyemi, 2010; Koli-Keaikitse, 2012; Alkharusi, 2011; Owusu-Mensah, 2019). Therefore, to have a comprehensive understanding of teachers' assessment practices, there

is the need to examine the contribution of teachers' demographic characteristics towards their assessment (testing) practices.

Background to the Study

Assessment refers to a broad term which includes a well organised collection of information that is useful for making decisions about the properties of individuals, events and objects (Reynolds, Livingstone, Wilson, 2009). Assessment includes not only data collection, but also a method that test users use to determine the knowledge, comprehension, capabilities of students in accordance with a well-defined set of planned learning outcomes or prescribed study program (Marriot & Lau, 2008). Assessment as a concept is delineated to five major dimensions; (1) why assessment is carried out? This dimension focuses on the essence of assessment in terms of why it is conducted and what results it will yield. (2) what aspect of the learning target to be assessed? Thus, coming to terms with the reasons for people to be assessed. (3) how to effectively carry out the assessment? Choosing from among available alternative ways of performing an assessment, considered to be accurate and appropriate for different valued information. (4) how to offer appropriate interpretation to the assessment result? Making meaning from the observations or measurement we carry out through different assessment procedures. (5) How to respond / find effective ways to express our reaction to the observations that are made and the best to convey it to the individuals concerned (Rowntree 1987, p. 11).

These factors feed into the structure in which classroom assessment practices should be conceptualised. All these point to the fact that classroom assessments play an integral part of the teaching and learning learning. Several

techniques are available for carrying out classroom assessment. Through these approaches, teachers will gain an insight into the thinking of the students. Based on these tests, teachers gather different types of information to make knowledgeable, accurate and reliable correct evaluations of students' learning outcomes (Koloi-Keaikitse, 2012).

The importance of assessment in our educational system can not be overemphasised, as it is an approach that evaluates both the learner and the teacher's classroom behaviours and results. It is a strong mechanism that, depending on how it is implemented, can either enhance or hinder learning. Hence, assessment is often viewed as the bedrock of teaching and learning in every educational setting. Assessment is all the activities that teachers use to help students learn and judge progress (Opara, Onyekuru & Njoku (2015).

In the conduct of classroom assessment, teachers use different strategies to test the knowledge of students in order to gauge the level of mastery or the amount of learning achieved, in relation to the predetermined specific subjects. Assessment strategies are broadly categorised into two; traditional types of assessment and performance-based assessment (Birenaum, & Feldman, 1998). Birenaum, and Feldman (1998), pointed out that the traditional type of assessment method is used when teachers intend to assess knowledge-based objectives. Examples of such methods include multiple-choice questions, short answer essays or structured answers and standardised tests but use the performance-based assessments when students are required to perform a task rather than choose from the options provided, and students are assessed on the basis of their performance outcomes and the extent to which those outcomes relate to the rubrics or feedback instruments. Based on the

different focus and applicability of these two types of classroom assessment, it is important to note that for a teacher to organise classroom assessment that is valid and reliable, the teacher should have the knowledge and commitment towards the planning, construction, administration, scoring and communicating assessment feedback to students, in line with the guidelines espoused by Measurement and Evaluation Experts.

Practically speaking, education and assessment must be deemed to be two sides of the same coin, and therefore, are inseparable and almost impossible for anybody to study in educational system without being exposed to several forms of educational and psychological assessment procedures. This is so, because always in an educational system, some important decisions have to be made about students, curricula and programmes, and educational policies (Nitko, 1996). Nitko points out that, such important decisions include: decisions on students include handling teaching in the classroom, placing students in various types of programs, placing them in suitable groups, directing and counselling them, selecting them and certifying their competence for educational opportunities and qualifications.

Also, Curriculum and program decisions include conclusions about how effective the students learning or curriculum has been (summative assessment) and ways of improving them (formative assessment). Decisions on educational policies are determined at national level in Ghana. However, it is worth noting that educational assessments are dominated by the use of tests in the Ghanaian education system, in order to provide some of the information needed for these types of decisions. The responsibility of obtaining accurate information on the learning progress of each student, especially, at the basic

level, directly rests on the shoulders of the teacher, who in the case of Ghana mostly uses tests for the assessment.

According to the Standard for Educational and Psychological testing, National Council on Measurement in Education (NCME, 2014) “a test is a device or procedure in which a sample of an examinees behaviour in a specified domain is obtained and subsequently evaluated and scored using a standardized process” (p. 2). The examinee’s behaviour in this case, means some specific psychological characteristics that are believed to be imbibed in the student after certain period of well planned and executed lessons by the teacher or instructor. It must be noted, however, that an individual’s psychological attributes cannot be directly measured like other physical attributes of individuals, such as height or weight. That is to say that such measures only indicate an estimation of the presence of psychological constructs and hence, one may never be able to absolutely confirm that such a psychological construct exists.

Teachers in the educational setting would want to estimate the degree to which their students are characterised by the knowledge they have imparted to them within a given period. It is worthy to note that, not all the psychological attributes of such constructs may be observed by a single test. Nonetheless, a well-constructed test may essentially evaluate a reasonable amount of the construct from which to make a value judgment. Test is an essential tool that helps to quantify such constructs which help one to make a value judgment about the degree to which such constructs might probably exist in an individual, and therefore, educators and teachers must also be

mindful that the exam itself is prone to mistakes that could adversely affect its use in decision-making about students' learning.

Tests are essential tools in every educational system (Quaigrain, 1992). That is to say simply, that tests and teaching are interlaced. Quaigrain has stated that tests provide needed information for assessment. Without assessment, there cannot be feedback and knowledge of results. Without knowledge of results there cannot be any systematic improvement in learning. Standardised achievement, aptitude, and intelligence tests seen in developed countries like the United States of America (USA), Canada, and Great Britain are essentially non-existent in the Ghanaian educational system. The assessments carried out at the terminal points of the educational system by the West African Examinations Council (WAEC) cannot be said to be standardised because they do not follow all the basic characteristics of standardised performance tests. The basic education certificate examination (BECE) and the Senior Secondary School Certificate Examination (SSSCE) are examples of the WAEC tests conducted in Ghana.

Unless students' learning levels are properly measured and interpreted, teachers and school administrators will not be able to provide the right educational opportunities and support each student's needs. Testing provides feedback on which educational decisions are made. These decisions may require information on the success of learning programmes or students who have attained specific skills and knowledge levels (Izard, 2005). It is generally believed that, accurate and valid information about students' achievement is essential for effective instruction, as it allows teachers to provide appropriate feedback and adapt their instruction to suit student needs.

Classroom assessment is an essential part of the teaching and learning process in every educational setting, as it helps teachers make decisions in the classroom. These assessments are not only a way of awarding grades and deciding whether students achieve goals but also a learning mechanism (Watering, Gijbels, Dochy, & Rijt, 2008). In education, learning is the ultimate dependent variable (Nenty, 1992), everything done in the name of education has a positive or negative impact on learning. Nenty argues that it is through assessment that learning and hence quality education is defined, so any improvement in these ultimately depends on assessment quality. Education should be seen as a tool that can improve the global skills of young people, placing them in an advantageous position among emerging global children.

Tony Wagner (as cited in Conlan, Posner, & Conlon, 2008), embolden the call of these researchers by listing seven survival skills of the 21st century; critical thinking and problem solving, collaboration using different networks and providing leadership through influence, agility and flexibility, initiative and business mindedness, effective spoken and written communication, accessing and analysing information, Curiosity and imagination. To laden his call, Tony Wagner opines that for survival in the 21st century, we:

expect all teachers to teach all students how to think and communicate effectively, and they need to assess these skills and benchmark expectations to what the world will require of our high school graduates. And this needs to happen every day in every class and at all grade levels. If we do this in all of our schools, while also stimulating curiosity and imagination, then

all students will have the skills they need to get and keep a good job and be a contributing citizen, while our country will have a workforce that can continually produce innovations. An economy based on innovation will be more competitive and successful than any other in the 21st century Conlan, Posner, & Conlon, 2008, pp. 1).

This means that assessment is viewed as tool or a double-edged sword that is used by the teacher on daily basis, to gauge the progress of the learner. Therefore, depending on how the teacher uses it on daily basis, it can lead to helping the learners to maximise their potentials, if it is used effectively or it can lead to a shuttered potential, if it is abused.

In line with this global vision, Ghana, keen on promoting education and with particular emphasis on quality education for all, has recently rolled out a new national curriculum system for pre-tertiary education. The new curriculum has incorporated a number of key skills as part of Ghana's national curriculum goals and expectations, including: Focus on the 4Rs: Reading, writing, arithmetic and creativity as a catalyst for achieving rapid sustainable developmental changes, Focus on the essential knowledge, skills and competences that Ghana's young people need to become educated citizens, Focus on mathematics and science as the fundamental building blocks for success in the era of technological advancement and Ensuring that basic school lays the solid foundation necessary for tertiary education as well as preparations for early entry into the work place (National Pre-Tertiary Education Curriculum Framework, 2018, P. 7).

Considering the critical role classroom assessment plays in ensuring that every student at every stage of the educational ladder, maximises his or her potential, several researchers globally, have studied and continue to explore how teachers go about classroom assessment as well as some issues that possibly impact such classroom activities of instructors all over the world. Regrettably, several investigations into the assessment practices of teachers found that teachers have poor test construction skills. It has been observed that Such teachers either construct poor items or often rely on past questions to assess their students (Ololube, 2008; Onyechere, 2000; Frey & Schmitt, 2010).

Similarly, in Ghana, the classroom assessment practices of teachers have been studied and the results showed that Ghanaian teachers have limited skills in constructing the objective and essay type tests, (Quagrains, 1992; Oduro-Okyireh, 2008; Anhwere, 2009; Quansah, Amoako & Ankomah, 2018; Quansah, Amoako & Ankomah, 2019). The findings of these researchers suggest that teachers in Ghana, at all levels of the educational ladder have reported that their knowledge in assessment is low and therefore do not follow basic principles of assessment. Other teachers have reported having negative attitude towards classroom assessment.

The teacher is a significant pillar in the achievement of every nation's educational goals. Teachers have been shown to have a significant impact on the academic fortunes of the students at all levels of the educational process. That is to say that, teachers play a crucial role in the realisation of educational goals of any nation, part of the teacher's core duties is to break down policy (curriculum) into actionable activities based on values while they engage students on daily basis (Afe, 2001). The teacher serves as a conduit through

which knowledge, values and skills are passed on to the student in the learning process. This therefore, means that if the teacher is unable to carry out this responsibility successfully, students will find it difficult to make progress academically under the tutelage of the teacher. It is worth knowing that measuring the learning process and development can only be evaluated meaningfully with a proper classroom assessment, of which the test typically dominates.

The ability of the teacher to succeed in effectively carrying out classroom assessment is largely influenced by the knowledge and skills in assessment that he or she has acquired, as part of his or her initial training in an educational institution. For example, every professionally trained teacher in the Ghanaian classroom has taken at least a course in educational assessment. Therefore, the knowledge gained from the course is expected to influence the way he/she goes about classroom assessment.

Researchers, however, have hinted that apart from the influence of the knowledge and skills of teachers on classroom assessment practices, their demographic characteristics also, have a relationship with the way teachers carry out classroom assessment (Adeyemi, 2010; Kolo- Keaikitse, 2012; Alkharusi, 2011). Adeyemi, for instance has argued that teachers' qualification, years of teaching experiences and gender have a relationship with the way the teachers carry out classroom assessment. In the view of Adeyemi (2010), that is to say that teacher characteristics influence teaching and learning of pupils in classrooms. Owusu-Mensah (2019) indicated that teachers' age, years of teaching experience and academic qualification accounted for the variances in the classroom assessment practices of teachers.

For instance, Studies have shown that teaching experience is positively and significantly linked to teacher effectiveness. Teachers continue to improve their ability to increase student performance through the second and often third decade of teaching, but then later in their career to level off or decrease their effectiveness (Kini, & Podolsky, 2016). Some authors confirm that brand new teachers are on average less successful than those with some experience (Harris & Sass, 2011; Ladd and Sorensen, 2017). Also, the issue of professionalism in teaching has been on course for quite some decades ago. Scholars argued for the necessity of skilled teachers for effective learning. For example, Ngada as cited in Fajonyomi (2007) stated that the success or failure of any educational programme rests majorly on the adequate availability of qualified (professional), competent and dedicated teachers. Seweje and Jegede (2005) noted that the ability of a teacher to teach is not derived only from pedagogical skill acquired but also one's academic background.

In Ajayi (2009) point of view, the professional qualities of a teacher have to do with the following: Ability to involve the students in meaningful activities throughout the period of teaching; Management of the details of learning; Frequent monitoring of students' progress through tests, formal and informal, written and oral quizzes. In a similar way, there had been divergent views and reports as to the comparative ability of male and female in human endeavours, especially in education. An example of such such views came from Fauth (2016), who noted that women have been found to be more concerned than men about the academic achievement of students and participate more in professional growth activities. The views of the aforementioned authors on the relationship between teachers' demographics

and their effectiveness, suggests that the demographic characteristics may influence the job of the teacher, of which assessment is key component.

Statement of the Problem

The problems with teachers' practice of classroom assessment continuous to be an issue of concern to many stakeholders in education worldwide. The findings of previous studies on the assessment practices of teachers suggest that they leave much to be desired (Ololube, 2008; Onyechere, 2000; Frey & Schmitt, 2010). That is to say, such teachers either construct poor items or often rely on past questions to assess their students. Therefore, the results from these defective classroom assessments make the interpretations and uses to which these scores are put unreliable. In other words, scores from such tests may fail to meet the intended purpose or the uses to which these results will be put to may not be appropriate.

Similarly, in Ghana, accusing fingers have been pointed at teachers' skills in constructing, administering and scoring of good classroom tests (Quagrain, 1992; Anhwere, 2009; Oduro-Okyireh, 2008; Amoako, Asamoah & Bortey, 2019). A careful study of the previous studies suggests that teachers' assessment practices are mainly underpinned by their knowledge of educational assessment that they received as part of their initial training. It is worthy to note that while teachers' knowledge in educational assessment may influence their classroom assessment practices, literature has shown that other factors such teachers demographic characteristics may also influence the way teachers practise assessment.

Besides, I have observed personally, that while testing practices of Ghanaian teachers have been explored by several researchers, such studies,

mostly, focused on teachers in the Southern part of Ghana (Quagrain,1992; Anhwere, 2009; Oduro-Okyireh, 2008). It appears no empirical study has been reported on the testing practices of teachers in the Sissala East Municipality and for that matter the Northern part of Ghana.

Also, it is important to point out that even though previous researchers have studied teachers' assessment practices, they mostly focused on just finding out if any differences or associations exist in the assessment practices of teachers on the basis of their demographic characteristics. It is important to indicate that merely finding out if differences or association exist in teachers' assessment practice may not necessarily point out if those demographic characteristics contribute to the way teachers practice assessment.

The current study, therefore, found it relevant to investigate the testing practices of teachers and specifically, the contribution of teachers' demographic characteristics; gender, years of teaching, teacher qualification and subject teachers teach towards teachers' adherence to testing principles among selected Junior High Schools in the Sissala East Municipality.

Purpose of the Study

The main purpose of the study was to investigate the contribution of teachers' demographic characteristics towards their adherence to testing principles. Specifically, the study examined whether:

1. Gender predicts how teachers adhere to the principles of testing; construction, administration and scoring.
2. Years of teaching predicts how teachers adhere to the principles of testing; construction, administration and scoring.

3. Teacher qualification predicts how teachers adhere to the principles of testing; construction, administration and scoring.
4. Subject teachers teach how teachers adhere to the principles of testing; construction, administration and scoring.
5. Gender, years of teaching, teacher qualification and subject teachers teach combined together, will predict how teachers adhere to the principles of testing; construction, administration and scoring.

Research Hypotheses

The following hypotheses were tested for the study;

1. H_0 : Gender does not significantly predict teachers' adherence to principles of testing.
 H_1 : Gender significantly predicts teachers' adherence to principles of testing.
2. H_0 : Teachers' years of teaching does not significantly predict teachers' adherence to principles of testing.
 H_1 : Teachers' years of teaching significantly predicts adherence to principles of testing.
3. H_0 : Qualification of teachers does not significantly predict teachers' adherence to principles of testing.
 H_1 : Qualification of teachers significantly predicts teachers' adherence to principles of testing.
4. H_0 : Subject teachers teach does not significantly predict teachers' adherence to principles of testing.
 H_1 : Subject teachers teach significantly predicts teachers' adherence to principles of testing.

5. H_0 : Gender, years of teaching, teacher qualification and subject teachers teach combined together do not predict teachers' adherence to principles of testing.

H_1 : Gender, years of teaching, teacher qualification and subject teachers teach combined together predicts teachers' adherence to principles of testing.

Significance of the Study

Considering the vital role that assessment plays in every educational setting, especially in Ghanaian schools in terms of the uses to which the results of such tests are put, there was the need to investigate the contribution of teachers' demographic characteristics towards their testing practices. It is my belief that the findings from this current study would contribute to the following:

Evidence from this study would reveal to teachers, particularly, JHS teachers in the Sissala East Municipality about some of the flaws associated with their testing practices. When teachers get to know some of the flaws in their testing practices, it may help them improve on their assessment of students' learning, thereby helping teachers to make more valid and reliable decisions on students learning as well as curricula and instruction delivery approaches since some suggestions will be provided by this study on how to improve some of the bad spots that will be discovered on teachers' testing practices.

Also, findings from the study could encourage policy makers such as the Municipal Directorate of the Ghana Education Service to intensify interventions such as seminars, and workshops for teachers on how to improve

their classroom testing competences, by particularly, taking into consideration other factors that may influence the assessment practices of teachers, other than their demographic characteristics.

Finally, the current study utilised a significant amount of information from literature in terms of previous empirical works on teachers' classroom assessment practices globally. Therefore, this current study would contribute to expanding knowledge in existing literature on the teachers' classroom testing practices. Particularly, by adding to literature that though there may be differences and associations in assessment practices of teachers based on their demographic characteristics, these characteristics do not significantly influence their testing practices.

Delimitations

The study was confined to gender, years of teaching, teacher qualification and subject area as predictors of teachers' adherence to principles of testing among selected Junior High Schools in the Sissala East Municipality. Also, the study was delimited to the principles guiding the construction, administration and scoring of the teacher-made tests. Finally, the study focused on only teachers who teach English Language, Mathematics, Integrated Science and Social Studies in the Junior High Schools in the Sissala East Municipality. Geographically, the five (5) circuits which were randomly selected for the study, as well as the subject teachers were also used for the study.

Limitations

Inevitably, every research work comes with it, some inherent weaknesses bordering on different aspects of the study and usually span the

planning through to the end of the final report. This study is therefore, not an exception.

The design for the study was a descriptive cross-sectional survey. Therefore, the instrument deemed appropriate for collecting data was a questionnaire, and hence, its shortcomings cannot be ignored. Human participants were used for the research studies on the psychological assumption that they will report honestly about themselves. However, when questionnaires are given to respondents, some may not have enough time to read them carefully and answer them accurately. This may lead to giving responses that that will portray the positive aspect of the problem and ignoring the negative aspect. Particularly, when the issues bother on their integrity (Bratton & Mattes, 2005). This according to the duo may affect the results of the study.

Assumptions of the Study

Several assumptions undergird this study.

1. The researcher assumes that the respondents who took part in this study constitute a representative sample of the JHS teachers teaching the four core subjects in the Sissala East Municipality.
2. Teachers' demographic characteristics such as gender, years of teaching, qualification of teacher and subject area have an influence in their classroom assessment practices.
3. Practice makes perfect. Therefore, the more the number of years teachers teach and assess their students, the more they improve on their classroom assessment practices.

Definition of terms

Assessment: The use of different strategies and techniques in gathering evidence of students' learning.

Test: A tool teachers use to determine the amount of attribute, abilities, skills or performance of an individual or group.

Organisation of the Study

This research report is organised into five chapters. Chapter one, which is the introductory chapter, comprised the background to the study, the statement of the problem, purpose of the study, the research objectives, research hypotheses, significance of the study, delimitations and limitations. Chapter two of this report presents the literature related to the study. The literature entails both theoretical and empirical reviews. Chapter three describes the methodology adopted for the study. Thus, it examines the research design, the study area, the population, the sample, the sampling procedure, the research instruments, the validity and reliability, the pilot-testing procedure, the data collection procedure and the analysis of data. Chapter four presents the research results and discussion of the findings in relation to the reviewed literature. Chapter five focused on the summary, conclusions and recommendations based on the research findings.

CHAPTER TWO

LITERATURE REVIEW

Introduction

The motivation behind the study was to investigate whether the demographic characteristics of junior high school teachers in the Sissala East Municipality predict their testing practices. Literature relevant and related to this the topic were reviewed. This chapter consisted of three sections: The theoretical review, the conceptual review and the empirical review.

Theoretical Review

Classroom Assessment Environment Model (Brookhart, 1997)

Brookhart (1997) developed the model of the classroom assessment environment by reviewing and combining literature from classroom assessment literature, theories in social cognitive learning and motivation theories. The model encompasses students' conceptualisation of the classroom climate and their experience, in relation to the assessment purposes established by the teacher. The assignment of assessment tasks, setting of performance criteria and standards, communication of feedback, and monitoring of learning outcomes. Based on this model, the expectations of students about the classroom assessment environment are assumed to have an effect on the motivation of students' behaviours and the achievement-related outcomes of the students. This theory was tested on students at the primary, middle and high school, and for college (Brookhart & Bronowicz, 2003; Church, Elliot, & Gable, 2001).

Nonetheless, the majority of the assessment process has used individual student scores as the unit of measurement instead of the average student performance in the classroom (Church et al., 2001). The supporters of this model contended that the students in the same classroom perceive and interpret classroom assessment practices differently, as a result of the different entry behaviours they brought to the classroom (Ames, 1992). Meanwhile, Church et al. asserted that “composite measures of perceived classroom [assessment environment] have been shown to be internally consistent, and composite indicators yield a more comprehensive assessment of the perceived classroom environment than do individual indicators” (p. 51). Similarly, Brookhart (2004, p. 444) (as cited in Alkharusi, 2015) emphasised that “classes have an assessment ‘character’ or environment” that emerges from the teacher’s assessment practices, and that “students construct their own meaning of the classroom assessment environment based in part on their group experiences” (p. 445).

Stiggins and Conklin (1992), initially, introduced the concept of a classroom assessment environment based on observations they made about the assessment practices of some selected four teachers in three sixth grade classrooms. Stiggins and Conklin, argue that the classroom assessment environment involved eight key elements, which were outlined as follows: The purpose of the assessment, the assessment methods to be used, the criteria for choosing the assessment methods, the quality of assessment, assessment results and feedback, teacher’s assessment background and preparation, teacher’s perception of students, and assessment policy (Stiggins & Conklin, 1992). “Students’ perceptions of the meaning” of the classroom assessment

practices is at heart of classroom environment research (Ames, 1992b, p. 264; Stiggins and Conklin's, 1992). The conceptualisation of the classroom assessment environment focused more on teacher practices than on the perceptions of these practices by the students (Brookhart & Durkin, 2003). As such, based on a synthesis of classroom assessment and motivation literature, Brookhart (1997) developed a theoretical framework for the role of student motivation and achievement in classroom assessments. In this context, the classroom assessment setting was described as a situation encountered by students as the teacher sets assessment targets, assigns assessment tasks, sets performance requirements and expectations, provides input and tracks results (Brookhart, 1997).

In the context of the present study, the researcher was interested in finding out the contribution of teachers' demographic characteristics; the gender, years of teaching, academic qualification and subject teachers teach towards their testing practices. The classroom assessment environment was deemed appropriate to provide a broader lens through which teachers' testing practices can be viewed. It was envisaged that teachers with high academic qualification and those with more years of will establish a more suitable classroom assessment environment that will motivate students learning and achievement compared with their counterparts with low academic qualification and lesser number of years of teaching.

Classical True Score Theory

A test theory or test model is a theoretical reflection of the variables that affect the observed test scores and the assumptions underlying them. The concept classical true score theory, evolved from early works of Edward Lee

Thorndike in the 1904 in his first textbook on test theory. The Classical true score theory is a simple model that describes how measurement errors can influence observed scores of test takers. Proponents of the theory state that, the measurement of any construct, the observed score (X) is the sum of a true score, or true underlying ability (T), and the amount of error that emanates from the measurement of the construct, denoted as E. Thus, it is stated mathematically as $X = T + E$ (Crocker & Algina, 2008). It is believed that when students write a particular test, aimed at measuring a construct in two or more successions, it is unlikely that the scores they will obtain in the test will be the same. This is often due to the influence of some factors such as, inappropriate item writing, inappropriate test administration practices, fatigue, guessing, careless marking, or improper scoring that occurs during the testing process.

Similarly, a different form of the test could also yield differences in scores because of variation in content. These inconsistencies in individual scores that may result from the sampling of tasks and other factors must be regarded as measurement error (Crocker & Algina, 2008). According to Crocker and Algina, the “True Score” can be understood as the average of the observed scores obtained by a test taker, if the same test is taken over an endless number of times. In the classroom setting, the “true score” is the mark a teacher would assign if he were to take the average score of an individual from an unlimited number of test administrations. In reality, however, researchers argue that every test is meant to serve a purpose, and hence, an individual cannot administer a test an infinite number of times. They assert instead that, what is common in most testing situations is that the individual

has the opportunity of taking the test once and then the scores as well as the decisions based on the scores also determined. Therefore, reliability coefficients are used to approximate both true and error variance associated with our observed test scores (Crocker & Algina, 2008). The classical true score theory provides the basis for the development of most aptitude, achievement and other personality inventories used in contemporary times (Crocker & Algina, 2008).

In the application of the classical true-score theory, a lot of assumptions are employed on the relationship among these three components (Observed Score, True Score and Error Score) (Allen & Yen, 1979; Crocker & Algina, 1986). Most of the standard procedures for creating and evaluating classroom teacher made test are based on a set of operational assumptions on the Classical true-score theory.

Assumptions of the Classical True Score Theory

The model holds some conditions to be true. Thus, if these assumptions are reasonable, then the conclusions that are made from the model are reasonable. However, if the conditions are not reasonable, then the use of the model leads to questions about conclusions derived from the measurement. Among the several assumptions include: First, the raw score (X) comprises the true score (T) and the measurement error (E). Assuming a teacher administers a test to a student over an unlimited number of times, the average of the scores will be the student's best or true score (T) (Allen & Yen, 2012). In estimating the true score of a student, it is expected that the random errors around the student's score be normally distributed. And since the random errors are normally distributed, the expected value of the error is 0.

Also, the random errors are believed to be uncorrelated with each other (Novick, 1996).

Similarly, the random errors are uncorrelated with the true score, T . There is no systematic relationship between a true score (T) and whether or not that test taker will have positive or negative errors (Birnbaum, Lord & Novick, 1968). All of these assumptions of random errors led to the evolution of the Classical Test Theory. The standard deviation of the distribution of random scores around the true score is termed the standard error of measurement. The lower it is, the more closely packed the random errors are, around the true score.

The link between Classical Test Theory and this study

As regards this present study, it can be seen that when teachers construct, administer and score tests with weaknesses such as: unclear meanings, poor sentence structure, insufficient time limits, inappropriate arrangement of items, poor sitting arrangement, warning students about dire consequences of failure, making persistent comments unrelated to the tests, scoring test with the use of the scoring guide, scoring test when one is not sound physically or mentally (Amedahe & Asamoah- Gyimah, 2003); they introduce errors into the scores of students. They argue that the presence of such poor practices in classroom testing tends to affect the reliability of the scores observed from students using tests. They emphasised that the use of such faulty and unreliable scores from students tend to affect the reliability of any decisions that is made from such scores. The application of this theory in the measurement of psychological constructs, has the validity and reliability of such tests. The implication of this theory, therefore, means that in order to

achieve the reliability and validity of teacher made test, the principles of the theory need to apply. In the context of this study, it was envisaged that teachers with high academic qualification and those who have taught more years should be plan quality classroom assessment compared with their counterparts who have low academic qualification and lesser number of years in teaching.

Conceptual Review

Brief History of the Development of Testing

People who take a cue from history, often stand a chance of avoiding the mistakes of the past. Therefore, in reviewing literature on the state of the art of classroom assessment of human abilities, it is imperative to look back in time on some of the forces that have triggered the development of these measures of intellectual abilities, with the view to enhancing understanding of why such measures matter even in the contemporary classroom.

Talking about the origin of tests, it is believed that China set the pace for the development of tests. The practice of testing can be traced to China long before 1115 B.C (DuBois, 1970; Ebel, & Frisbie, 1972). According to DuBois (1970), China initiated the first organised form of testing as far back as 2200 B.C. These tests covered content on civil law, military affairs, revenue, agriculture, and geography. Civil servants were tested in every three years. DuBois pointed out that this testing programme lasted for about 4000 years. However, it was discontinued when Alfred Binet developed the scale for measuring intelligence. India was first to adopt the Chinese method in recruiting civil servants and later the USA also adopted it.

In the US, tests were developed to place employees in many occupations. As a means of validating tests, statistical methods were used to establish the correlation between performance of employees on the test and later achievement on the job. However, formal testing did not begin until the 12th century by the Arabs based on the Chinese method. The use of written test as a means of assessment was done by Jesuits at St. Ignatio (DuBois, 1970). He noted that the development and use of academic test were pioneered in the Britain, specifically, in the University of London. Testing then was used as the legitimate basis of awarding degrees and the making other decisions.

Based on the history of testing aforementioned, it can be seen that the use of testing in institutions has received global attention for several decades. Therefore, it is important that in carrying testing, especially in educational institutions, conscious effort has to be made by to ensure that the principles that guide testing are adhered to strictly.

Concept of Classroom Achievement Tests

Classroom achievement also, called teacher-made tests are designed and administered as part of the teaching and learning process in an educational setting (McDaniel, 1994). Such assessments are used by teachers to measure how much students learn or how much they gain at the end of a unit of the syllabus, term or at the end of an academic year (Amedahe, 1989). According to Mehrens and Lehmann (1991), tests planned by teachers usually measure students' achievement in a given course of study, at a particular level of learning. The predominance of teacher-made tests in every educational set up is given credence by the conclusions of studies by Herman and Dorr-Bremme, and Stiggins and Bridgeford (as cited in Mehrens & Lehmann, 1991), that in

the face of the ever-increasing number of the use of portfolios and performance assessments to gauge student progress, all these evaluations are based on teacher-made tests.

Measurement experts have clearly defined the main purpose of the teacher-made tests (Ebel & Frisbie, 1991; Etsey, 2004; Gronlund, 1988; Mehrens & Lehmann, 1991). Such authorities accepted that the main purpose of a teacher-made test is to obtain true, accurate, and useful information about how much students have mastered a concept over a period of time. Information gained through classroom achievement tests helps teachers to examine the overall teaching and learning in the classroom. There are several classifications of achievement tests (Mehrens & Lehmann 1991). In the view of Mehrens and Lehmann, the most common classification of teacher-made tests, to test experts is the classification based on the type of item format used, which classifies tests into objective-type tests and tests of the essay forms (Cunningham, 1986; Etsey, 2004). The aforementioned test experts have argued that essay-type tests can either be the extended or restricted response types, while objective-type tests can take the form of short, true-false, matching or multiple-choice tests. Testing in educational institutions is designed to assess either curriculum based (classroom instructional) achievement or a variety of students' traits other than curriculum-based achievement. Tests such as career interest, attitudes, and personality tests assess a variety of students' traits other than curriculum-based achievement (Nitko, 2001). Stainback and Stainback (1996) argued that, depending on how it is interpreted an example of a curriculum-based assessment (CBA) could be the measurement of almost any student performance deriving from or linked to

the classroom curriculum, including performance testing. It should be stressed that, the main reason for testing students' achievement, is to gauge the students' level of mastery, based on the scope of the content areas they have learned, which are derived from the school curriculum.

Etsey (2012) stated that achievement test “measures the extent of present knowledge and skills. In achievement testing, test takers are given the opportunity to demonstrate their acquired knowledge and skills in specific learning situations” (p. 41). A comprehensive literature review posits two key types of tests on achievement. These are teacher-made tests and external tests (Nitko, 2001). The assessment procedures used by teachers to estimate the amount of knowledge and understanding students have attained is called teacher-made tests. Teachers construct these tests to assess gain an understanding of how much a student has learned (Amedahe, 1989).

External tests or “extra-classroom assessments” on the other hand, include assessment tools developed and/or graded by persons who are not associated with the training schools of the students (Lissitz & Schafer, 2002). For example, Commercial test publishers, departments of education, and local school jurisdictions, usually develop external test. External test can take the form of textbook accompaniments, survey tests and mandated tests (Nitko, 2001)

Construction of Classroom Achievement Tests

Over the years, a variety of educational measurement experts have developed the basic principles for the construction of teacher-made tests (Amedahe, 1989). While some of the test testing principles are general and apply to any type of test, others are specific and apply solely to the particular

type of test under construction. From the available literature, the principles of testing, which researchers found most extensive and feasible in the classroom test situation, were those posited by Adom, Mensah, and Dake (2020), Tamakloe, Atta and Amedahe (1996) and Etsey (2004). These authors have outlined the following eight guidelines for planning a good test.

- a) The purpose of the test must be stated.
- b) Must determine an appropriate item format to be used.
- c) Must determine specifically, what is to be tested.
- d) The individual items are drafted.
- e) Each item is reviewed.
- f) Must prepare a scoring guide or marking scheme
- g) Must provide clear directions, and
- h) Must evaluate the whole test.

According to Gronlund (1988), the key to effective achievement testing is careful planning (p. 15). That is to say that the purpose of the test must be decided during the planning stage. As already pointed out in the literature, tests can be used for a number of purposes. Nonetheless, it is worth noting that each type of test use typically requires some modification of the test design and thus decides what type of item format to use. The second step of the planning stage is the determination of the item format to use. As stated earlier in the literature, the most common item formats, teachers used are the essay- and the objective-types. According to Etsey (2004), it is better to use more than a single item format in planning a test. This is because one item format cannot be used exclusively for measuring all learning outcomes, depending on the purpose of the test. According to Mehrens and Lehmann

(1991), The choice of a suitable item format depends on factors such as what the test wants to achieve, how much time is available to develop and do the scoring of the test, the number of students the test is meant for, the skills and competencies to be tested, the level of difficulty required, the facilities available for the test, the age of the students and the ability of the teacher to compose the various types of item. The final step of the planning stage is the determination of what is to be tested or measured. According to Etsey (2004), at this level, the teacher will decide the chapters or units of the course content to be covered by the test and the abilities, skills or attitudes to be assessed. The educational objectives must be defined in terms of student behaviours and related to what has been emphasised in the class. A test plan should be guided with table of specifications. The table of specifications aligns the content with the learning objectives (Etsey, 2004). The use of the specification table helps to ensure that the number of items on the test is fairly distributed, and also, helps avoid overlapping when constructing the test items.

After the planning stage, the actual writing of the individual test items follows. Tamakloe et al. (1996) and Etsey (2004) have pointed out that whichever test item types that are being constructed, must follow the basic principles laid down for them. There are, however, general guidelines that according to Mehrens and Lehmann (1991) and Etsey (2004), apply to all types of tests. These include:

1. The table of specifications must be kept before the teacher and constantly referred to as he/she writes the items.
2. The test items must be aligned or related to the specific learning objectives.

3. Must construct items which are precise and unambiguous.
4. Avoid the use of excessive verbiage and use of complex sentences.
5. The test items must be targeted at information that the students should know.
6. The initial draft of the items should contain more items than would be needed.
7. Items of varying levels of difficulty must be used. This, however, depends on the purpose of the test.
8. The test items and the scoring guide must be written just after the concept is taught.
9. In order to permit review and editing, the test items must be written in advance (at least two weeks) of the date the test will be taken

It is worthy of note that when it comes to the ideal number of items for a test, researchers offer different opinions. For example, while Mehrens and Lehmann (1991) argue that the initial number of items should be 25% more, Hanna (as cited in Amedahe, 1989) has suggested 10% more items than are actually needed should be added.

After the items have been written, Tamakloe et al. (1996) call the next stage the item preparation stage. At this stage the test items must be reviewed and edited. Etsey (2004) indicated that the items have to be reviewed thoroughly for at least one week after writing them. He has emphasised that the test items should be given other teachers to double check for possible errors and clarity. Reviewing and editing the items is to ensure that items that are poorly constructed are modified, to check the language difficulty level of items, and to check the number of items on the test.

Assembling of the test is done, after reviewing and editing the test items. In assembling test items, it is advised that the following points must be considered (Etsey, 2004; Mehrens & Lehmann, 1991; Tamakloe et al., 1996).

1. The items should be arranged in sections by item formats. The sections must progress from easier formats (true-false) to more difficult formats (interpretive exercises and essay).
2. Within each section or format, the items must be arranged in order of increasing difficulty. One way of achieving this is to group items in each format according to the instructional objectives being measured and make sure that they progress from simple to complex. According to Mehrens and Lehmann (1991), such a grouping has the advantage of helping the teacher to ascertain which learning activities appear to be most readily understood by students, those that are least understood and those that are in-between. Hambleton and Traub (as cited in Mehrens & Lehmann, 1991), argue that, ordering items in ascending order of difficulty leads to better performance than either a random or hard-to-easy ordering. Lafitte (as cited in Mehrens & Lehmann, 1991) on the other hand, has reported inconclusive data. Although, empirical evidence is also inconclusive about the effectiveness of using statistical item difficulty as a means of ordering items, Sax and Cromack (as cited in Mehrens & Lehmann, 1991), Mehrens and Lehmann (1991) recommended that items should start from the easy to the difficult for long or timed tests in order to instil confidence in the examinee, particularly at the start. Determining the difficulty indexes by the classroom teacher seems impracticable to a large extent (Tamakloe et

al., 1996). This is because statistical item difficulty data are always gathered after test administration or test try-outs and teacher-made test items are usually not pre-tested. Mehrens and Lehmann (1991) however, recommended that subjective judgement must be relied on to determine difficulty level of items. They have stated that -teachers could only categorise their items as difficult, average or easy.

3. The items must be spaced and numbered consecutively so that they are not crowded and can easily be read.
4. All stems and options must be together on the same page and if possible, diagrams and questions must be kept together.
5. If a diagram is used for a multiple-choice test, the diagram must be placed above the stem.
6. A definite response pattern to the correct answer must be avoided.

In addition to the above, Etsey (2004) has recommended that for objective-type tests, the options must be written vertically below the stem rather than across the page. In addition, Etsey (2004) indicated that test items can also be presented in the order in which they were taught in the classroom or in the order in which the material appeared in the textbook. After the test items have been assembled, the next task is the preparation of the scoring key, the marking scheme or the scoring rubric (Etsey, 2004). According to Etsey (2004), and Amedahe and Gyimah (2003), the marking scheme must be prepared when the items are still fresh in the mind of the teacher and always before the test is administered. This way, defective items that do not match their expected responses would be recognised and reviewed. For objective-type tests, correct responses to items should be listed. For essay-type tests,

points or marks should be assigned to various expected qualities of responses. Mehrens and Lehmann (1991) noted that if the teacher finds it appropriate to have differential weighting for specific essay questions, considerations such as the time required to respond, the difficulty of the issue and the content area which emphasised in the class during the during the delivery process should be considered. Immediately following the preparation of the marking scheme is the writing of clear and concise directions for the entire test and sections of the test. Here, the time limit for the test must be clearly stated. A good working rule is to try to set a time limit so that about 90 percent of students feel they have adequate time to complete the test (Nunnally, 1964; Ebel & Frisbie (1991). Directions, according to Etsey (2004), must include penalties for undesirable writings, a number of items to respond to, where and how the answer should be written, credits for the orderly presentation of material (where necessary), and modes of identification of examinees.

The last stage of the test construction process is the evaluation of the test on the criteria of clarity, validity, practicality, efficiency and fairness. Amedahe and Gyimah (2003), Etsey (2004) and Tamakloe et al. (1996) posit that Clarity of a test relates to how the elements are written simply and clearly, relative to the test tester's ability level and the material that is being measured. It also refers to the kinds of information that the test is measuring and how the test items contribute adequately to the material and course objectives. Validity bothers on how closely the test represents the material presented in the course unit or chapter and how faithfully the test reflects the difficulty level of the material taught in class. The issue of validity here establishes the content

validity evidence of the test (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996).

On practicality, consideration is given to whether students will have enough time to complete the test. It also bothers on whether there are enough materials (chairs, tables, answer booklets) to present the test and complete it effectively (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996). Efficiency bothers on finding out whether the test is the best way to measure the desired knowledge, skill or attitude. Consideration must also be given to the problems that might arise due to material difficulty or shortage and these expected problems well catered for (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996).

On the fairness criterion, consideration is given to whether students have been given advance notice of the test, whether students have been adequately prepared for the test, and whether students understand the testing procedures. Consideration is also given to how the lives of students are affected based on the possible ways by which the test scores will be used (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996). After this comprehensive evaluation of the test, the test can be submitted to be processed for subsequent administration.

Validity of Test Items

Validity is “the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses” of a test (AERA, APA, & NCME, 2014, p. 11). The concept of validity according to Nitko (2001) is the “soundness of one’s interpretation and uses of students’ assessment results”. This means that for teachers in the Junior High Schools to

produce valid results of their students, the students' results must be supported with many evidences. The results must be devoid of errors and therefore, the soundness of the results. The emphasis here is not necessarily on scores or test questions, but rather on the meanings that is derived from the scores. This means that the behavioural interpretations made from the student's performance is of paramount concern.

Validity is an integrated evaluative judgment to support the adequacy and appropriateness of inferences and actions based on test scores and other assessment methods (Messick, 2003). The validity of classroom assessment depends on appraising the intended learning and all its embedded components, providing a good match between the assessment strategies, the learning targets and the decisions taken by teachers and learners on learning, ensuring that the assessment accurately covers the desired learning outcomes like content, thought processes, skills and attitudes (Northern Canadian Protocol for Collaboration in Education, 2006) (As cited in Lorna & Manitoba. School Programs Division, 2006).

A measure itself is neither valid nor invalid; rather, the validity question concerns the meanings and uses of test scores. The explanations and uses of the results of one's assessment are also true only when the values they indicate are acceptable. In general, the interpretations and uses of the assessment results are also true when the implications of these interpretations and uses are compatible with acceptable values. Here, when the values of the assessment are not in accordance with the consequence of the assessment then this principle is violated. Another significant assumption of validity concept is that validity is a matter of degree, it is not an "all-or-none" question. That is,

rather than merely true or false, the validity of a test when it comes to the inferences made from the test scores should be conceived in terms of being strong or weak. Validity should be a key factor for research consumers when selecting psychological assessments. Although these choices are based on a variety of practical, theoretical, and psychometric considerations, a test should only be chosen if the expected interpretation and usage are supported by good enough evidence. A third critical aspect of validity is that an understanding of a test is founded on proof and theory. To be confident in the interpretation and use of test scores, a test user must have empirical evidence to support interpretation and use. Further, Contemporary validity views underline the significance of basing the analysis and use of the test on a defensible psychological theory). There are three different types of validity evidence. Namely, criterion validity, construct validity and content validity

Validity Evidence

The AERA/APA/NCME, (2012), outlined three categories of validity evidence; Content validity, Criterion-related validity and Construct validity.

The Concept of Content Validity

Content Validity is often characterised as the degree to which the sample of items, tasks or test questions is representative of the content domain (Moss, 1992). Validity according to Nitko (2001) is the “soundness of your interpretations and uses of students’ assessment results”. Nitko, clarified that in talking about validity of students’ assessment, emphasis must be placed on how one interprets the results, not the instrument or procedure itself. Bollen (as cited in Drost, 2011), defined content validity as a qualitative form of validity in which the concept domain is clearly defined and the analyst

assesses whether the measures fully reflect the domain (p.185). But, William (1993) argues that "content validity should be concerned not just with test questions, but also with the answers elicited, and the relationship between them" (p. 4). That is to say, for content-related evidence, William advocates the inclusion of elicited responses that actually correlate to the targets of the assessment task.

Content validity is usually seen as a general property of a test. The test developer, who determines the domain of content and writes test items to represent the domain, achieves some success to some extent in achieving its objective.

According to Miller, McIntire and Lovler (2011) ((as cited in Mall (2014), there are evidence of validity to be demonstrated based on test content during test development. These evidences include:

1. Defining the test universe which involves the amount of knowledge or behaviour that a test reflects. They further asserted that, the step involves analysing similar instruments that measure the same concept, consulting other people who have deeper understanding of the construct. This is done to ensure you understand the concept clearly and can define it clearly, the way you will measure it. According to Groth-Marnat (1997), evidence of validity bases on test content requires that the test covers all the major aspect of the testing universe in the correct proportion.
2. Developing the test specifications/blue print which involves a documented plan containing details about the test's content. The specification delineates, the thinking processes the test is to measure

with their given proportion, the content area with respect to the subject matter the test is to be measured and the number of questions that will be included to assess each content,

3. Establishing an appropriate test format in which the test will be constructed to elicit the construct of interest,
4. Constructing the test questions. Here test developers are to be careful that each question represents the content area and the objective it is intended to measure (pp. 196-197)

Criterion-Related Validity

Criterion-related validity is the degree of correspondence between a test measure and one or more external referents (criteria), usually measured by their correlation. Criterion-related evidence answers the question, how well the results of an assessment can be used to infer or predict an individual's standing on one or more outcomes other than the assessment procedure itself. Here, the outcome is called the criterion (Etsey, 2012). There are two types of criterion-related evidence. These are concurrent validity and predictive validity. We talk about concurrent validity when the criterion exists at the same time as the measure. Concurrent validity refers to the ability of a test to predict an event in the present. In concurrent validity, one is asking whether the test score can be substituted for some less efficient way of gathering criterion data (such as using a score from a group scholastic aptitude test instead of a more expensive-to-gather individual aptitude test score).

Again, for concurrent validity, data are collected at approximately the same time and the purpose is to substitute the assessment result for the scores of a related variable. For instant, a test of swimming ability verses swimming

itself. When the criterion occurs in the future, we talk about predictive validity. Predictive validity evidence refers to the extent to which individual's future performance on a criterion can be predicted from their prior performance on an assessment instrument. For predictive validity, data are collected at different times. Scores on the predictor variables are collected prior to the scores on the criterion variables (Etsey, 2012). The purpose is to predict the future performance of a criterion variable. For instant, using first year GPA to predict the final CGPA of a University student. Another example is to use students GMAT scores to predict their GPA in a graduate programme. We would use correlations to assess the strength of the association between the GMAT score with the criterion (i.e., GPA). Although concurrent and predictive validity differs in the time period when the criterion data are gathered, they are both concerned with prediction in a generalizability sense of the term. In this study, both concurrent and predictive reliability would aid one to tell whether an individual behaviour should be reinforced concurrently or based on one's behaviour, one will be able to perform a particular task in the future.

Construct Validity

DeVellis (1991) explains that the construct validity of a measure "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).

Messick's (1989) definition of construct validity captures the breadth of the concept of validity; "validity is an integrated evaluative judgement of

the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (p, 13.). In the view of Moss (1992), he points 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 summaries" (p, 233). This means asking whether the interpretation given to the test score truly summaries the behaviour. That is, a construct needs to be both operationalised and syntactically defined in order to measure it effectively (Benson, 1998; Crocker & Algina, 1986). The operationalizing of the construct involves developing a series of measurable behaviours or attributes that are posited to correspond to the latent construct. Defining the construct syntactically involves establishing assumed relationships between the construct of interest and other related constructs or behaviours (Benson, 1998; Crocker & Algina, 1986).

If a relationship is causal, what are the specific cause and effect behaviours or constructs involved in the relationship? Construct validity refers to how well you interpreted and converted the operationalisation of the concept, idea, or behaviour which is built into a working and functional reality (Trochim, 2006). Here, the study will lay emphasis on how teachers measure the constructs of students when assessing them.

Principles of Constructing Classroom or Teacher-Made Tests

Test plays a cardinal role in the assessment processes in educational settings. Good and quality test items are not just constructed by test constructors or experts. They require adequate and extensive planning so that the instructional objectives, the teaching strategy to be employed, the textual

material, and the evaluative procedures are all related. Ideally, every test should be reviewed critically by other teachers to minimize the deficiencies identified in it by an expert. Without adequate and careful planning, one can be fairly certain that one's test will not be very good (Tinkelman, 1971). According to Tinkelman (p. 46) "at the very least, inattention to planning can lead to waste and to delay due to failure to coordinate properly the various phrases of test construction."

Based on the above, Mehrens and Lehmann (1991) (as cited in Anhwere, 2009), outlined the following stages and steps as being important to test the construction of the classroom or teacher-made test;

1. Specify the course or unit content
2. List the major course or unit objectives
3. Define each objective in terms of students' behaviour
4. Discard unrealistic objectives
5. Prepare a table of specifications
6. Decide on the type of item format to be used
7. Prepare test items that match the instructional objectives

In addition to the basic principles of test construction, Mehrens and Lehmann (1991) indicated that, in writing, of any classroom or teacher-made tests, it is prudent the teacher considers the following;

1. Identify the purpose of the test i.e. What the teacher wants to achieve by the test.
2. Select the test item type that will best measure the learning outcome.
3. Obtain a representative sample of student behaviour which the teacher would want to evaluate (e.g. in the affective domain; does the teacher

wants to know how a student has received a classroom activity, responding, valuing, organization, characterisation by a value complex?)

4. Construct test items of the proper level of difficulty.
5. Try to eliminate factors that are extraneous.

Classroom achievement tests are generally teacher-made tests (McDaniel, 1994). These tests are constructed by teachers to test the amount of learning done by students or their attainment at the end of a course unit, term or at the end of an academic year (Amedahe, 1989). According to Mehrens and Lehmann (1991), teacher-made tests usually measure attainment in a single subject in a specific class or form or grade. The predominance of teacher-made tests in every educational set up is given credence by the conclusions of studies by Herman, Dorr-Bremme, Stiggins and Bridgeford (as cited in Mehrens & Lehmann, 1991) that, in the face of the ever-increasing use of portfolios and performance tests to assess student progress, teacher-made tests are mostly the major basis for evaluating student progress in school.

The main purpose of teacher-made tests has been delineated by measurement experts (Etsey, 2004; Gronlund, 1998; Mehrens & Lehmann, 1991). All these authorities have agreed with the fact that the main purpose of a teacher-made test is to obtain valid, reliable, and useful information concerning students' achievement and thus contributes to the evaluation of educational progress and attainments for the total improvement of classroom teaching and learning. Teacher-made tests can be classified in a variety of ways. According to Mehrens and Lehmann (1991), one type of classification is based on the type of item format used - essay-type versus objective-type.

Another classification is based on the stimulus material used to present the tests to students-verbal versus non-verbal, while other classifications may be based on the purposes of the tests and the use of the test results-criterion-referenced versus norm-referenced, achievement versus performance, and formative versus summative.

Administration of Classroom Achievement Tests

The guiding principle in test administration is to provide all examinees with a fair chance to demonstrate their achievement on what is being measured (Gronlund, 2006; Tamakloe et al., 1996). The need to maintain uniform conditions in test administration cannot be over-emphasised. This is especially essential for the test to yield consistent, reliable and valid scores without much influence of chance errors. This calls for ensuring a congenial psycho-physical atmosphere for test taking (Tamakloe et al., 1996, p. 214). This was also emphasised by Airasian (cited in Amedahe & Gyimah, 2003) that test administration is concerned with the physical and psychological setting in which students take their tests.

The first and foremost task of the teacher is to prepare his students in advance for the test (Etsey, 2004). Etsey has emphasised that for students' maximum performance, they should be made aware of when (date and time) the test will be given, the conditions (number of items, place of test, open or closed book) under which the test will be given, the content areas (study questions or list of learning targets) that the test will cover, the emphasis or weighting of content areas, the kinds of items (objective-types or essay-types) on the test, how the test will be scored and graded, and the importance of the results of the test.

The physical conditions that need to be in place to ensure maximum performance on the part of students include adequate work space, quietness in the vicinity, good lighting and ventilation and comfortable temperature (Etsey, 2004; Gronlund, 2012;) (as cited in Tamakloe et al., 1996). Adequate work space is very essential for test administration because when tables and chairs are closely arranged together, students will not have the independence to work on their own. This will in no doubt lead to students copying from each other. In addition, tables provided for the examination must be conducive to the testing materials being used. For example, in Practical Geography examinations where topographical sheets are used, each student could use two tables or desks in order to get an adequate work space (Tamakloe et al., 1996).

Noise and distraction in the testing environment should be kept at the barest minimum if not eliminated completely. Interruptions within and outside the testing room has the tendency of affecting students' performance (Mehrens & Lehmann, 1991; Tamakloe et al., 1996). Etsey (2004) has pointed out that it is helpful to hang a -Do Not Disturb. Testing in Progress sign at the door of the testing room to warn people to keep off. Good lighting is important in the effective test administration. This facilitates students' reading of instructions and test items without straining their eyes, thereby working faster (Gronlund, 2006) — Good ventilation and comfortable temperature should be assured since absence of these factors could create unrest or uneasiness in test takers, thereby making concentration difficult (Tamakloe et al., 1996, p. 215). Other basic physical conditions are that, all testing equipment must be in the room and readily available, and also, all possible emergencies during test administration must be expected and well catered for.

The psychological conditions in test administration, on the other hand, include the position of the invigilator, timing of the test, threatening behaviours of invigilators, and interruption to give instructions and announcements (Etsey, 2004; Bernstein, cited in Amedahe, 1989; Gronlund, 2012; Tamakloe et al., 1996). A study on the examiner as an inhibiting factor, carried out by Bernstein (1953) and reported by Amedahe (1989) found out that, the presence of the examiner tended to inhibit the performance of those students who were nervous. The crux of the matter is that if the mere presence of the examiner or invigilator could affect the performance of students who are nervous, then there is no doubt that the position of the invigilator is very significant to the performance of students on examinations. Etsey (2004) has recommended that the invigilator should stand where all students could be viewed and move among the students once a while to check malpractices. Such movements should not disturb the students. He must be vigilant. Reading novels or newspapers, making of and listening to telephone calls, dozing off and chatting are not allowed.

The timing of tests is very important. Tests must not be given immediately before or just after a long vacation, holidays or other important events where students are involved either physically or psychologically. Tests must also not be given when students would normally be doing something pleasant such as having lunch, athletics or other sporting activities as this will hamper students' concentration (Amedahe & Gyimah, 2003; Etsey, 2004).

Interruptions during testing, such as giving instruction, must be kept to the barest minimum and should always relate to the test. The time spent and time left to complete the test must be announced at regular intervals to enable

students apportion their time to the test items. Where practicable, the time should be written on the chalkboard at 15-minute intervals until near the end of the test when it could be changed every five minutes. Further, students should start the test promptly and stop on time (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996).

Teachers should always work on minimizing test anxiety in students during testing. They should, therefore, avoid, warning students to do their best because the test is important, telling students that they must work faster in order to finish on time, threatening dire consequences of failure in the test, and threatening students with tests if they do not behave (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996).

Guidelines in Administering Achievement Tests

According to Etsey (2004), in administering test items, classroom teachers are to consider that, the following information are essential in maximising students' performance.

1. Students must be made aware of the rules and regulations covering the conduct of the test.
2. Penalties for malpractice such as cheating should be clearly spelt out and clearly adhered to.
3. Avoid giving tests immediately before or after a long vacation, holidays or other important events where all students are actively involved physically or psychologically/emotionally.
4. Avoid giving tests when students would normally be doing something pleasant e.g. Having lunch, etc.

5. The sitting arrangement must allow enough space so that pupils will not copy each other's work.
6. Adequate ventilation and lighting are expected in the testing room.
7. Provision must be made for extra answer sheets and writing materials.
8. Pupils should start the test promptly and stop on time.

Scoring of Classroom Achievement Tests

According to Etsey (2004), essay tests can be scored by using the analytic scoring rubrics (also known as the point-score method) or holistic scoring rubrics (also called global-quality scaling or rating method). In analytic scoring, the main elements of the ideal answer are identified and points awarded to each element. This works best on restricted response essays. In holistic scoring, the model answer serves as a standard. Each response is read for a general impression of its adequacy as compared to the standard. The general impression is then transformed into a numerical score. To check the consistency of the scoring, a first reading is done to sort the responses into several piles (mostly five A, B, C, D, E) according to the different levels of quality. The analytic, point-score or the trait method basically involves the use in scoring of an already prepared list of points or ideas considered essential to a good answer to the question, together with the number of points (marks) allotted to each idea raised or discussed in the answer (Nitko, 2001; Mehrens & Lehmann, 1991). This is known as a marking scheme, a scoring rubric or a scoring key (Amedahe & Gyimah, 2003; Etsey, 2004).

The Holistic scoring rubric requires the marker to make judgement about the overall quality of each student's response. Teachers do not mark

each specific content elements that student included in the answer. According to Nitko (2001), “the Holistic scoring is probably more appropriate for extended respond essays involving a student’s abilities to synthesise and create and when no definite correct answer can be specified” (p. 195). The Holistic method is less objective than the Analytic method unless you have specified scoring criteria.

The scoring of essay-type tests according to Etsey (2004), is a highly important issue due to the fact that no matter how careful one is in writing the items, without equally taking careful steps to ensure consistency of scoring, the scores will not be reliable. The main reason for utmost care in the scoring of essay-type tests is the subjectivity involved. This is a major difference between the essay- and objective-type tests (Amedahe & Gyimah, 2003; Etsey, 2004; Gronlund, 2008). According to Mehrens and Lehmann (1991), the decision on a method of scoring for essay-type tests depends to some extent on the type of score interpretation desired (norm-referenced or criterion-referenced) and the amount of diagnostic information needed about an individual’s responses. It also depends on the time and facilities available for reading the papers and whether the essay is of the restricted- or extended response type.

In order to improve objectivity in the scoring and reliability of the scores of essay-type tests, Mehrens and Lehmann (1991); Amedahe and Gyimah (2003); and Etsey (2004) have suggested the following techniques or principles to be adopted by scorers.

1. Constantly follow the marking scheme when scoring. It is one thing deciding to score all papers uniformly using a scoring guide and

actually following the scoring guide constantly to achieve uniformity. Scorers should follow the marking scheme constantly as they score, as this reduces rater drift, which is the likelihood of either not paying attention to the scoring guide or interpreting it differently as time passes.

2. Prepare a form of scoring guide. This could either be an analytic scoring guide or a holistic scoring guide.
3. Comments should be provided and errors corrected on the answer scripts for students to facilitate learning. This is especially important in formative assessments, where the comments should focus on students 'weaknesses and strengths' while they are answering various items.
4. Scorers must also avoid being influenced by the first few papers they score since this can let them become too lenient or harsh in scoring other papers.
5. Score all responses item by item rather than script by script. Here, scorers must take one item at a time, and score all the responses to it throughout before going to the next item. This principle is to minimise the carryover effect on the scores and thereby ensure consistency.
6. Score the scripts anonymously. Scripts should be identified by code numbers or any other means instead of the names of students. This principle is to reduce the halo-effect. This happens when a scorer's general impression of a person influences how the paper is scored.
7. Keep previously scored items out of sight when scoring the rest of the items. This principle is to minimise the carryover effects and ensure consistency of the scores.

8. Randomly reshuffle the scripts when beginning to score each set of items. This will minimise the bias introduced as a result of the position of one's script. Research by Hales and Tokar (as cited in Mehrens and Lehmann, 1991) has shown that a student's essay grade will be influenced by the position of the paper, especially if the preceding answers were either very good or very poor. Mehrens and Lehmann (1991) have pointed out that randomly reshuffling of scripts is especially significant when teachers are working with high- and low-level classes and read the best scripts first or last.
9. Try to score all responses to a particular item without interruption. This is to avoid unreliability of the scores as a result of the grader's standards varying markedly due to excessive interruptions in the course of scoring.
10. Score essay-type tests only when you are physically sound and mentally alert. This is to say that essays must be scored at a congenial time. This is because it is known that consistency in scoring essay tests is a function of the time the paper is scored (Karpicke & Roediger, 2008). Over excitement, depression, and any type of psychological or mental disequilibrium will affect the consistency of the scores of essay-type tests.
11. The mechanics of expressions such as correct grammar usage, flow of expression, quality of handwriting, orderly presentation of material and spelling should be judged separately from subject matter correctness.

Empirical Review

Teachers Assessment Practices

Globally, several researchers have explored the assessment practices of teachers across different levels of the education latter, and in many different institutions. Among such studies discussed in this study include:

Sofo, Ocansey, Nabie, and Asola (2013), carried out a study on the assessment practices of physical education teachers among secondary schools in Ghana. Using purposive sampling, a total of 63 secondary physical education teachers, comprising 43 males and 20 females was selected as participants. An open-ended questionnaire was used in data collection. Data were analysed using qualitative content analysis comprising of both inductive and deductive analyses. The findings of the study indicated that teachers used observation, skill test, knowledge test, demonstration, peer observation, and oral report as assessment practices. The deductive analysis of the study further indicated that the highest percentage of assessment practice utilised by the teachers in practical lessons was teacher observation while that with the lowest percentage was oral report. Teachers also reported knowledge test and individual report as the highest and lowest percentage of assessment practices used in their theory lessons respectively. The findings of the study also revealed that most of the teachers who participated in the study used assessment for documenting learning, rather than for accountability purposes.

In another study in Morocco, Benzehaf (2017) conducted a study, by exploring teachers' assessment practices and skills. The study focused on High school English teachers in the town of El Jadida. The researcher employed both quantitative and qualitative methods in gathering the data for the study.

Specifically, the study made use of both questionnaires and interview guides for data collection. Forty (40) respondents which consisted of 24 male teachers and 16 female teachers were randomly sampled for the study. In the findings of the study, the teachers reported that they used varied number of assessment strategies ranging from home assignments to in-class written tests but mainly for summative purposes. With reference to frequency of assessment, some respondents indicated that they assessed their students once a month; whereas others stated they did so, twice a semester. The findings of the study further revealed that respondents also used a number of assessment practices. Among the assessment practices used by teachers included essay questions, true/false questions, and fill-in-the-blanks.

Besides the foregoing, Zhang, and Burry-Stock (2003) investigated teachers' assessment practices and self-perceived assessment skills within the framework of classroom assessment literature and the Standards for Teacher Competence in Educational Assessment of Students. The factor analytical technique was applied to study the relationship between the constructs of assessment practices and self-perceived assessment skills.

Teachers' assessment practices and self-perceived assessment skills were examined in a MANOVA design to determine how they may vary as a function of teaching level, content area, teaching experience, and measurement training. Their operationalisation of the constructs of assessment practices and self-perceived assessment skills overlap to some extent in the underlying dimensions they measure, yet each contains a certain degree of uniqueness. The similarity between assessment practices and self-perceived assessment skills was supported by a strong correlation coefficient of .71 and by similar patterns

of item loadings on four of the underlying dimensions they measure (paper-pencil test; standardized testing, test revision, and instructional improvement; performance assessment; and non-achievement-based grading). Where the two factor structures differ, the construct of assessment practices does a better job of subsuming inherently related activities under the same dimension than does that of self-perceived assessment skills. The findings of the study showed that teachers differ in their assessment practices due to the nature of classroom assessment outlined by teaching levels. The findings also revealed that there was a general difference between elementary and secondary teachers in terms of the assessment methods used and teachers' concerns for assessment quality. While secondary teachers rely mostly on paper-pencil tests and were concerned about the quality of assessment, elementary teachers often use performance assessment as an alternative.

Furthermore, Ogunkola (2013), investigated the instructional assessment practices pattern, techniques and challenges of science teachers in Barbados. The cross-sectional survey design was adopted for the study. The population for the study included all science teachers in secondary schools in Barbados. Science teachers included teachers of Biology, Chemistry, Physics, Integrated Science, Agricultural Science, and Human and Social Biology. Twelve out of 22 secondary schools were randomly selected. All the science teachers at the twelve schools formed the sample for the study. Data was collected from fifty-five secondary school science teachers.

For the data analysis, the Descriptive statistics (frequency and percentages), one-way Analysis of Variance (ANOVA), as well as one-sample t-test were used in analysing the research questions. The findings of the study

revealed that all the teachers used collaborative assessment practices. Such practices include discussing assessments with peers before administering and learning from other teachers through sharing of ideas about assessment practices that work. Ninety-six percent of the teachers reported that they assess students' skills and follow the guidelines of the assessment development process. The findings of the study further revealed that there was no significant difference in the instructional assessment practices of the male teachers and the female teachers. That is to say, both males and females did not differ in terms of their assessment practices. The findings of the study also revealed a non-significant difference in the teachers' instructional assessment practices and their experience, professional qualification as well as their academic qualification.

In addition, In Nigeria, Agu, Onyekub and Anyichie (2012) conducted a study for the development and validation of an instrument for assessing test construction skills of secondary school teachers. The study developed and validated a Test Construction Skill Inventory (TCSI) for assessing secondary school teachers' competencies in constructing classroom-based tests. After carrying out Factor analysis, the researchers found that 25 items were factorially valid. In the study, five hundred and forty-three (543) secondary school teachers in the Onitsha education zone, Anambra state, Nigeria were used as respondents, using the proportionate stratified random sampling. The sample comprised of 120 male and 423 female teachers. Teachers sampled for the study with less than ten years teaching experience were taken as 'less experienced' (n = 200) while those that have been in the teaching profession from ten years upwards were taken as 'experienced teachers' (n = 323).

The Cronbach's alpha was used to analyse data for research questions 1 and 2 while mean and standard deviation statistical procedures were used to analyse data for research question. The Test Construction Scale Inventory was found to be reliable with a coefficient of 0.73 and the secondary school teachers found almost all the 25 items important skills for quality classroom-based test construction. The finding of the study revealed that almost all the teachers took the following skills into consideration while constructing test items: outlined the content covered for the term before setting test items form the content covered, prepared a test blueprint as a guide in the test construction process, ensured that the items are measuring the determined objectives, prepared a marking guide while constructing the test items, consulted standard text books in the subject for guide, gave clear instructions to guide the test takers, submitted tests meant for promotional examinations for expert editing on time. The findings of the study, however, revealed that quite a number of the respondents did not consider the following skills while constructing test items: avoid the use of interlocking items, avoid items that measure opinion.

Ololube (2008), evaluated test construction skills of professional and non-professional teachers in Nigeria and reported that professional teachers tend to construct effective evaluative instruments more than the non-professional teachers. Ololube emphasised that professional teachers have the propensity to employ the various assessment techniques correctly, which is unlikely to happen in the case of non-professional teachers.

In a similarly vain, Quansah and Amoako (2018), developed and validated a standardised instrument for measuring teachers' attitude towards test construction. The instrument was developed based on literature as well as

personal experiences of the researchers. Through an exploratory factor analysis, four dimensions were obtained which include: planning, item construction, item review and assembling. A confirmatory factor analysis was then conducted to examine the factor loadings of the items. After critical evaluation, the items on the instrument remained 32 which was on a four-point Likert scale. The study further used the instrument to explore the attitude of teachers towards test construction. The instrument was administered to 432 Senior High School teachers in the Cape Coast Metropolis. The results showed that overall, teachers reported having a negative attitude towards test construction. Based on this result, is it not possible that this negative attitude of teachers will influence the way they practise testing in the classroom.?

The relationship between teacher characteristics and their classroom assessment practices among teachers in the Atwima-Nwabiagya south district in the Ashanti Region was investigated by Owusu-Mensah (2019). The descriptive survey, using a random sample of 219 teachers, was used for the study. For the data collection, the researcher adapted a 41-item self-perceived assessment practices questionnaire. In all, four research questions and one hypothesis were posed to guide the study. For data analysis, the frequencies, percentages, independent samples t-test, one-way ANOVA and the hierarchical multiple regression were used.

It was found that teachers in the district used a variety of assessment formats, but the most common assessment format used among the teachers was test. It was also revealed that teachers in the district considered the purpose of the test before developing test items, preparing marking scheme and also matched instructional objectives with the test. However, it was found

that the teachers reported they did not follow the test specification table while writing items. Further, when it comes to teachers' characteristics and their assessment practices, it was found that teacher characteristics had no significant effect on their assessment practices.

Gender and Classroom Assessment Practices

Alkharusi (2011) investigated the influence of gender, subject area, grade level, teaching experience, teaching experience and in-service assessment training on teachers' classroom assessment skills. The study employed a survey design using 213 public school teachers in Turkey. A 25-item self-perceived assessment skills inventory was used for the data collection. The results showed a significant gender difference in self-perceived assessment skills. Specifically, female teachers reported, on average a higher level of self-perceived skills in writing test items and communicating feedback than male teachers. The study also found that across subject areas, science teachers reported, on average a higher level of self-perceived skilfulness than English language teachers in writing test items, using performance assessment, and grading. However, English teachers reported, on average lower level of self-perceived assessment, communicating feedback to students than science teachers' science teachers. Similarly, a significant difference in self-perceived classroom assessment skills was found with respect to teacher qualification.

According to the results, while grade six teachers reported, on average, higher levels of self-perceived skilfulness than grade 10 teachers when comes to analysing test items and grading, Grade 10 teachers reported on average higher levels of self-perceived skilfulness in communicating classroom assessment feedback. In a similar fashion, a significant difference in

self-perceived classroom assessment skills was found among teachers based on teaching experience. Specifically, the findings showed that teachers with more than 10 years of teaching experience reported, on average, higher levels of self-perceived skilfulness in analysing test items, communicating assessment, feedback and writing test items than both teachers with 1-5 years of teaching and those with 6-10 years of teaching experience. Also, teachers with 6-10 years of teaching experience reported on average, higher level of self-perceived skills communicating assessment feedback than teachers with 1-5 years of teaching experience. However, when it comes to analysing test items, writing test items and using performance assessment, the results showed no statistically significant difference in self-perceived skills between teachers with 1 -5 years of experience and those with 6-10 years of teaching experience. The findings of this study, suggest that the demographic characteristics of teachers are associated with their testing practices and therefore, may influence they way they actually carry out testing in schools.

In Ethiopia, teachers' perception of classroom assessment was explored by Chalachew and Terefe (2020). The study specifically, investigated the influence of gender and teaching experience on teachers' perception of classroom assessment among High Schools of the South West Shewa Zone. The study adapted the Zhang and Burry-Stock (2003) Assessment Practice Inventory (API) for data collection. The inventory was administered to a randomly selected sample of 353 teachers, drawn from seven high schools. After pilot testing the instrument, modifications were made to the inventory and the researcher in person, finally distributed a total of 195. For data analysis, the study used the independent sample t-test to assess the gender

difference in self-perceived assessment skills, while one-way ANOVA was used to examine differences in self-perceived assessment skills across years of teaching. The results showed a significant gender difference in self-perceived skills in giving assessment feedback. The study also revealed a statistically significant difference on the basis of years of teaching, when it comes to constructing tests items, analysing tests results and communicating feedback.

Asamoah, Songnalle, Sundeme and Derky (2019) explored the gender differences in formative assessment knowledge of senior high school teachers in the Upper West Region. A descriptive survey was adopted for the conduct of the study. Using the simple random sampling technique, a total of 295 SHS teachers was selected. A questionnaire developed by the researchers was used for the data collection. The data were analysed using the independent samples t-test. The result showed a significant difference in the formative assessment knowledge of male and female SHS teachers. Specifically, the results showed that male SHS teachers reported on average a higher of formative knowledge than their female counterparts. The researchers did well by using a relatively large sample size (295). I believe that the use of this sample may permit the generalisability of the findings to a reasonable extent. However, the exploration of the teachers' formative assessment knowledge alone may not reveal more about how the teachers SHS teachers in the Upper West Region usually practice formative assessment in the classroom. Also, the researchers sought only to find out gender differences in formative assessment knowledge. The findings on gender differences alone may leave more to be desired, as far as SHS teachers' practice of formative assessment is concerned. With respect to gender and classroom assessment, the finding of Asamoah, Songnalle,

Sundeme and Derky (2019), supports the findings made by Mohuiddin (2015), who also examined attitudes, competence, knowledge and practices of teachers about educational assessment practices in Bangladesh, and the results showed that there was significant difference in educational assessment knowledge with respect to gender. However, the findings of both Asamoah, Songnalle, Sundeme and Derky (2019), and Mohuiddin (2015) disagree with the findings of Opara (2018), who investigated teachers' characteristics and how they could determine teachers' attitude towards the practice of continuous assessment in Ohio-Akpor Local Government Area of River States, and found that gender was not a good determinant of teachers' attitude towards the practice of continuous assessment. It was also found that Female teachers tended to report having, on average, a higher level of educational assessment knowledge than their male counterparts. The current study therefore, intends to fill this gap by investigating the contribution of gender and other equally important teachers' characteristics such as academic qualification, years of teaching experience and subject area to classroom achievement practices of Junior High School teachers in the Sissala East Municipality of the Upper West Region of Ghana.

Years of Teaching and Assessment Practices

Sahinkaraka (2012) explored the role of teaching experience on teachers' perception of language assessment. The study employed metaphors using the qualitative approach. The sample for the study included 53 students in pre-service education and 47 English language teachers. Out of the 47 practising- teachers, 23 had less than 5 years of teaching experience and 24 had more than 5 years of teaching experience. The findings showed that there

was not difference among the three groups of participants; However, the results revealed that more experience teachers conceived assessment as a negative factor more than less experienced teachers. The reason for this difference, I guess, could be that the more experienced teachers have had the opportunity of participating in several seminars and workshops and have realized the frustration students endure during examinations, especially high-stake examinations. Also, the use of metaphors in measuring perceptions of teachers in itself, may be problematic since interpretations of data from such metaphors may not be objective enough to reduce errors of measurement. The use of quantitative methods such as the use of questionnaire for data collection where simple means and standard deviations could have been computed since the study sought to explore differences in teachers' perception of assessment.

Mohuidin (2015) examined attitudes, competence, knowledge and practices of teachers about educational assessment practices in Bangladesh. The study employed the descriptive survey. In all, the study sampled 80 teachers, teaching various subject areas in the secondary and higher secondary levels and higher secondary colleges in Chittagong city. A questionnaire was used for the data collection. The analysis of the data, the study used the Multiple Analysis of Variance (MANOVA), Factorial Anova and the Pearson product-movement Correlation Coefficient. The results showed that though teachers reported a positive attitude towards classroom assessment, they demonstrated a low level of knowledge in educational assessment. The study also, revealed that difference in teachers' assessment practices was attributed to years of teaching experience. The study further revealed that there was significant difference in educational assessment knowledge with respect to

gender. Female teachers tended to report having, on average, a higher level of educational assessment knowledge than males. In a similar vein, Mathematics teachers reported, on average a higher level of educational assessment knowledge than English language teachers. The study further showed that while female teachers reported also to be more skilful in writing test items than male teachers, Science teachers similarly perceived themselves more skilful in developing and analysing assessment results than English Language teachers. Besides, years of teaching experience was reported to correlate positively with teachers' self-perceived assessment skills.

The weakness of this study first, has to do with the sample. The sample size of 80 for a survey appears to be relatively smaller to permit generalization. Also, the study merely focused on, and revealed differences in teachers' classroom assessment practices based on gender and years of teaching, but does not tell contributions of teacher factors such as gender, years of teaching and subject to the classroom assessment practices of teachers. This current study, therefore hopes to fill this void by using a relatively larger sample size to investigate the unique contributions of teacher characteristics such as gender, teaching experience, and subject area for classroom assessment practices of teachers.

Susuwele-Banda (2005) investigated the teachers' perception of classroom assessment in mathematics and their classroom assessment practices in Malawi. The study adopted a mixed approach, thereby using questionnaires, observations, interviews and document analysis of the data collection. Six teachers comprising 3 males and 3 females served as respondents in the study. The results showed that what teachers reported about

their assessment practices did not reflect on the actual classroom practice. Teachers perception of classroom assessment was also found to have an influence on the classroom assessment practices of teachers. However, while teacher's academic qualification indicated that it has has an influence on teachers' perception of classroom assessment, teachers' years of teaching experience did not seem to contribute to teachers' perception of classroom assessment. The researcher did well by using both the quantitative and qualitative methods of data collection in order to triangulate the findings. However, the relatively small sample size used by the researcher may make generalisations of the findings problematic. Especially, findings on the influence of teachers' academic qualification and years of teaching experience on the perception of classroom assessment practices. Particularly so, if the findings were obtained through the interview and observation only. This current study, therefore, intends to use a relatively more sample size in order to enhance the generalisability of the findings.

Anhwere (2009) investigated whether teachers in the teacher training colleges of Ghana follow the basic laid down principles of testing in three specific stages of planning a classroom assessment; test construction, administration and scoring of teacher-made tests. The study employed the descriptive survey design, involving 20 teacher training colleges, using a sample size of 310 teachers. The study used a sample of 310, comprising 230 male and 80 female teacher training college tutors. For the data collection, the study made use of a questionnaire. Means, standard deviations, frequency, percentages and the independent samples t-test were used for the data analysis. The result showed that the teacher training college tutors adhere to the basic

principles of testing, in the area of test construction of classroom assessment. The findings also, showed that the tutors of teacher training colleges found the conduct of classroom assessment as an extra burden. Further, the result showed that there was not statistically significant difference in the way the tutors follow the basic principles of test construction and administration on the basis on years of teaching experience. Specifically, no significant difference in the construction and administration of teacher-made tests was found between tutors who taught for 1 to 3 years and those who taught above three years. The results of Anhwere (2009), affirm the findings of Mohuiddin (2015), Sahinkaraka (2012), and Alkharusi (2011), whose studies found no statistically significant difference in teachers' self-perceived assessment skills, with respect to years of teaching experience. For instance, the findings Alkharusi (2011), showed that teachers with more than 10 years of teaching experience reported on average higher levels of self-perceived skillfulness in analysing test items, communicating assessment feedback and writing test items than both teachers with 1-5 years of teaching and those with 6-10 years of teaching experience. Also, teachers with 6-10 years of teaching experience reported, on average, higher level of self-perceived skills, communicating assessment feedback than teachers with 1-5 years of teaching experience. However, when it comes to analysing test items, writing test items and using performance assessment, the results showed no statistically significant difference in self-perceived skills between teachers with 1 -5 years of experience and those with 6-10 years of teaching experience. These findings (Anwhere, 2009; Mohuiddin, 2015, Sahinkaraka, 2012; & Alkharusi, 2011) however, are not consistent with the findings of Susuwele-Banda (2005), who

observed that teachers' years of teaching experience did not seem to contribute to teachers' perception of classroom assessment. A weakness of the study is that a cursory visual observation of the sample will easily reveal a huge difference in the sample size of the male tutors compared to the female tutors. The relatively unequal sample sizes of the independent groups of tutors (males and females) may affect the use of the independent samples t-test in testing the gender differences in tutors' testing practices. Another weakness of the study is that while it explored differences in testing practices of tutors with regards to the teachers' years of teaching experience, it failed to explore the proportion of the tutors testing practices that is explained by their years of teaching experience. A gap this current study hopes to fill.

Teacher Qualification and Assessment Practices

In Botswana, a study by Koloji-Keaitkitse (2012), explored the discrepancies between teachers' perceived skills and use of classroom assessment practices of primary and secondary schools. A descriptive survey design was adopted for the study. For data collection, the study used a questionnaire known as the Classroom Assessment Practices and Skills (CAPS) questionnaire. The instrument was administered to a total of 691 teachers sampled from primary, junior and senior secondary schools from various educational regions and inspectorate areas in Botswana. The results showed that teachers with only a certificate in teaching needed more training in assessment applications, statistical applications and criterion referenced testing, while the more experienced teachers reported higher on average on perceived skills and use of classroom assessment practices.

In Nigeria, Opara (2018) investigated teachers' characteristics and how they could determine teachers' attitude towards the practice of continuous assessment in Ohio-Akpor Local Government Area of River States. The study adopted an ex-post facto research design. A sample of 391 secondary school teachers was used as participants. The proportionate stratified random sampling technique was used to select the sample, based on educational qualification, years of teaching experience and gender in the study. An instrument titled Continuous Assessment Teachers' Attitude Scale, developed by the researcher was used for the data collection. The reliability of the instrument was reported to be 0.78. The data were analysed using percentages, means, standard deviation, chi-square, independent samples t-test and one-way analysis of variance. The result showed a significant difference in the proportion of teachers with positive and negative attitudes towards continuous assessment. Also, the findings revealed that while educational qualification and years of teaching experience were found to be good determinants of teachers' attitude towards continuous assessment practices, gender was not found to be a good determinant of teachers' attitude towards the practice of continuous assessment.

In another study, Gameda and Getachew (2019), studied instructors' perception of continuous assessment in the Madda Walabu University in South East Ethiopia. The study employed a descriptive survey (cross-sectional) method. Using the simple random sampling technique, a sample size of 225 was selected. For data collection, the study used a questionnaire, which had both closed and open-ended questions. With the aid of SPSS version 21.0, the data were analysed using frequencies and analysis of variance. The results

showed that generally, instructors in the Mada Walabu University had a positive perception about continuous assessment. The result also, showed that there was a statistically significant difference in continuous assessment perception among the instructors. With regard to educational qualification, the results revealed that first degree holders reported lower perceptions of continuous assessment, compared to PhD and Master degree holders. The weakness of this study was that the study only sought to find differences in instructors' perception of continuous assessment but did not explore the contribution of educational qualification to the instructors' perception of towards continuous assessment. This current study, therefore seeks to fill this void by examining the unique contribution of teachers' educational qualification and other characteristics such gender, years of teaching experience and subject area to their (teachers') classroom achievement test practices.

Subject Teachers Teach and Assessment Practices

Saefurrohman and Balinas (2016) investigated high school English teachers' classroom assessment practices in ELL class. The study was done using Filipino and Indonesian high school English teachers. Using both the quantitative and qualitative approach (mixed method design), sampled 48 Filipinos and Indonesian junior high school English teachers as respondents. For the data collection, the questionnaire, interview and observations were used. Frequencies, percentages, means and standard deviations were used to analyse the quantitative data while the qualitative data were analysed using thematic data analysis using Creswell model. The result showed that the majority of the Filipino junior high school English teachers prepared and

developed their own assessment while Indonesian junior high school English teachers used items from published textbooks mainly for constructing classroom assessment items. The weakness of this study is that the sample size was relatively small. Thereby, making generalisations from results of the sample problematic. Also, the study also did not explore the contributions of teacher characteristics and their contribution to classroom assessment practices of the English teachers. This present study, therefore, hopes to fill this gap by using a relatively more sample size and also, examine the contribution of teachers' characteristics to their classroom assessment practices in the Sissala East Municipality.

Muhiuddin (2015) conducted a study on the educational assessment practices of secondary and higher secondary teachers in Bangladesh. The study employed a descriptive survey design. For the data collection, a self-report questionnaire was administered to 80 in-service teachers who served as respondents, randomly selected from secondary school and higher secondary colleges in Chittagong city, Bangladesh. For data analysis, the study basically used descriptive statistics such as frequencies, percentages, means and standard deviation. Also, the study made use factorial analysis of variance (Factorial Anova) and multivariate analysis of variance. The results showed that generally, teachers have a positive attitude toward educational assessment. However, the findings showed that there was not statistically significant difference in attitude of teachers towards educational assessment with respect to teacher characteristics such as, gender, teaching class, pre-service training in assessment and inter-service training in assessment. Meanwhile, the results showed a statistically significant difference in attitude of teachers towards

educational assessment regards to teaching subject. Specifically, teachers teaching English language and social studies tended to have on average less favourable attitude towards educational assessment than those teaching other subjects such as Bangla Language, Business studies, and Mathematics.

Quansah, Amoako and Ankomah (2019) examined the test construction skills of senior high school teachers in the cape coast metropolis. Using qualitative document analysis, the researchers examined samples of End-of-Term of term examination papers of in three subject areas; Integrated science, core mathematics and social studies in three randomly selected Senior High School. The results showed that teachers have limited skills in the construction of end of term examination. The findings of Quansah, Amoako and Ankomah (2019) is consistent with Mohuiddin (2015), who examined attitudes, competence, knowledge and practices of teachers about educational assessment practices in Bangladesh and found that Mathematics teachers reported on average a higher level of educational assessment knowledge than English language teachers.

Conceptual Framework

Figure 1 illustrates the conceptual framework that undergirded this study. The study adopted the four teachers' demographic characteristics /dimensions to explain how these variables help predict teachers' testing practices with specific reference to classroom testing practices.

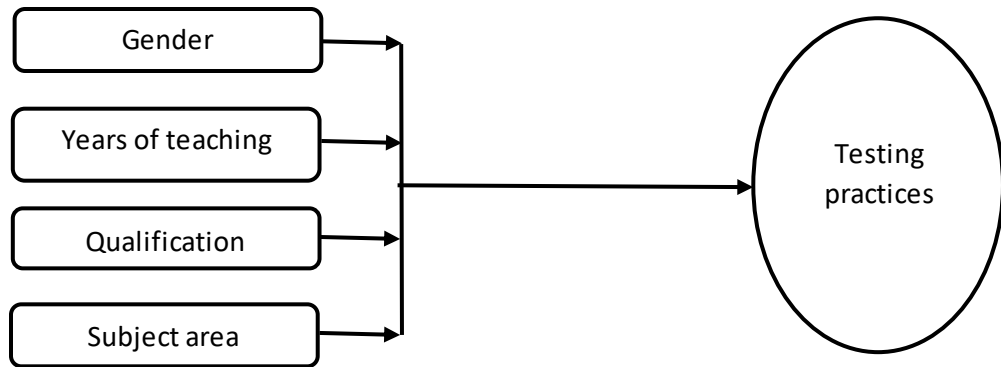


Figure 1: Conceptual framework depicting the relationship between teachers' demographic characteristics and teachers' testing practices

Source: Researcher's construct adapted from literature.

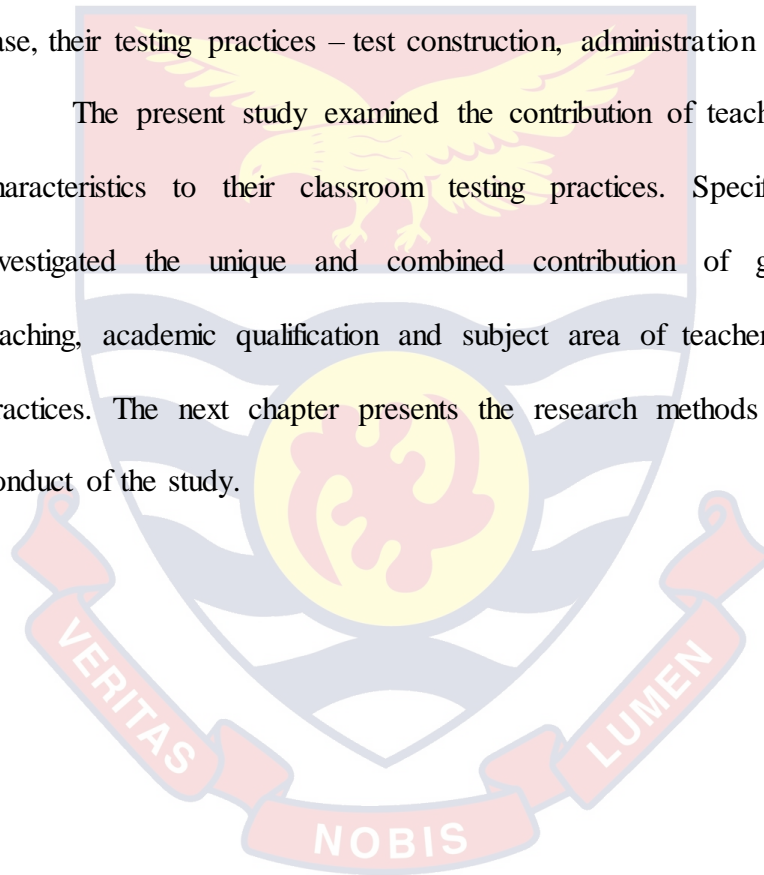
The conceptual framework suggests that teachers testing practices have a relationship with their demographic characteristics. Put another way, teachers' gender, thus, being male or female has an influence on how he/she adheres to testing principles when carrying out classroom assessment. Also, the number of years a teacher has taught influences the testing practices of teachers. Similarly, the level of education of the teacher also, has a relationship with how he/she is likely to follow testing principles during classroom assessment. Finally, the teaching subject of the teacher, thus, being a particular subject teacher influences how the teacher is likely to follow testing principles.

Summary of Literature Review

The chapter focused on the empirical literature reviewed on the relationship between teacher demographic characteristics and their classroom assessment practices. It emerged clearly from the studies that the previous researchers tackled the issue, using three different approaches; while first group of researchers focused on the attitude of teachers towards the practice of classroom assessment, the second group approached it by focusing on

teachers' perception about their assessment practices, and the third group focused on the differences and correlations between teachers' demographic characteristics and their assessment practices, even though the majority of the studies were conducted in other where the educational system may be different from Ghanaian, the findings in each case, were contradictory. It appears, however, that there is no known study that focused on the contribution of teachers' demographic characteristics to their assessment practices, in this case, their testing practices – test construction, administration and scoring.

The present study examined the contribution of teachers' demographic characteristics to their classroom testing practices. Specifically, the study investigated the unique and combined contribution of gender, years of teaching, academic qualification and subject area of teachers to their testing practices. The next chapter presents the research methods employed in the conduct of the study.



CHAPTER THREE

RESEARCH METHODS

This chapter covers the methodology employed in carrying out this study. The methods and approaches as described in the chapter are under nine sub-sections. These are: The Research Design, the Study Area, Population, Sample and Sampling Procedure, Validity and Reliability of the Instrument, Pilot-testing Procedure, Training of Research Assistants, Data Collection Procedure and Data Analyses.

The purpose of the study was to investigate whether teachers' demographic characteristics would predict their testing practices among Junior High Schools in the Sissala East Municipality of the Upper West Region of Ghana. Specifically, the study sought to examine the proportion of teachers' testing practices that is explained by their demographic characteristics. That is to say that the study sought to find out the contribution of teachers' demographic characteristics towards the way they practice classroom assessment. Thus, when it comes to the way teachers follow the basic principles of constructing, administering, and scoring classroom achievement tests.

Research Design

Polit and Beck (2004) defined a research design as the overall plan that guides the process of obtaining answers to the questions being studied and for responding to some of the difficulties that the researcher will encounter during the research process. Every research is unique and therefore, research designs

are developed to cater for the unique requirements of a study. Some researchers argue that in selecting a good research design, consideration should be whether the design does the job of providing reliable answers to the research questions (LoBiondo-Wood & Haber, 1998; Polit & Beck, 2004).

To achieve the objectives of the study, a quantitative research research appr. According Polit and Beck (2004), a quantitative research lends itself to studying phenomena that require precise measurement and involves quantifying using rigorous and controlled design. Quantitative designs basically rest upon numbers converted into statistics in order to enable the researcher to make interpretations from obtained data and research conclusions (Cormack, 2013). He indicated that this feature of the research study in accordance with the quantitative research paradigm.

In this study, I considered the descriptive (cross – sectional) survey design to be the most suitable research design to to be used. In the view of Cohen, Mansion, and Morrison (2007), a descriptive research is concerned with associations that exist, beliefs, point of views or attitudes that are held, practices that prevail and processes that are ongoing. The term survey, here is used to connote any research activity in which the researcher gathers data from a proportion of a population with the aim of examining the characteristics, opinions or intentions of that population (Couchman & Dawson, 1995; Polit & Beck, 2004). The descriptive survey was used for this study because of its high degree of representativeness and the ease with which the researcher could obtain the participants' opinion (Polit & Beck, 2004). Similarly, Amedahe (2004), indicated that descriptive research is research which specifies the nature of a given phenomenon (p. 50). Gay (as cited in Amedahe, 2004),

explains that descriptive research allows for the collection of data in order to test hypotheses or answer research questions concerning the current status of the subjects of the study.

According to Murphy (2009), the use of descriptive survey provides multiple sources of data collection. He indicates that the use of this type of survey has several advantages. For example, a survey can provide statistics about an event while also illustrating how people experience that event. Again, he states that the descriptive research design also offers a unique means of data collection. Notwithstanding the strengths of the descriptive survey as aforementioned, Fraenkel and Wallen (2000) cautioned that the use of a descriptive survey comes with several weaknesses, which include (1) difficulty in ensuring that all questions to be answered are clear and not misleading; (2) getting respondents to answer diligently, thoughtfully and honestly; (3) getting a sufficient number of questionnaires well completed and returned for the researcher to make meaningful analysis is often a major setback.

Study Area

The study was conducted in the Sissala East Municipality. The Sissala East Municipality, until the year 2004, was part of the Sissala District. The Sissala District was divided into two districts: Sissala East and Sissala West. In about fourteen years later, thus, from the year 2004 to the year 2018, the status of the district changed and attained the Municipality status in the year 2018. The municipality is one of the 11 Municipalities and Districts in the Upper West Region. The Municipality is located between Latitude: 10° 52' 58.80" N Longitude: -1° 58' 58.80" W. The Municipality is bounded to the

North -Western part of Ghana in the Upper West Region. Tumu is both the traditional and administrative capital of the municipality. The Municipality shares a 300-kilometer border with Burkina Faso to the north.

The Municipality shares close boundaries to the east with the kasena Nankana Municipal and Builsa District, both in the Upper East Region. To the south-eastern portion, it shares a boundary with the Mamprugu Moaduri District in the Northern Region and to the west, the Municipality shares boundaries with her sister District; Sissala west district and to the south, with WA East District. The population of the municipality stands at 56,528 with 27,503 males and 29,025 females (Feed the Future Ghana District Profile Series, 2017). The municipality, has, in all, one College of Education, one Midwifery Training College, two senior high schools, one vocational institute and 135 basic schools, with a total teacher population of 723 (GES, 2020). Out of this number, 483 are teachers in the 78 Junior High Schools in the municipality.

The Municipality was considered for the study because it is one of the Municipalities in the Upper West that has a college of education and also a study centre for training teachers. Therefore, teachers in the municipality have more access to upgrading themselves academically compared to their counterparts in the other districts and municipalities within the Region.

Population

Agyedu, Donkor, and Obeng (1999) defined a population as the entire set of individuals (subjects, objects, events) which bear certain peculiar features that the researcher is interested in studying. Fink (2015), argues that in a survey, the conditions for including a unit in a survey are based on the

characteristics of respondents who are eligible for inclusion in the survey. The target population is about the group of people, which the researcher is interested in collecting information and drawing conclusion about. Thus, the group of individuals who have certain similar characteristics, that the researcher is interested (Amedahe, 2004). The target population for this study was all the public basic schools (135) in the Municipality and the accessible population public Junior High Schools (78) in the Sissala East Municipality, of the Upper Region which was spread across in nine (9) educational circuits. Out of the 483 teachers, 196 (41 %) were females and 287 (59%) were male teachers (GES, 2020).

Sample and Sampling Procedure

A sample consists of a carefully selected unit that comprises all the categories of the population (Sarantakos 2005). Sarantakos asserts that estimation of the sample size varies significantly, with some researchers making a case for quantity, whilst others argue for quality and yet others advocating for combination of the two methods. Corroborating this, Field (2011), indicated that in determining the sample in regression, there should be ten (10) or 15 cases of data per each predictor. Green (1991) (as cited in Field, 2011), suggests two rules of thumb in determining the sample size for regression. He noted that to test for the overall fit of the regression model, a minimum sample size of $50 + 8K$, where, he explained K as the number of predictors. However, to test for the individual contribution of the predictors, he suggests a minimum sample size of $104 + K$. According to Green, most researchers are usually interested in testing for both the overall model fit as well as the contribution of the individual predictors. Hence, he argues that

both the minimum sample size for the individual contribution of the predictors and the overall model fit be calculated and the one with the larger value be chosen for the regression.

The study made use of a multi-stage sampling procedure. First, a simple random sampling technique was used by the researcher to select five (5) out of the nine circuits in the Municipality. The five selected circuits had a total of 61 schools. According to Krejcie and Morgan (1970) sample size determination table, for a survey research, a sample size of 52 randomly selected from a population of 61 is sufficient for the study. However, the study made use of 55 schools. The number was increased to 55 schools because the researcher believed that would enhance the return rate. The five circuits were selected randomly for the study because; the circuits in the Municipality are generally similar in characteristics. For, examples, teachers in all the circuits in the Municipality, were trained in similar colleges of education in Ghana. Similarly, social interventions rolled out by the Central Government or other Non-Governmental Organisations (NGOs), with the view to improving the educational system in the Municipality, are usually, evenly distributed among all the schools in the Municipality and therefore, all teachers benefit similarly. Moreover, all the teachers in the various circuits of the Municipality are supervised by the same Municipal Education Directorate staff. The second stage involved selecting 55 schools, from the five selected circuits, using a proportionate stratified sampling technique. The proportionate stratified sampling was used to select 55 schools from the five selected circuits. This was done because of the differences in the numerical strength of schools in the five selected circuits. Therefore, in order to enhance representativeness, I

chose to select the schools proportionately using the stratified random sampling technique. Based on the proportions, 12 out of 13 schools were selected from circuit one (1). In circuit two (2), 11 out of 12 schools were selected. In circuit three (3) and four (4), 10 out of 11 schools were selected, and finally, in circuit five (5), 12 out of 14 schools were selected.

The last stage involved purposively selecting the four-core subject (English Language, Mathematics, Integrated Science and Social Studies) teachers in each of the selected schools to serve as respondents of the study. In all, the 55 schools gave a total of 220 teachers as the final sample.

Data Collection Instrument

For the data collection, a Questionnaire developed by Oduro-Okyireh (2008) was adapted for the data collection. The instrument was used to measure the testing practices of Secondary School teachers in the Ashanti Region of Ghana. The original questionnaire was structured based on three aspects of testing: Test construction, Test administration and Test scoring. In the instrument was reported to have a Cronbach's Alpha Reliability Coefficient of 0.70. A questionnaire with a reliability of 0.70 and above is appropriate for data collection on the construct(s) of interest (Pavet, Diener, Colvin & Sandvik, 1991). According to Sarantakos (2005), the use of questionnaire for data collection affords greater assurance of confidentiality and anonymity to respondents. Besides, the researcher chose to use questionnaire for the study because it would offer the researcher the opportunity to select large sample of respondents. The items on the original questionnaire were modified based on the objectives of my study, in order to elicit the needed information.

In the original questionnaire, 10 items were used to measure test construction practices. However, in my study, I used 9 out of the 10 items in the original questionnaire without any modification. But the remaining item which was stated as “I evaluate the test as whole to find out whether: the test items are simple and clear; the test is a representative sampling of the material taught; the students will have enough time to complete the test; the test is the best instrument to measure the desired knowledge; the students have prepared adequately for the test, etc, before submitting for typing”. In my study, I modified this item by turning each of the statements into an item. This modification was necessary because each of the five (5) statements that were lumped together in the tenth item could stand as a major principle of test construction. Also, with advice from measurement and evaluation experts in my department, five (5) additional items measuring test construction practices were crafted, reviewed and added to the instrument, thereby bring the total number of items measuring test construction to nineteen (19). For test administration, the original questionnaire had 18 items. In my study, 11 of the items were adapted by rewording them. The items were reworded because in the original questionnaire, the items were crafted using measurement jargons that I believed the respondents in this study would have difficulty understanding them. For example, measurement jargons like ‘test items’; ‘scoring rubric’ were reworded as ‘questions’ and ‘marking scheme’ respectively. It is worthy to note that the seven (7) other items were omitted because they were eliciting the same information as those adapted, from the respondents. Therefore, there was no need using all of them. Similarly, the original questionnaire had 10 items for test scoring, which were stated in a

dichotomous form using Yes or No. In my study, I adapted all the 10 items in but stated them in the continuous form using the Likert-type scaling. Again, with expert advice from measurement and evaluation experts in my department, 4 additional items measuring scoring were added. These additional items were deemed to very relevant for assessing how tests are scored. However, these items were missing in the original questionnaire.

In all, the adapted questionnaire for this study was a closed-ended type. The questionnaire was a five-point Likert-type scale made up of a total of 44 (19 for test construction, 11 for test administration and 14 for test scoring) items with responses ranging from “Always-A, Very Often-VO, Sometimes-ST, Rarely- R and Never-N”. Out of the 44 items on the questionnaire, 35 of them were positively stated while the remaining 9 of them were negatively stated. To score the responses on the scale, all responses to the positively stated items took a score 5 for Always, 4 for Very Often, 3 for Sometimes, 2 for Rarely and 1 for never. However, all responses to the negative items were scored in the reversed order. Thus, always took a score of 1, Very Often took a score of 2, Sometimes, a score of 3, Rarely took a score of 4 while Never took a score of 5. For negatively stated items, the weighting is reversed because disagreement with a negative statement is psychologically equivalent to an agreement with a positive statement (Ary, Jacobs & Sorensen, 2010).

The final instrument for this study was organised into four sections (A, B, C, and D). Section ‘A’ comprised the background information of the teachers. In this section teachers were asked to provide information on gender, number of years in the teaching service, the highest academic qualification and subjects that they teach. The Section, ‘B’ comprised of 19 items asking

teachers in Junior High Schools in the Sissala East Municipality to indicate how frequently they follow the basic principles that guide test construction ranging from 'Always' to 'Never'. "Section C" constituted 11 items on how frequently Junior High School teachers in the Sissala East Municipality adhere to the basic principles that guide test administration, ranging from 'Always' to 'Never'. "Section D" was the final section of the instrument, and this section focused on how Junior High School teachers in the Sissala East Municipality adhere to the basic principles of scoring a test. The section consisted of 14 items with responses ranging from 'Always' to 'Never'.

Validity and Reliability of the Instrument

According to Dambudzo (2005), the idea of validity hinges on the extent to which research data and the methods of obtaining the data are considered accurate, honest and on target. Practically, the validity of an instrument is assessed in relation to the extent to which evidence can be generated in support of the claim that the instrument measures the construct targeted in the proposed research. To ensure the content relatedness of the questionnaire, the instrument was evaluated by Measurement and Evaluation experts in the Department of Education and Psychology, University of Cape Coast. All modifications based on errors and other expert-advice to the instrument were made.

Subsequently, to establish the reliability of the instrument, the Cronbach's Alpha Correlation Coefficient was used to estimate the internal consistency. Reliability reveals that when procedures of the study are repeated, the exact same results are expected (Mugenda & Mugenda, 2003). A reliability test was carried out with the purpose of testing the consistency of

the research instruments. For the reliability of the instruments, a pilot-test of the instrument was carried out on teachers in the Sissala West District. The Sissala West District was selected for the pilot -test because teachers in that District share similar characteristics. Most teachers in both the Sissala East Municipality and the Sissala West District had their pre-service training from the same College of Education (Tumu College of Education). Both set of teachers usually benefit in similar ways any time some interventions are designed and implemented by the few Non- Governmental Organisations in the Upper West Region, in order to enhance the capacity of teachers in line teaching and learning.

Pilot-Testing

For the pilot-testing, 20 schools in the Sissala West District were randomly chosen. The District was used for the pilot-test because it has similar characteristics with the Sissala East Municipality (current study setting). The respondents for the pilot -testing were randomly selected using the lottery method from 20 purposively selected schools in the District. Teachers teaching the four core subjects (English Language, Mathematics, Integrated Science and Social Studies) were purposively selected. In all, 80 (55 male and 25 female) teachers responded. The questionnaire was given to the teachers by the researcher. Before then, the purposive of the questionnaire was explained to the respondents. The respondents were told to discuss verbally and frankly with me any ambiguity, incoherence or miscomprehensions that they experienced detected on any aspect of the questionnaire. The questionnaire was retrieved after one week. The necessary corrections were made after the trial testing based on some the errors and challenges suggested by the

respondents. The responses were coded analysed with the aid of the Statistical Product for Service Solution (Version 23.0), to determine the reliability of each of the dimensions instrument with the Cronbach's Alpha measure of internal consistency. The Cronbach's Alpha Correlation Coefficients were as follow; test construction (.70), administration (.71) and scoring (.68).

Data Collection Procedure

In the conduct of the study, I followed strictly, the essential issues guiding the conduct of any research work. For instance, confidentiality, anonymity, and privacy were adhered to throughout the conduct of the study. First, I thoroughly explained the purpose of the study to the respondents (teachers), after which the consent of each of them was sought. Before responding to the questionnaire, respondents were made aware that engagement in the study was by choice and not compulsion. The data that were collected were kept confidentially, names of respondents or their schools were not disclosed in any part of the work (Maree, 2007). The data collected were analysed as a group, and, for that matter, it was not possible to trace responses to respondents. The data collected were well-managed and kept secretly to avoid the accessibility of other people). The discussion of the findings was also based on the trends that emerged from the data and not from any preconceived ideas.

A letter of consent to participate in the study was given to each of the teachers. An introductory letter (see Appendix A) from the Department of Education and Psychology and an ethical clearance letter ((see Appendix B) authorised by the Ethical Review Board, both from the University of Cape Coast were used to apply for permission from the Municipal Education

Directorate before going to the selected schools for the data collection. The permission letter from the Municipal Education Directorate was shown to the teachers as evidence indicating to them that they had the permission to participate in the study.

Before going to the field for the data collection, I trained two research assistants to aid in the collection of the data. The questionnaire was administered to the respondents within a period of three weeks. Before administering the questionnaires, the permission letter from the District Directorate was first shown to each headteacher of the selected schools. Also, I, met all the teachers who teach the four core subjects and discuss the purpose of the study with them and also went through the questionnaires with them for any possible clarification on the instrument. I sought permission from the heads of the schools in order to administer the questionnaire after the introductory letter and a permission letter from the Municipal Directorate office had been presented to the head teachers for their cooperation. On the 17th to 21st of February, 2020, the questionnaire was administered to circuits one and two by the researcher with the assistance of the research assistants. The questionnaire for circuits three and four were administered on the 24th to 28th of February, 2020 and then finally, two days in the third week were used to administer the questionnaire and the three remaining days were used to go back to each school to retrieve the questionnaires and spanned from 2nd to 6th of March, 2020.

In each school, I explained the purpose of the study to the head and the sampled teachers, and assured them of anonymity and confidentiality of their participation in the study. The teachers also had the opportunity to seek

clarification on the questionnaire before responding to them. This was because it would help to reduce respondents' biases and prejudices (Trochim, 2000). This ensured good contact with the teachers to further explain the purpose of the study so that I would obtain the commitment and cooperation of the teachers. This was done to ensure that the teachers responded to items on the questionnaire and submitting them in good time. In all, the researcher used three weeks in collecting the data for the study.

Data Processing and Analysis

The data was sorted, coded and cleaned for errors. Data on demographical variables of respondents were analysed using percentages and Frequencies. These tools were used since the researcher only wanted to describe the demographics of the respondents.

Hypotheses 1, 2, 3 and 4 were tested using Multivariate Multiple Linear Regression. Hypothesis 5 was tested, using the Hierarchical Multiple Linear Regression. The researcher used these data analysis tools, since, first, the researcher sought to find out if there was any relationship between teachers' demographic characteristics and their testing practices. Also, since, the researcher wanted to find out the proportion of teachers' testing practices that was explained by their demographic characteristics. Therefore, these data analysis tools were considered most appropriate.

CHAPTER FOUR

RESULTS AND DISCUSSION

The chapter sought to examine teachers' demographic characteristics as predictors of testing practices among Junior High schools in the Sissala East Municipality. Questionnaire was used to gather data from the respondents. Out of the 220 questionnaires administered, 144 of them were completely responded to and returned. This resulted in a response rate of 65%. In the view of Fincham (2008), the main target of researchers is to obtain a response rate of approximating 60%. This view, corroborates the assertion by Amedahe and Asamoah-Gyimah (2015), that a 5% to 20% of the population size is enough for the purpose of generalisation for quantitative research. Hence, this number was deemed valid for the purpose of the study and therefore, all the analyses in the chapter were based on the 144 respondents. The chapter has the description of the sample characteristics, followed by the analysis of the main data and then the discussion of the result.

Demographic Characteristics of Respondents

This section presents results on the respondents based on the demographic distribution. The demographic information includes gender, years of teaching experience, academic qualification and teaching subject. Table 1 presents details of the demographic distribution.

Table 1- *Distribution of Respondents by Demographic Characteristics (n =144)*

Variable	Frequency	Percentage (%)
Gender		
Male	123	85.4
Female	21	14.6
Years of teaching		
5 Years or below	53	36.8
6-10 years	52	36.1
11-15 years	27	18.8
16-20 years	6	4.2
20 years or above	6	4.2
Highest academic qualification		
Cert A	1	0.7
Diploma in education	70	48.6
Bachelors with education	62	43.1
Bachelors without education	8	5.6
Masters with education	2	1.4
Masters without education	1	0.7
Teaching subject		
English language	35	24.3
Mathematics	37	25.7
Integrated science	36	25.0
Social studies	36	25.0

Source: Field Survey (2020)

Distribution of Respondents Based on Gender

The results in Table 1 shows that, out of a sample of 144, 123 (85.4 %) of them were males whereas 21 (14.6 %) were females. This shows that there were more male teacher respondents than their female counterparts. This is not surprising, since the population of teachers in the Municipality is dominated by male teachers, compared to their female counterparts. Besides, out of the five circuits randomly sampled for the survey, four (4) of the them are far apart in the Sissala East Municipality. In my view, the reason for the male teacher dominance in these circuits is not surprising, because, during the data collection, I observed that traveling to such schools is mostly problematic for

even the male teachers who must ride motor-cycles before they are able to go to work. The reason for this situation, largely is due to the long distance between the communities in the Municipality and generally, due to the poor road network connecting the communities. As a result of this, teachers posted to such schools are mostly males.

Distribution of Respondents Based on Years of Teaching

The results in Table 1 reveal that majority of the teachers, 53 (36.8%) have taught for 5 years or less. The results in Table 1 also show that 52 (36.1) of teachers reported that they have taught for at least, 6 to 10 years. In addition, the results show that 27 (18.8%) of the respondents have reported that they have taught for at least 11 to 15 years. Based on these results, it can be seen that 79 (54.9%) have reported that they have taught for between 6 to 15 years. This therefore, suggests that majority of the teachers in the Sissala East Municipality sampled for the study, may have enough experience when it comes to assessing their students. In this case, when it comes to the construction, administration and scoring of classroom assessment.

Distribution of Respondents Based on Educational Qualification

Regarding academic qualification, the results in Table 1, show that 70 (48.6%) of the teachers reported they have Diploma in education certificate, and closely following this number, was 62 (43.1%) of the respondents who also reported that they have the bachelors with education certificate. Based on the results, it can be seen that majority of the teachers 132 (91.7%) have either diploma in education or bachelors with education. Only 8 (5.4%) had bachelors without education. Also, 2 (1.4 %) of the teachers have masters with

education while only 1 (0.7%) each of the teachers reported having masters without education and Certificate 'A' respectively.

Distribution of Respondents Based on Subjects Teachers are Teaching

The data in Table 1 show that majority 37 (25.7 %) of the teachers reported that they teach Mathematics, follow by 35 (24.3%) who also reported that they teach English language. For integrated science and social studies, 36 (25%) reported they are teaching these subjects respectively. When it comes to the subjects the teachers are teaching, the results in Table 1 show that the teachers are evenly distributed across the four subject areas in the Sissala East Municipality.

Testing Practices of Teachers

The study sought to examine teachers' demographic characteristics as predictors of testing practices. Testing, essentially, is broad and with several dimensions. For example, test construction, administration, scoring, grading, feedback. However, the study focused on only three of these dimensions; test constructions, test administration and scoring.

Test construction Practices of Teachers

The dimension of test construction practices was measured using nineteen (19) items on a five-point Likert -type scale. Out of the 19 items, fourteen (14) were positively stated, where, (Always-5, Very Often-4, Sometimes-3, Rarely-2, Never-1). The rest (5) of the items were negatively stated where, (Always-1, Very Often-2, Sometimes-3, Rarely-4, Never-5). Mean and standard deviation were used to analyse the data gathered on this construct (test construction). A midpoint of 3.0 was used as cut off score, such that a mean value above 3.0 indicated that the teachers stated that they follow the principles of testing

when constructing test items while a mean value less than 3.0 is an indication that the teachers are not adhering to the basic principles guiding test construction. Results on the test construction practices of teachers are presented in Table 2.

Table 2– *Test Construction Practices of Teachers*

Statement	Mean	SD
I relate instructional objectives of subject matter to test	4.73	.58
I inform students in advance before test	4.44	.84
I inform students about uses of the test scores after the test is written	3.62	1.01
I select test format suitable for testing stated objectives	4.47	.79
I select test format convenient for scoring the test	4.22	.99
I prepare and use table of specification	3.48	1.14
I inform students what chapters the test will cover	4.08	1.13
I prepare more questions than needed	2.90	1.29
I write questions in advance (at least two weeks) for review	3.47	1.26
I write questions a day before the test to avoid leakage	3.40	1.38
I review questions (at least one week) after they have written	3.48	1.29
I ensure questions are challenging to high achievers	3.23	1.35
I prepare marking scheme immediately after the test is written	3.82	1.34
I write clear and concise directions for the entire and all sections of the test	4.58	.73
I ensure the questions are simple and clear	4.17	.94
I ensure test is a representative sampling of content covered	4.67	.66
I ensure students have enough time to complete the test	4.20	1.01
I ensure test is the best instrument to assess the knowledge or skills taught	4.09	1.04
I prepare the students adequately for the test	4.51	.80
Overall Mean/Mean of Means (MM)	3.98	1.03

Source: Field Survey (2020)

The results in Table 2, show that thirteen (13) out of the nineteen (19) principles of test construction had mean values above the cut-off mean of 3.0 implying that the teachers adhere to principles of test items writing. examples of some of the principles of test construction that the teachers reported that they adhere to were, “I relate the instructional objectives of the subject matter to the test” ($M = 4.73$, $SD = .58$), “I inform students in advance before the test” ($M = 4.44$, $SD = .84$), “I select test format suitable for testing stated objectives” ($M = 4.47$, $SD = .79$), “I inform students what chapters the test will cover” ($M = 4.08$, $SD = 1.13$). The data in Table 2 show that the teachers indicated that they did not follow one (1) of the principles of test construction. Thus, “I prepare more questions than needed for the test” ($M = 2.90$, $SD = 1.29$). Also, the results in Table 2 show that the teachers in the Sissala East Municipality reported that they flouted five (5) basic principles guiding test construction. This is evident in the means of their responses, when I asked the teachers the following questions; “I inform students about how the scores will be used, immediately after the test is written” ($M = 3.62$, $SD = 1.01$), “I select the test format that is convenient for scoring the test” ($M = 4.22$, $SD = .99$), “I write the test questions a day before the test to avoid leakage of questions” ($M = 3.40$, $SD = 1.38$), “I ensure that the test questions are challenging to the high achievers who think they are very intelligent” ($M = 3.23$, $SD = 1.35$), and “I prepare the marking scheme immediately after the test is written” ($M = 3.82$, $SD = 1.34$). A mean of means value of Mean = 3.98, $SD = 1.03$ however, shows that teachers generally adhere to the principles of test item writing (construction) practices.

Test Administration Practices of Teachers

The dimension of test administration practices was measured using eleven (11) items on a five-point Likert -type scale. Out of the 11 items, seven (7) were positively stated, where, (Always-5, Very Often-4, Sometimes-3, Rarely-2, Never-1). The rest (4) of the items were negatively stated where, (Always-1, Very Often-2, Sometimes-3, Rarely-4, Never-5). Mean and standard deviation were used to analyse data gathered on this construct (test administration). A middle point of 3.0 was used as cut off score, such that a mean value above 3.0 indicated that the teachers stated that they follow the principles of testing when administering test while a mean value less than 3.0 is an indication that the teachers are not adhering to the basic principles guiding test administration. Results on the test administration practices of teachers is presented in Table 3.

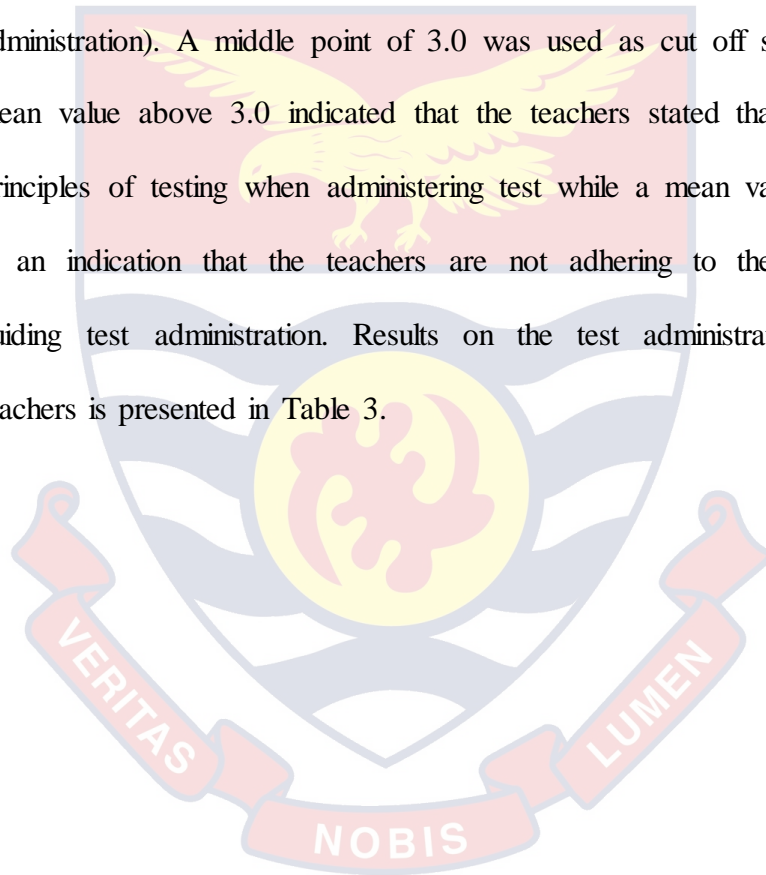


Table 3– *Test Administration Practices of Teachers*

Statement	Mean	SD
I give test immediately before long vacation or other important events	3.51	1.39
I give test immediately after long vacation or other important events	2.43	1.12
I ensure sitting arrangement allows enough space to avoid copying	4.38	.95
I ensure there is adequate ventilation and lighting in the testing room	4.79	1.45
I use-'DO NOT DISTURB' test in progress sign when students are taking test	3.38	1.47
I expect and cater for all possible emergencies	4.39	.96
I announce remaining time (left to complete work) at regular intervals	4.67	.70
I stand where I can view all students and move once a while to check malpractices	4.71	.65
I keep all remarks in the testing room to a minimum and make sure they are related to the test	3.91	1.10
I ask students to work faster during testing in order to finish on time	3.55	1.28
I tell students the dire consequences of failure in the test they are taking	3.43	1.41
Overall Mean/ Mean of Means (MM)	3.92	1.13

Source: Field survey (2020)

The data in Table 3 show that five (5) of the items measuring the test administration practices of teachers recorded mean values above 4.0. They ranged from Mean = 4.38, *SD* = 0.95 to Mean = 4.79, *SD* = 1.45, which is far above the cut-off mean of 3.0, implying teachers adhere to such principles

when administering their test. Where the item, “I ensure that there is adequate ventilation and lighting in the testing room” was put, a highest Mean = 4.79, $SD = 1.45$ was obtained. The item which recorded the next highest Mean = 4.71, $SD = 0.65$ was the item “I stand where I can view all students and move among the students once a while to check on malpractices”. It can also be seen from the data in Table 3 that two (2) other items also had high means ranging from Mean = 3.38, $SD = 0.95$ to Mean = 3.91 to $SD = 1.10$, also implying that teachers adhere to the test administration principles. Also, teachers kept all remarks in the testing room to a minimum and make sure they are related to the test ($M = 3.91$, $SD = 1.10$). The data also show that teachers do not give test give tests immediately after a long vacation, or other important events ($M = 2.43$, $SD = 1.12$). It can also be seen in Table 3 that teachers flouted three (3) of the principles of test administration practices. These were, “I ask students to work faster during the time of testing in order to finish on time” ($M = 3.55$, $SD = 1.28$), “ I give tests immediately before a long vacation, or other important events” ($M = 3.51$, $SD = 1.39$), and “ I tell students the dire consequences of failure in the test they are taking” ($M = 3.43$, $SD = 1.41$). A mean of means value of Mean = 3.92, $SD = 1.13$ however, shows that teachers generally adhere to the principles of test administration practices.

Test Scoring Practices of Teachers

In all, fourteen (14) items were used to measure this dimension of teachers' testing practices (test scoring) on a five-point Likert -type scale. Out of the 14 items, twelve (12) were positively stated, where, (Always-5, Very Often-4, Sometimes-3, Rarely-2, Never-1). The rest (2) of the items were negatively stated, where, (Always-1, Very Often-2, Sometimes-3, Rarely-4, Never-5). Mean and standard deviation were used to analyse data gathered on this construct (test scoring). A mid- point of 3.0 was used as cut off score, such that a mean value above 3.0 indicated that the teachers stated that they follow the principles of testing when scoring test items while a mean value less than 3.0 is an indication that the teachers are not adhering to the basic principles of testing when it comes to scoring of tests. Results on the test scoring practices of teachers are presented in Table 4.

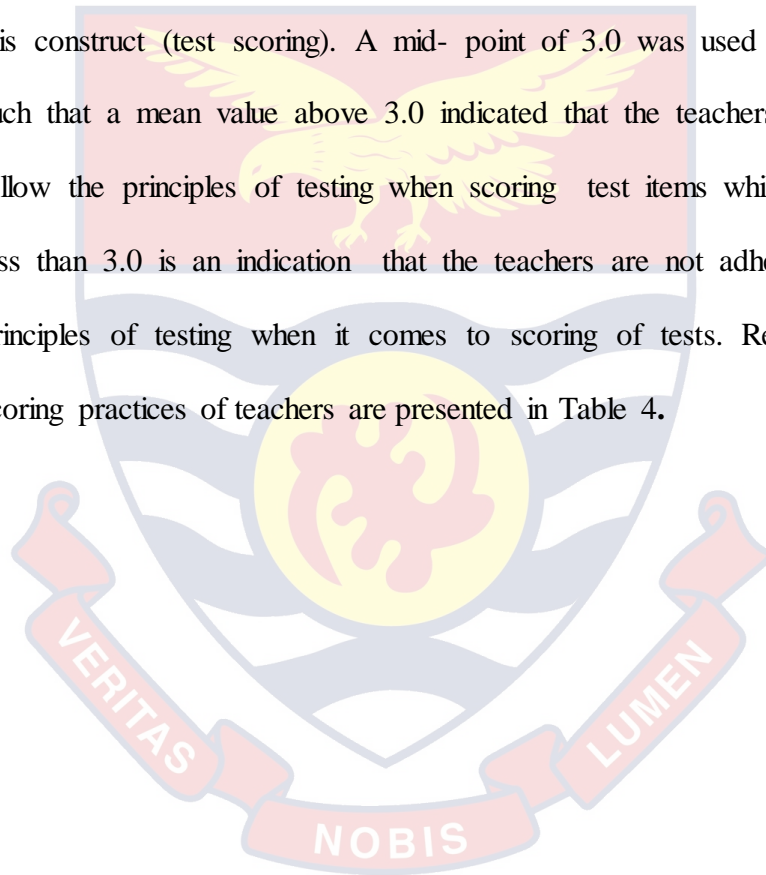


Table 4– *Test Scoring Practices of Teachers*

Statement	Mean	SD
I mark papers just after the test	4.19	.94
I prepare scoring guide and use it	4.58	.79
I make sure test takers are kept anonymous	3.19	1.32
I score the essay type responses, question by question	4.03	1.14
I keep scores of previous questions out of sight	3.28	1.29
I periodically rescore previously scored questions	2.56	1.26
I shuffle scripts before scoring	2.95	1.39
I score essay test when I am physically sound, in a sound environment	3.90	1.30
I score essay test when I am mentally alert, in a sound environment	3.75	1.45
I constantly follow scoring guide	4.43	.97
I am influenced by scores of the first few papers read when scoring test items	2.10	1.28
I score a particular question on all papers at a sitting, but break when fatigue sets in	3.12	1.36
I provide comments and correct errors on scripts	4.20	.96
I give extra marks to students based on handwriting; grammar etc. but not desired by the test.	2.42	1.28
Overall Mean/Mean of Means	3.48	1.20

Source: Field survey (2020)

The data in Table 4 show that five (5) of the items measuring the test scoring practices of teachers recorded mean values above 4.0. They ranged from Mean = 4.03, *SD* = 1.14 to Mean = 4.58, *SD* = 0.79, which is far above the cut-off mean of 3.0, implying teachers adhere to such principles when

scoring their test. Where the item, “I prepare scoring guide and use it” was put, a highest Mean = 4.58, $SD = 0.79$ was obtained. The item which recorded the next highest Mean = 4.43, $SD = 0.97$ was the item “I the constantly follow scoring guide”. It can also be seen from the data in Table 4 that five (5) other items also had high means ranging from Mean = 3.12, $SD = 1.36$ to Mean = 3.90 to $SD = 1.30$, also implying that teachers adhere to the principles of scoring test. E.g. the items, “I score essay test when I am physically sound in a sound environment” ($M = 3.90$, $SD = 1.30$) and “I keep scores of previous questions out of sight” ($M = 3.28$, $SD = 1.29$). Also, according to the data in Table 4, teachers were not influenced by scores of the first few papers read when scoring test items ($M = 2.10$, $SD = 1.28$) and they did not give extra marks to students based on handwriting, grammar etc but which are not desired by the test ($M = 2.42$, $SD = 1.28$). It can also be seen in Table 4 that teachers flouted two (2) of the principles of test administration practices. These were, “I periodically rescore previously scored questions” ($M = 2.56$, $SD = 1.26$) and “I shuffle scripts before scoring” ($M = 2.95$, $SD = 1.39$). A mean of means value of Mean = 3.48, $SD = 1.20$ however, shows that teachers generally adhere to the principles of test scoring practices.

Hypotheses Testing

The study tested five hypotheses. The criterion variable (testing practices) in the study was multi-dimensional. It comprised construction, administration and scoring of tests. Therefore, prior to testing the hypotheses, the normality assumption of the data was assessed for all the three dimensions. This assessment was done using several techniques. These techniques included:

The Shapiro-Wilk statistic, normal Q-Q plot, and the Histogram. Details of the Shapiro-Wilk statistics results are presented in Table 5.

Table 5- *Test for Normality*

Dimension	Shapiro-Wilk		
	Statistic	Df	Sig.
Construction	.978	107	.079
Administration	.980	107	.116
Scoring	.986	107	.322

Source: Field survey (2020)

The results in Table 5 showed that, the distribution of scores is normally distributed for all the three dimensions-construction, administration and scoring from the criterion variable. This is evident in the non-significant (values **sig.** values of more than .05) for Shapiro-Wilk statistics for the three dimensions, indicating a non violation of the assumption of normality.

Also, the normal Q-Q plots for all three dimensions of the variable were assessed. The details are displayed on Figure 2, Figure 3 and Figure 4 respectively.

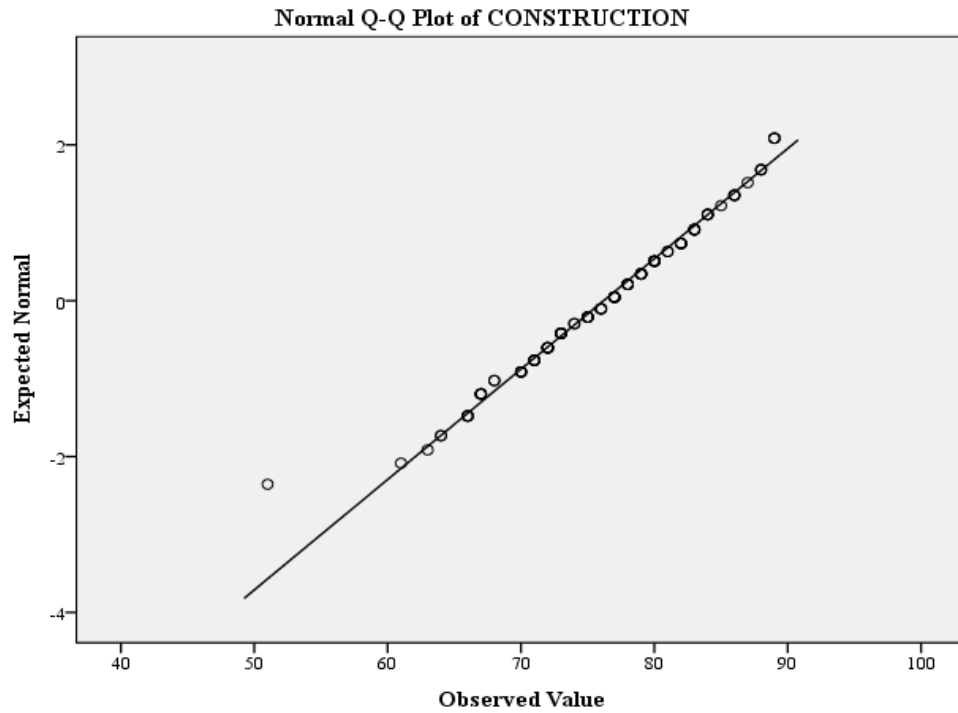


Figure 2- Test for Normality for Test Construction.

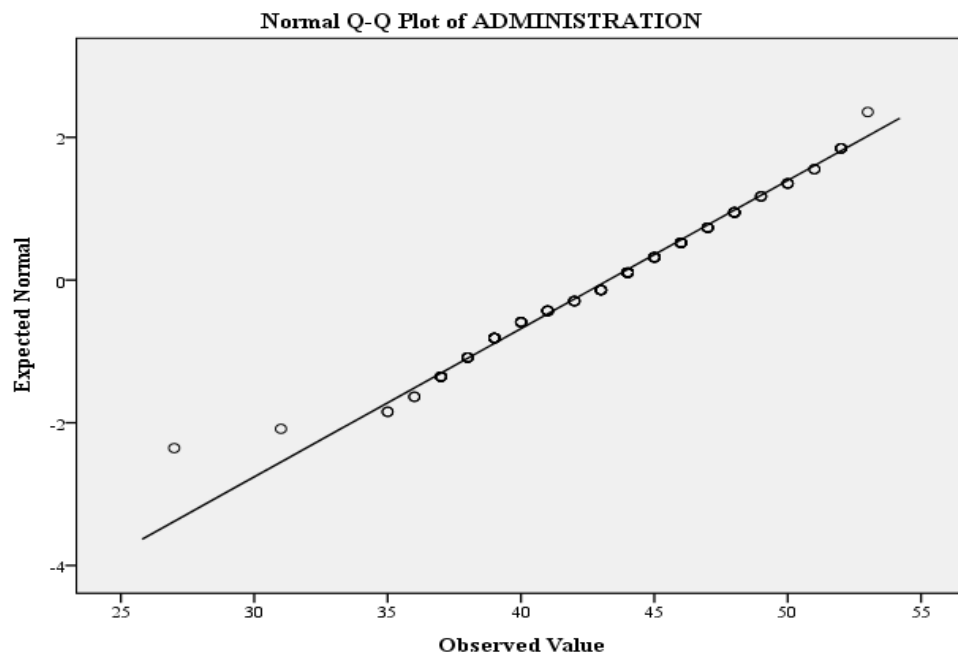


Figure 3- Test of Normality for Test Administration

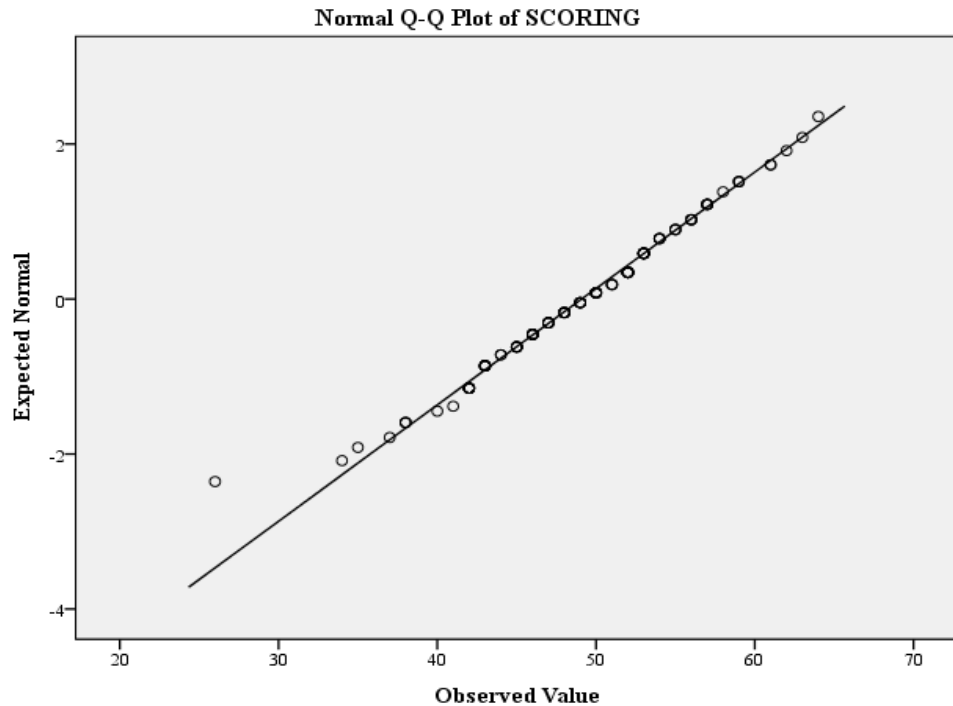


Figure 4- Test of Normality for Test Scoring

The outputs from Figure 2, 3, and 4 showed that the data was normally distributed. Based on Figure 2, 3, and 4, the normal Q-Q plots for all the dimensions of testing practices; construction, administration and scoring showed that the distribution of all the scores were close to the straight line.

In order to further ensure that the normality assumption of the data was not violated, I examined the output from the histograms of all the three dimensions of the construct. The results suggested further, that the normality assumption was not violated for any of the dimensions (see Appendix D). From Appendix D, the results showed that the data was reasonably normally distributed for each of the dimensions, with most of the scores occurring in the centre, narrowing out towards the extreme ends of the Histogram (see Appendix D).

Also, before the data on the five hypotheses were finally tested, the assumptions of multivariate regression analysis were checked. For, example,

the multivariate normality was checked and the results showed that the assumption was not violated. This was evident, because all the standard residuals were closely clustered at the centre (see Appendix E). Similarly, the multivariate linearity assumption was checked and the results again, showed that the assumption of multivariate linearity was met. The reason being that the residuals for both the predictor and the criterion variable were closely packed in a linear fashion (see Appendix E)

Hypothesis 1

H₀: Gender does not significantly predict teachers' adherence to principles of testing.

H₁: Gender significantly predicts teachers' adherence to principles of testing.

The main focus of this hypothesis was to test the relationship between female and male teachers in terms of their testing practices. The predictor variable (gender) was measured on nominal basis. Dummies were created and male was made the reference category. The criterion variable (testing practices), which had three dimensions; construction, administration and scoring, was measured on continuous basis.

Using an alpha level of .05, the results showed that the overall multivariate model after combining the criterion (testing practices) was not significant $F(3, 103) = .19, p = .90$, partial eta squared = .005 (see Appendix F).

Also, using the Bonferroni's alpha adjustment of .017, the model for univariate test results for the predictor (gender) and criterion (testing practices) showed that female teachers compared to male teachers in the models were not statistically significant. Thus, for construction $F(1, 105) = .43, p = .513$, partial

eta squared =.004; administration $F(1, 105) = .42, p = .520$, partial eta squared =.004; and scoring $F(1, 105) = .12, p = .731$, partial eta squared =.001 (see Appendix F).

The details of the regression coefficients for hypothesis 1 are illustrated in Table 6. The Table contains the unique contribution of female teachers compared to male teachers for the three dimensions of the criterion variable, using the Bonferroni's alpha adjustment of .017.

Table 6— *Parameter Estimates for Test Practices*

Dependent Variable	Parameter	B	Std. Error	T	Sig.	Partial Eta Squared
Construction	Intercept	76.39	.74	102.94	<.001	.990
	Female	-1.26	1.92	-.66	.513	.004
Administration	Intercept	43.41	.51	85.76	<.001	.986
	Female	-.84	1.31	-.65	.520	.004
Scoring	Intercept	49.19	.70	70.18	<.001	.979
	Female	-.62	1.81	-.34	.731	.001

Source: Field Survey (2020); *Significant, $p < .017$

The results in Table 6 shows that female teachers compared to male teachers was not a significant predictor of testing practices – construction, $B = -1.26, t = -.66, p = .515$. This result implies that female teachers compared to male teachers are less likely to adhere to the basic principles of test construction by 1.26 units. Also, female teachers compared to male teachers was not a significant predictor of testing practices- administration, $B = -.84, t = -.65, p = .520$. This means that in test administration, female teachers compared to male teachers are .84 units less likely to adhere or follow the basic laid down principles of test administration. Similarly, female teachers

compared to male teachers was not a significant predictor of testing practices-scoring, $B = -.62$, $t = .34$, $p = .73$. Like test construction and administration, this result means that female teachers compared to male teachers, are less likely to follow the basic principles of test scoring by .62 units.

Based on these results, it can be concluded that female teachers compared to male teachers is not a significant predictor of their testing practices, and therefore, I failed to reject the null hypothesis because there was no gender effect on the criterion variable (testing practices).

Hypothesis 2

H₀: Teachers' years of teaching does not significantly predict teachers' adherence to principles of testing.

H₁: Teachers' years of teaching significantly predicts adherence to principles of testing.

This hypothesis sought to examine the relationship between teachers' years of teaching and their testing practices. The predictor variable (years of teaching) was measured on nominal basis. The variable (years of teaching) was categorised into five. The categories were; those who have taught for 5 years or less, 6-10 years, 11- 15 years, 16-20 years, and 21 years or more. Since the variable was measured on nominal basis, dummies were created for it and those who have taught for 5 years or less was made the reference category. However, the criterion variable (testing practices), which had three dimensions; construction, administration and scoring was measured on continuous basis.

The results showed that the overall multivariate model after combining the criterion (testing practices) was not significant $F(3, 100) = 1.72, p = .17$, partial eta squared = .05 (see Appendix G).

Using the Bonferroni's alpha adjustment of .017, the model for univariate test results for the predictor (years of teaching) and criterion (testing practices) showed that all other categories of the predictor variable compared to the reference category (five years or less) in all the models were not statistically significant. That is to say that, for teachers who taught for between 6 and 10 years compared to those who taught for 5 years or less, was not statistically significant; construction $F(1, 102) = .26, p = .614$, partial eta squared = .003; administration $F(1, 102) = .002, p = .967$, partial eta squared = <.001 and scoring, $F(1, 102) = .042, p = .839$, partial eta squared = <.001. Also, teachers who taught 11 to 15 years compared to those who taught for 5 years or less; construction $F(1, 102) = 1.57, p = .213$, partial eta squared = <.001; administration $F(1, 102) = .96, p = .329$, partial eta squared = .009 and scoring, $F(1, 102) = .23, p = .636$, partial eta squared = .002. Further, between teachers who taught for 16-20 years and those who taught for 5 years or less; construction $F(1, 102) = 5.79, p = .018$, partial eta squared = .054; administration $F(1, 102) = .99, p = .323$, partial eta squared = .010 and scoring, $F(1, 102) = .55, p = .460$, partial eta squared = .005 and for teachers who have taught for 21 years or more, compared to those who taught for 5 years or less; construction $F(1, 102) = 3.67, p = .058$, partial eta squared = .035; administration $F(1, 102) = 2.56, p = .113$, partial eta squared = .024 and scoring, $F(1, 102) = .17, p = .679$, partial eta squared = .002 (see Appendix G).

The details of the regression coefficients for hypothesis 2 are illustrated in Table 7. The Table contains the unique contribution of teachers who taught for 6-10 years, 11- 15 years, 16-20 years, and 21 years or more, compared those who taught for 5 years or less for the three dimensions of the criterion variable, using the Bonferroni’s alpha adjustment of .017.

Table 7– *Parameter Estimates for Test Practices*

Dependent Variable	Parameter	B	Std. Error	T	Sig.	Partial Eta Squared
Construction	Intercept	75.40	1.10	68.29	<.001	.979
	Six	-.79	1.56	-.51	.614	.003
	Eleven	2.29	1.82	1.25	.213	.015
	Sixteen	8.61	3.58	2.41	.018	.054
	Twenty-one	6.21	3.24	1.92	.058	.035
Administration	Intercept	42.74	.78	54.75	<.001	.967
	Six	.05	1.10	.05	.962	.000
	Eleven	1.26	1.29	.98	.329	.009
	Sixteen	2.51	2.53	.99	.323	.010
	Twenty-one	3.66	2.29	1.60	.113	.024
Scoring	Intercept	48.87	1.10	44.63	<.001	.951
	Six	-.32	1.55	-.20	.839	.000
	Eleven	.86	1.81	.46	.636	.002
	Sixteen	2.63	3.55	.74	.460	.005
	Twenty-one	1.33	3.21	.42	.689	.002

Based on the results in Table 7, the four categories of the predictor variable (years of teaching), thus, teachers who taught for 6 -10 years, 11- 15 years, 16-20 years, and 21 years or more, compared those who taught for 5 years or less were not statistically significant for all the models. For instance, when it comes to test construction, the results showed that all the four categories were not statistically significant; for 6 to 10 years, $B = -.79$, $t = .51$, $p = .003$; for 11 to 15 years, $B = 2.29$, $t = 1.25$, $p = .015$; for 16 to 20 years, $B = 8.61$, $t = 2.41$, $p = .05$ and for those who taught for 20 years or more, $B = 6.21$, $t = 1.92$, $p = .035$. Also, test administration; those who taught for 6 to 10 years,

$B = .05, t = .05, p < .001$; for those who taught for 11 to 15 years, $B = 1.26, t = .98, p = .009$; for those who taught for 16 to 20 years, $B = 2.51, t = .99, p = .010$ and for those who taught for 20 years or more, $B = 3.66, t = 1.60, p = .024$. Similarly, for scoring, those who taught for 6 to 10 years, $B = -.32, t = -.20, p = .839$; for those who taught 11 to 15 years, $B = .86, t = .475, p = .636$; for those who taught for 16 to 20 years, $B = 2.63, t = .74, p = .460$ and for those who taught for 20 years or more, $B = 1.33, t = .42, p = .679$. Based on these results, it can be concluded that teachers' years of teaching, and thus, teachers who taught for 6-10 years, 11- 15 years, 16-20 years, and 21 years or more, compared to those who taught for 5 years or less is not a significant predictor of their testing practices, and therefore, I failed to reject the null hypothesis, which states that the predictor variable (teachers' years of teaching) does not significantly predict teachers' adherence to testing principles.

Hypothesis 3

H₀: Qualification of teachers does not significantly predict teachers' adherence to principles of testing.

H₁: Qualification of teachers significantly predicts teachers' adherence to principles of testing.

This hypothesis sought to examine the relationship between teachers' academic or professional qualification and their testing practices. The predictor variable (teacher qualification) was measured on nominal basis. The variable (teacher qualification) was categorised into six. These categories included those who have Teachers' Certificate A, Diploma with Education, Bachelors with Education, Bachelors without Education, Masters with

Education, and Masters without Education. Since the variable was measured on nominal basis, dummies were created for it. Teachers with Diploma in Education were made the reference category. However, the criterion variable (testing practices), which had three dimensions; construction, administration and scoring was measured on continuous basis.

Using the alpha level of .05, the results showed that the overall multivariate model after combining the criterion (testing practices) was not significant $F(3, 100) = .64, p = .589$, partial eta squared = .019 (see Appendix H).

Also, an evaluation of the model for univariate test, using the Bonferroni's alpha adjustment of 0.017 was done. The results for the predictor (teacher qualification) and criterion (testing practices) showed that all other categories of the predictor variable compared to the reference category (Diploma with Education) in all the models were not statistically significant. That is to say that, for teachers who have only a Teachers' Certificate A, compared to those Diploma with Education, was not statistically significant; construction $F(1, 102) = 2.24, p = .14$, partial eta squared = .021; administration $F(1, 102) = .13, p = .175$, partial eta squared = .001 and scoring, $F(1, 102) = .19, p = .663$, partial eta squared = .002. Also, teachers who have Bachelors with Education compared to Diploma with Education; construction $F(1, 102) = .54, p = .463$, partial eta squared = .005; administration $F(1, 102) < .001, p = .973$, partial eta squared = .001 and scoring, $F(1, 102) = .16, p = .688$, partial eta squared = .002. Further, between teachers who have Bachelors without Education compared those with Diploma with Education; construction

$F(1, 102) = 2.810, p = .097$, partial eta squared = .027; administration $F(1, 102) = 1.56, p = .21$, partial eta squared = .015 and scoring, $F(1, 102) = 3.593, p = .061$, partial eta squared = .034 and for teachers who have Masters with Education, compared to those who have Diploma with Education; construction $F(1, 102) = .96, p = .329$, partial eta squared = .009; administration $F(1, 102) = 1.84, p = .177$, partial eta squared = .018 and scoring, $F(1, 102) = .73, p = .395$, partial eta squared = .007 (see Appendix H).

The details of the regression coefficients for hypothesis 3 are illustrated in Table 8. The Table contains the unique contribution of teachers' qualification. Thus, teachers who have Teachers' Certificate A, Bachelors with Education, Bachelors without Education, Masters with Education, and Masters without Education, compared to those who have Diploma with Education for the three dimensions of the criterion variable, using the Bonferroni's alpha adjustment of .017.

Based on the results in Table 8, the five categories of the predictor variable (teacher qualification), thus, teachers who have Teachers' Certificate A, Bachelors with Education, Bachelors without Education, Masters with Education, and Masters without Education, compared those who Diploma with Education were not statistically significant for all the models.

Table 8– *Parameter Estimates for Test Practices*

Dependent Variable	Parameter	B	Std. Error	t	Sig.	Partial Eta Squared
Construction	Intercept	75.44	.99	76.36	<.001	.983
	Certificate	10.56	7.06	1.50	.138	.021
	Bachelors education	1.04	1.41	.74	.463	.005
	Bachelors without	5.06	3.02	1.68	.097	.027
	Masters Education	-4.94	5.04	-.98	.329	.009
	Masters Without	0
Administration	Intercept	43.22	.68	63.41	<.001	.975
	Certificate	1.78	4.87	.366	.715	.001
	Bachelors education	-.032	.97	-.033	.973	.000
	Bachelors without	2.61	2.08	1.26	.212	.015
	Masters Education	-4.72	3.48	-1.36	.177	.018
	Masters Without	0
Scoring	Intercept	49.08	.936	52.44	<.001	.964
	Certificate	2.92	6.68	.44	.663	.002
	Bachelors education	-.54	1.34	-.403	.688	.002
	Bachelors without	5.42	2.86	1.90	.061	.034
	Masters Education	-4.08	4.77	-.86	.395	.007
	Masters Without	0

Source: Field Survey (2020); *Significant, $p < .017$

For instance, when it comes to the testing practices – Construction, the results in Table 8 shows that all the four categories were not statistically significant; for Teachers’ Certificate A, $B = 10.56$, $t = 1.50$, $p = .138$; for Bachelors with Education, $B = 1.04$, $t = .74$, $p = .463$; for Bachelors without Education, $B = 5.06$, $t = 1.68$, $p = .097$ and for those with Masters with Education, $B = -4.94$, $t = -.98$, $p = .329$. Also, for test Administration; for Teachers’ Certificate A, $B = 1.78$, $t = .34$, $p = .72$; for Bachelors with Education, $B = -.03$, $t = -.033$, $p = .973$; for Bachelors without Education, $B = 2.61$, $t = 1.26$, $p = .21$ and for those with Masters with Education, $B = -4.72$, $t = -1.36$, $p = .177$. Similarly, for Scoring, those with Teachers’ Certificate A, $B = 2.92$, $t = .44$, $p = .663$; for Bachelors with Education, $B = -.54$, $t = -.40$, $p =$

.688; for Bachelors without Education, $B = 5.42$, $t = 1.90$, $p = .061$ and for those with Masters with Education, $B = -4.08$, $t = -.86$, $p = .395$.

Based on these results, it can be concluded that teachers' qualification, and thus, teachers who have Teachers' Certificate A, Bachelors with Education, Bachelors without Education, Masters with Education, and Masters without Education, compared those who Diploma with Education is not a significant predictor of their testing practices, and therefore, I failed to reject the null hypothesis, which states that the predictor variable (teacher qualification) does not significantly predict teachers' adherence to testing principles.

Hypothesis 4

H₀: Subject area of teachers does not significantly predict teachers' adherence to principles of testing.

H₁: Subject teachers teach significantly predicts teachers' adherence to principles of testing.

This hypothesis sought to investigate the relationship between teachers' subject area and their testing practices. The predictor variable (teacher qualification) was measured on nominal basis. The variable (subject teachers teach) was categorised into four. These categories included: Teachers who teach the four core subject areas; English Language, Mathematics, Integrated Science and Social Studies. Since the predictor variable was measured on nominal basis, dummies were created for Mathematics was made the reference category. However, the criterion variable (testing practices), which had three dimensions; construction, administration and scoring was measured on continuous basis.

The overall multivariate model after combining the criterion (testing practices) was analysed, using the alpha level of .05. The results showed that the overall multivariate model was not significant $F(3, 101) = .26, p = .852$, partial eta squared = .008 (see Appendix I).

Also, the model for univariate test, using the Bonferroni's alpha adjustment of 0.017 was analysed. The results for the predictor (Subject teachers teach) and criterion (testing practices) showed that all other categories of the predictor variable compared to the reference category (Mathematics) in all the models were not statistically significant. For English Language, compared to Mathematics; construction $F(1, 103) = 5.22, p = .024$, partial eta squared = .048; administration $F(1, 103) = .11, p = .738$, partial eta squared = .001 and scoring, $F(1, 103) = .61, p = .436$, partial eta squared = .006.

The details of the regression coefficients for hypothesis 4 are illustrated in Table 9. The Table contains the unique contributions of Teachers who teach the four core subject areas; English Language, Integrated Science and Social Studies, compared to those who teach Mathematics, for the three dimensions of the criterion variable, using the Bonferroni's alpha adjustment of .017.

Table 9- *Parameter Estimates for Testing Practices*

Dependent Variable	Parameter	B	Std. Error	t	Sig.	Partial Eta Squared
Construction	Intercept	78.00	1.26	61.93	<.001	.98
	English	-4.10	1.80	-2.28	.024	.05
	Science	-3.38	1.96	-1.72	.088	.03
	Social	-.11	1.80	-.06	.952	.00
Administration	Intercept	43.63	.89	48.10	<.001	.96
	English	-.43	1.27	-.34	.738	.001
	Science	-.73	1.39	-.53	.601	.003
	Social	-.37	1.29	-.29	.773	.001
Scoring	Intercept	48.63	1.22	39.73	<.001	.94
	English	1.37	1.75	.78	.436	.01
	Science	-.78	1.91	-.41	.685	.002
	Social	.96	1.78	.54	.591	.003

Source: Field Survey (2020); *Significant, $p < .017$

The results in Table 9 shows that the three categories of the predictor variable (Subject teachers teach), thus, teachers who teach the three core subject areas; English Language, Integrated Science and Social Studies, compared to those who teach Mathematics were not statistically significant for all the models. For instance, when it comes to the testing practices – Construction, the results show that all the four categories were not statistically significant; English Language, $B = -4.10$, $t = -2.28$, $p = .024$; Integrated Science, $B = -3.38$, $t = -1.72$, $p = .088$; and for those who teach Social Studies, $B = -.11$, $t = -.06$, $p = .952$. For test administration, English Language, $B = -.43$, $t = -.34$, $p = .738$; Integrated Science, $B = -.73$, $t = -.53$, $p = .601$; and for those who teach Social Studies, $B = -.37$, $t = -.28$, $p = .773$ and test scoring, English Language, $B = 1.37$, $t = .78$, $p = .436$; Integrated Science, B

= -.78, $t = -.41$, $p = .685$; and for those who teach Social Studies, $B = .96$, $t = .54$, $p = .591$.

Based on the results in Table 9, it can be concluded that predictor variable (teachers' subject area), and thus, teachers who English Language, Integrated Science and Social Studies, compared those who teach Mathematics is not a significant predictor of their testing practices, and therefore, I failed to reject the null hypothesis, which states that the predictor variable (Subject teachers teach) does not significantly predict teachers' adherence to testing principles.

Hypothesis 5

H₀: Gender, years of teaching, teacher qualification and subject teachers teach combined together do not combined together predict teachers' adherence to principles of testing.

H₁: Gender, years of teaching, teacher qualification and subject teachers teach combined together predict teachers' adherence to principles of testing.

This hypothesis sought to examine the joint contribution of the four categories of the predictor variable; gender, years of teaching, teacher qualification and subject teachers teach to their classroom testing practices. The predictor variables were measured on nominal basis, and therefore, dummies were created for all the four categories of the predictor variable. The criterion variable (testing practices), which was multidimensional; test construction, administration and scoring was measured on continuous basis. The overall model was tested using the Hierarchical Multiple Regression for each of the three dimensions of the criterion variable (testing practices).

Before the analysis, the data on each of the three dimensions of the predictor variable was checked for autocorrelation and multicollinearity. The results from the Durbin-Watson test showed that there was no autocorrelation for each of the dimensions; construction ($d = 2.2$), administration ($d = 2.33$) and scoring ($d = 2.03$). Also, the VIF test showed that there was no multicollinearity for each of the three dimensions of the predictor variable.

Table 10 provides results of the hierarchical regressions for test construction.

Table 10– *Hierarchical Regression of Predictor variable and Testing Practices (Construction)*

Model		Sum of Squares	Df	Mean Squares	F	Sig.	R ²	R ² Change
1	Regression	6.17	1	6.17	.12	.735	.03	.001
	Residual	6710.57	125	53.69				
2	Regression	313.96	5	62.79	1.19	.320	.22	.05
	Residual	6402.80	121	52.92				
3	Regression	463.92	9	51.55	.97	.473	.26	.07
	Residual	6252.84	117	53.44				
4	Regression	953.43	12	79.45	1.57	.110	.38	.14
	Residual	5763.33	114	50.56				

Dependent Variable: Test construction

1. Predictor: Gender
2. Gender, Years of teaching
3. Gender, Years of teaching, Qualification
4. Gender, Years of teaching, Qualification, Subject Area

The results in Table 10 shows that the overall hierarchical regression model for test construction was not statistically significant $F(12, 114) = 1.57, p = .11$. This implies that of all the four categories (demographic characteristics of teachers) - gender, years of teaching, qualification and subject area of the predictor variable, jointly, did not contribute significantly, to the first dimension- test construction, of the criterion variable (testing

practices). However, from the results in Table 10 shows that, jointly, the four categories of the predictor variable contributed 38% of variance in the criterion variable, in this case, teachers' adherence to the basic principles of test construction.

The results in Table 11 shows that apart from English Language and integrated science, none of the predictor variables was found to be a predictor of the test construction practices of teachers.

Table 11– *Regression Coefficients for Predictor Variables (construction)*

Model 4	Unstandardized Coefficients		Standardized Coefficients	T	Sig.	VIF
	B	Std. Error	Beta			
(Constant)	76.99	1.48		52.17	.000	
Female	-1.58	1.97	-.08	-.80	.425	1.19
6-10 years	.49	1.78	.03	.28	.781	1.81
11-15 years	2.94	2.37	.11	1.24	.218	2.23
16-20 years	5.34	3.50	.16	1.52	.130	1.39
21 years/more	2.31	3.48	.07	.66	.509	1.37
Cert A	9.72	7.59	.12	1.28	.203	1.13
Bachelors with education	.279	1.80	.02	.16	.877	2.024
Bachelors without Education	3.00	3.15	.09	.95	.343	1.30
Masters with Education	-5.16	5.29	-.08	-.98	.331	1.09
English Language	-3.65	1.82	-.22	-2.01	.047	1.56
Integrated Science	-5.25	1.86	-.3	-2.83	.006	1.56
Social Studies	-1.46	1.81	-.09	-.81	.422	1.52

* Significant at .05 level

The results in Table 12 shows that the overall hierarchical regression model for test administration was not was not statistically significant $F(13,$

122) = .63, $p = .822$. This implies that of all the four categories (demographic characteristics of teachers) - gender, years of teaching, qualification and subject teachers teach of the predictor variable, jointly, did not contribute to the first dimension- test administration, of the criterion variable (testing practices). However, the results in Table 16 showed that, jointly, the four categories of the predictor variable contributed 6% of the variance in the criterion variable, in this case, teachers' adherence to the basic principles of test administration.

Table 12– *Hierarchical Regression of Predictor Variables and Testing Practices (Administration)*

Model		Sum of Squares	Df	Mean Squares	F	Sig.	R ²	R ² Change
1	Regression	32.22	1	32.22	.119	.277	.01	.01
	Residual	3624.42	134	27.05				
2	Regression	120.66	5	24.13	.89	.492	.03	.02
	Residual	3535.98	130	27.20				
3	Regression	227.39	10	22.74	.83	.602	.06	.03
	Residual	3429.26	125	27.43				
4	Regression	231.05	13	17.77	.63	.822	.06	.001
	Residual	3425.60	122	28.08				

1. Predictor: Gender
2. Gender, Years of teaching
3. Gender, Years of teaching, Qualification
4. Gender, Years of teaching, Qualification, Subject Area

The results in Table 13, shows that none of the predictor variable was found to be a significant predictor of test administration.

Table 13 – *Regression Coefficients for Predictor Variable (administration)*

Model 4	Unstandardized Coefficients		Standardized Coefficients		Sig	VIF
	B	Std. Error	Beta	T		
(Constant)	42.66	1.06		40.17	<.000	
Female	-1.40	1.40	-.10	-.99	.322	1.25
6-10 years	1.74	1.28	.16	1.36	.178	1.82
11-15 years	1.68	1.72	.13	.98	.330	2.27
16-20 years	.67	2.59	.03	.26	.796	1.37
21 years/more	3.76	2.77	.14	1.36	.177	1.32
Cert A	.53	5.63	.01	.094	.925	1.12
Bachelors with education	-1.14	1.31	-.11	-.88	.383	2.06
Bachelors without Education	1.34	2.40	.06	.56	.577	1.36
Masters with Education	-5.96	3.94	-.14	-1.52	.132	1.09
Masters Without Education	-3.13	5.58	-.05	-.56	.576	1.10
English Language	.136	1.30	.01	.11	.917	1.57
Integrated Science	-.36	1.33	-.03	-.27	.788	1.532
Social Studies	-.053	1.30	-.004	-.04	.968	1.470

The results in Table 14 shows that the overall hierarchical regression model for test scoring was not statistically significant $F(13, 110) = .89, p = .893$. This implies that of all the four categories (demographic characteristics of teachers) - gender, years of teaching, qualification and subject area of the predictor variable, jointly, did not contribute to the first dimension- test scoring, of the criterion variable (testing practices). However, the results in Table 15 shows that, jointly, the four categories of the predictor variable contributed 10% of the variance in the criterion variable, in this case, teachers' adherence to the basic principles of scoring test.

Table 14– *Hierarchical Regression of Predictor Variables and Testing Practices (Scoring)*

Model 4	Sum of Squares	Df	Mean Squares	F	Sig.	R ²	R ² Change
1 Regression	23.93	1	29.93	.48	.489	.004	.001
Residual	6070.42	122	49.76				
2 Regression	75.65	5	15.13	.30	.914	.01	.01
Residual	6018.70	118	51.01				
3 Regression	438.15	10	43.82	.88	.558	.07	.06
Residual	5656.19	113	50.06				
4 Regression	581.49	13	44.73	.89	.893	.10	.02
Residual	5512.86	110	50.12				

1. Predictor: Gender
2. Gender, Years of teaching
3. Gender, Years of teaching, Qualification
4. Gender, Years of teaching, Qualification, Subject Area

The results in Table 15, shows that none of the predictor variables was found to be a significant predictor of test scoring.

Table 15- *Regression Coefficients for Predictor Variable (scoring)*

Model 4	Unstandardized Coefficients		Standardized Coefficients		Sig	VIF
	B	Std. Error	Beta	t		
(Constant)	48.23	1.48		32.60	.000	
Female	-2.08	1.97	-.101	-1.05	.294	1.25
6-10 years	1.13	1.81	.08	.62	.536	1.87
11-15 years	1.10	2.43	.11	.82	.412	2.27
16-20 years	5.45	4.11	.14	1.33	.187	1.30
21 years/more	.72	3.71	.02	.19	.848	1.32
Cert A	-.65	7.58	-.01	-.09	.931	1.14
Bachelors with education	-2.43	1.87	-.17	-1.30	.196	2.09
Bachelors without Education	4.02	3.04	.14	1.32	.188	1.38
Masters with Education	-5.57	5.28	-.10	-1.05	.294	1.10
Masters Without Education	-9.70	7.49	-.12	-1.30	.198	1.11
English Language	2.43	1.85	.15	1.31	.192	1.59
Integrated Science	-.37	1.84	-.02	-.20	.839	1.51
Social Studies	1.72	1.80	.11	.96	.341	1.47

The results of the hierarchical multiple regression showed that there was no statistically significant contribution of the predictor variable (teacher demographics) – gender, years of teaching, qualification and subject teachers teach to the variance of all the three dimensions of the the criterion variable (testing practices). Thus, test construction, administration and scoring.

Discussions

Clearly, the success of any educational institution without proper assessment will be in doubt. It is worth noting that, the role of the teacher and characteristics related to him/her is quintessential to classroom assessment. That is to say that, teachers' demographic characteristics; gender of teachers, years of teaching, qualification of teachers and the subject teachers teach, have an influence on the way they may perceive and practice assessment. In the context of this study, the contribution of teachers' demographic characteristics to their classroom testing practices was explored. This section discusses the findings of this study in relation to teachers' demographic characteristics as predictors of testing practices.

Gender of Teacher and Testing Practices

The gender of an individual generally, is believed to have some influence on one's behaviour. This hypothesis sought to test the contribution of gender towards the testing practices of teachers in the Sissala East Municipality. That is to say, the hypothesis sought to examine whether being a male or female teacher would the testing practices of teachers.

The findings of the study from the model, the Bonferroni's alpha adjustment, revealed that female teachers compared to male teachers in all the models were not significant. This suggests that both female and male teachers

in the municipality practice, classroom assessment (testing) independent of the influence of their gender.

The finding from this study may raise a number of questions. For instance, could be that both male and female teachers have similar levels of knowledge, motivation and commitment when it comes to the practice of classroom assessment? While trying to figure out a possible answer to the question, it is important to note however, that the finding from this current study contradicts findings of some previous researchers (Asamoah, et al, 2011; & Mohuiddin, 2015) whose studies observed a significant difference in educational assessment knowledge between male and female teachers. However, it is important to say that even though the finding of this current study does not confirm the findings of the previous studies, I wish to put my evaluation of the findings in context. That is to say that while this current study explored teachers' practices of classroom testing, the previous studies focused on teachers' knowledge about classroom assessment. But it is clear that though, both constructs; teachers' knowledge and teachers' practice relate to classroom assessment, the two constructs vary significantly when it comes to their measurement.

The finding from the current study nonetheless, supports that of Opara (2018), who investigated teachers' characteristics and how they could determine teachers' attitude towards the practice of continuous assessment in Ohio-Akpor Local Government Area of River State in Nigeria, and found that gender not to be a good predictor of teachers' attitude towards the practice of continuous assessment.

Years of Teaching and Testing Practices

The hypothesis sought to examine the contribution of teachers' years of teaching to their assessment practices. The finding from the study revealed that years of teaching was not a significant predictor of testing practices. This implies that the adage that "experience is the best teacher", does not find expression among teachers in the Sissala East Municipality, as far as their testing practices is concerned. It is believed that the more anyone practices a skill, the more he/she becomes more knowledgeable and competent in that skill, and therefore, he/she is likely to bring the experience to bare, on a related job. The use of test in the teaching and learning process is very prevalent. Therefore, it was expected that the more teachers use tests for classroom assessment over a period of time (years), they more they should be able to master and follow the basic principles of developing tests that are valid for classroom assessment.

The study however, found that for the four categories of the predictor variable (years of teaching), thus, teachers who taught for 6-10 years, 11- 15 years, 16-20 years, and 21 years or more, compared those who taught for 5 years or less, jointly explains 4.9% of the variance of the combined testing practices. This simply means that both the more experienced and the less experienced do not adhere to the basic principles of testing differently. In my view, and based on my personal experience, the finding of the study is not surprising. The reason being that, the use of past questions and commercialised tests in our educational institutions has become a convention, specially, at the Junior High Schools. Therefore, heads and other stakeholders in education such as the Sissala East Education, are encouraged to intensify

their effort in organising workshop activities on effective ways of planning classroom assessment, particularly, in line with the recommended guidelines for test constructions, administration and scoring.

Susuwele-Banda (2005) investigated the teachers' perception of classroom assessment in mathematics and their classroom assessment practices in Malawi. The study adopted a mixed approach thereby using questionnaires, observations, interviews and document analysis for the data collection. The findings showed that teachers' years of teaching experience did not seem to contribute to teachers' perception of classroom assessment. The finding of the current study corroborates the findings of Susuwele-Banda. The finding of the current study also, affirms the finding of Anhwere (2009), who, investigated whether teachers in the teacher training colleges of Ghana follow the basic laid down principles of testing in three specific stages of planning a classroom assessment; test construction, administration and scoring of teacher-made tests and found that there was not any statistically significant difference in the construction and administration of teacher-made tests based of years of teaching experience. Even though findings of these studies are similar, context should be considered in evaluating these findings. The study of Susuwele-Banda was done in Malawi, on teachers' perception of classroom assessment, while, the current study and that of Anhwere, were carried out in Ghana.

In contrast with the finding of this current study, Mohuiddin (2015) found difference in teachers' assessment practices was attributable to years of teaching experience. The contradiction here could result from the use of teachers from two different education levels. While, the current study targeted

JHS teachers, that of Mohuiddin, focused on secondary school teachers. Also, the the finding of this study is inconsistent with Koloji-Keaitkitse (2012), who explored the discrepancies between teachers' perceived skills and use of classroom assessment practices of primary and secondary schools. A descriptive survey design was adopted for the study a questionnaire known as the Classroom Assessment Practices and Skills (CAPS) questionnaire. The study found a significant difference in teachers' perceived skills and use of classroom assessment, on the basis of years of teaching experience. The also, results showed that teachers with only a certificate in teaching needed more training in assessment applications, statistical applications and criterion referenced testing, while the more experienced teachers reported higher on average on perceived skills and use of classroom assessment practices.

Qualification of Teacher and Testing Practices

There is a saying that "to whom much is given, much is expected". Generally, higher education and training comes with higher qualification. The essence of higher education is to acquire new skills and competencies, or to improve on existing skills and competencies, which are relevant to one's career or job. The acquisition of skills in Classroom assessment is one of the core targets of all educational institutions, the world all over. Teachers get more grips of the broader scope and complexities of these skills through higher education. When it comes to classroom assessment therefore, teachers with higher qualifications are expected to plan classroom assessment, more effectively than those with lower academic qualification. This study tested the hypothesis whether teachers' academic qualification would predict their adherence to basic principles in testing.

The finding showed that the overall regression model after combining the criterion (testing practices) was not significant. This implies that for the four categories of the predictor variable (teacher qualification), thus, teachers who have Teachers' Certificate A, Bachelors with Education, Bachelors without Education, Masters with Education, and Masters without Education, compared those who Diploma with Education, jointly explains 1.9% of the variance of the combined testing practices.

The finding of the current study is inconsistent with the finding of Opara (2018), in Nigeria, who investigated teachers' characteristics and how they could determine teachers' attitude towards the practice of continuous assessment in Ohio-Akpor Local Government Area of River States. The results showed that educational qualification was found to be a good determinant of teachers' attitude towards continuous assessment practices.

The finding of Koli-Keaitkitse (2012), is also at variance with the finding of this study. Keaitkitse explored the discrepancies between teachers' perceived skills and use of classroom assessment practices of primary and secondary schools. The results showed that teachers with only a certificate in teaching needed more training in assessment applications, statistical applications and criterion referenced testing. The inconsistencies in these studies could be as a result of the differences in the context of the studies and the differences in focus of the researchers. While this study focused on teachers' adherence to the basic principles in test construction, administration and scoring, Keaitkitse's focus was on the discrepancies between teachers' perceived skills and use of classroom assessment practices.

The finding of this study however, affirms the finding of Gemedā and Getachew (2019), who studied instructors' continuous assessment perception in the Mada Walabu University in South East Ethiopia. With regard to educational qualification, the results found that first degree holders reported lower perceptions of continuous assessment, compared to PhD and Master degree holders. Here, again, it is important to indicate that context should be applied in evaluating the finding of the current study, relative to the finding by Gemedā and Getachew. For instance, the current study was done in Ghana, and focused on teachers' adherence to basic principles of testing, while that of Gemedā and Getachew was done in Ethiopia, and they focused on teachers' perception of continuous assessment. Therefore, the measurement of these two constructs, are very distinct.

Subject Teachers Teach and Testing Practices

In evaluating learning of pupils, it is common knowledge that the appropriateness of the assessment format that the teacher uses, depends in part, on the nature of the concept taught and the objectives that the test is meant to evaluate. This is to say that different concepts are taught in different subject areas at all levels of the educational system. Different subject teachers therefore, may apply principles of testing differently. This hypothesis sought to test whether the subject area of teachers would predict their adherence to basic testing principles in the classroom.

The results showed that the overall multivariate model was not significant. This means subject teachers in the municipality follow the basic principles of testing similarly. The finding showed that for the three categories of the predictor variable (Subject teachers teach), thus, English Language,

Integrated Science and Social Studies, compared to Mathematics, jointly explains 0.8% of the variance of the combined testing practices.

The finding of the current study corroborates the finding of Quansah, Amoako and Ankomah (2019) and Mohuiddin (2015). Quansah, Amoako and Ankomah, examined the test construction skills of senior high school teachers in the cape coast metropolis. Using qualitative document analysis, the researchers examined samples of end - of -term examination papers of in three subject areas; Integrated science, core mathematics and social studies in three randomly selected Senior High School. The results showed that teachers in all the three subject areas have limited skills in the construction of end of term examination. Similarly, Mohuiddin explored attitudes, competence, knowledge and practices of teachers about educational assessment practices in Bangladesh and the results showed that Mathematics teachers reported on average a higher level of educational assessment knowledge than English language teachers. It is important to say that, though the finding of the current study is similar to the previous ones, it is entreated that the context of these studies should guide the evaluation of the findings. For instance, though both the current study and that of Quansah, Amoako and Ankomah were conducted in Ghana, the current study was done in the Sissala East municipality in the Upper West Region while that of Quansah, Amoako and Ankomah was done in the Cape Coast Metropolis. These two regions may vary in terms of their internal policies.

Gender, years of teaching, qualification, and subject teacher teach and testing practices

Two heads they say are better than one. The contribution of each of the teachers' demographic characteristics to their testing practices was explored. The study also sought to examine the joint contribution of all the demographic characteristics to the testing practices of teachers. Therefore, the hypothesis was stated as: gender, years of teaching, teacher qualification and subject area would not jointly predict teachers' adherence to testing principles.

The analysis of the hierarchical multiple regressions showed that there was no statistically significant contribution of the predictor variable (teacher demographics) – gender, years of teaching, qualification and subject area to the variance in all the three dimensions of the criterion variable (testing practices). Thus, test construction, administration and scoring. The findings showed that jointly, 38% of the variance in test construction was explained by teachers' demographic characteristics. The model also explained 6% of the variance in test administration and 10% of the variance in test scoring. This implies that even though the contribution of the overall model was not significant, they model explains some reasonable percentage of the variance in the criterion variable, especially, when it comes to test construction and scoring. The result of the study simply means that, the role of teachers' demographic characteristics cannot be ignored completely, as far as classroom assessment practices are concerned.

The finding of the current study corroborates the finding of Owusu-Mensah (2019), who explored the relationship between teacher characteristics and their classroom assessment practices in the Atwima-Nwabiagya south

district in the Ashanti Region. The descriptive survey, using a random sample of 219 teachers, was used for the study. The study found that teachers' characteristics had no significant effect on their assessment practices. It is worthy to note that, though the finding of the current study and that of Owusu-Mensah are similar, the context of studies should be considered when evaluating the findings. Thus, though both the current study and that of Owusu-Mensah were conducted in Ghana, the current study was done in the Sissala East municipality in the Upper West Region while that Owusu-Mensah was done in the Ashanti Region. These two regions may vary in terms of their internal policies and the experiences of teachers.

Chapter Summary

It was found that gender of teachers in the Sissala East Municipality did not predict their adherence to testing practices. It was found that teachers' years of teaching did not predict their testing practices. It was reported that teachers' academic qualification did not contribute significantly to their testing practices. The study further found that subject area of teachers was not a significant predictor of testing practices. It was found that gender, years of teaching, academic qualification, and subject area of teachers, jointly, did not predict teachers' adherence to testing principles.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This chapter gives a summary of the study, the conclusions, suggestions and recommendations for further studies based on the findings.

Summary

Overview of the study

The main purpose of the study was to examine teachers' demographic characteristics as predictors of testing practices in the Sissala East Municipality. Specifically, the study sought to examine whether: (a) gender would predict how teachers adhere to the principles of testing, (b) years of teaching would predict how teachers adhere to the principles of testing, (c) teacher qualification would predict how teachers adhere to the principles of testing, (d) subject area would predict how teachers adhere to the principles of testing, and (e) gender, years of teaching, teacher qualification and subject area would jointly predict how teachers adhere to the principles of testing; construction, administration and scoring. The study was driven by these five objectives, which were translated into five research hypotheses.

The study employed the descriptive survey design (cross-sectional) in carrying out the study. The study targeted Junior High School teachers in the Sissala East Municipality. The study however, specifically, targeted teachers teaching the four core subjects; Mathematics. English Language; Integrated Science and Social Studies. Through a multi-stage sampling procedure, a

three-dimension adapted questionnaire, which comprised of 44 items was administered to 220 teachers.

The data collected were analysed using descriptive statistics-frequencies, percentages, means and standard deviations; and inferential statistics-multivariate multiple regression and hierarchical multiple regression.

Key Findings

The following were the findings from the study.

1. Gender of teacher did not predict teachers' testing practices.
2. Qualification of teacher did not predict teachers' testing practices.
3. It was found that years of teaching did not predict teachers' testing practices
4. Subject area of teachers did not predict their testing practices
5. The study found that gender, years of teaching, teacher qualification and subject area of teacher, jointly, did not predict teachers' testing practices.

Conclusions

Based on the key findings, it can be concluded that the classroom assessment practices of teachers in the Sissala East Municipality, is not influenced by their personal characteristics, such as the gender of the teacher and the qualification of the teacher. That is to say that whether teachers in the Municipality follow the basic principles of testing-construction, administration, and scoring or not, it is not as a result of their gender or qualification. This may be possible due to the fact that teachers in the municipality had similar academic and professional training as part of their initial training in the similar various educational institutions.

It can also be concluded that the adage, “experience is the best teacher”, does not apply to teachers in the Sissala East Municipality. That is to say, both the more experienced and less experienced teachers reported that they adhered to the basic principles of teaching similarly. In view my opinion, the possible reason for this could be that, both the less experienced and the more experience teachers depend on past questions and commercial test papers when conducting classroom assessment.

Again, it can be concluded that teachers in the Sissala East Municipality, irrespective of the subject they teach, carry out classroom assessment practices similarly. The implication is that subject teachers in the Sissala East Municipality are not different when it comes to how they carry out classroom assessment, particularly, in the area of test construction, administration and scoring.

Recommendations

Following the findings of the study, and the conclusions drawn, the following recommendations were made:

1. Prior to the hypotheses testing, the researcher explored how the teachers in the Sissala East Municipality adhere to the basic principles of testing (construction, administration, scoring). The results showed that teachers flouted some of the guidelines in testing. Therefore, the Directorate of the Ghana Education Service (GES) in the Sissala East Municipality is encouraged, as part of their training and development programmes, to continue to organise special seminars and workshops, which will sensitise teachers in the municipality on the essence of

developing quality classroom assessments, especially, in the area of test construction, administration and scoring.

2. The results from the hypotheses testing showed that teacher' demographic characteristics; gender, years of teaching, academic qualification and subjects teachers teach either individually or combined together, have no influence on their testing practices. Therefore, it is suggested that when planning interventions such as workshops on assessment practices for teachers in the Sissala East Municipality, the Directorate, GES, should not place premium on teachers' demographic characteristics.

Suggestions for Future Research

The following are the suggestions for further research:

1. Similar studies should be carried out to examine whether school factors – class size, infrastructure; government and school policy, and leadership style of headteacher, predict teachers' testing practices.
2. Future studies are needed to investigate the relationship between primary school teachers' personal characteristics and their testing practices.
3. There is also the need to replicate the study in other Districts, Municipalities, Metropolis and Regions in Ghana, since these geographical boundaries could vary in their internal policies, and therefore, may yield different study findings, hence leading to a broad and comprehensive understanding of the same issue phenomenon.

REFERENCES

- Adeyemi, B. (2010). Teacher Related Factors as Correlates of Pupils Achievement in Social Studies in South West Nigeria. *Electronic journal of Research in Educational psychology*, 8 (1): 313-332.
- Adom, D., Mensah, J. A., & Dake, D. A. (2020). Test, Measurement, and Evaluation: Understanding and Use of the Concepts in Education.
- Afe, J. O. (2001). Reflections on becoming a teacher and the challenges of teacher education. Inaugural lecture series 64. *Benin City: University of Benin.*
- Agu, N., Onyekuba, C., & Anyichie, A. C. (2012). Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. *Academic Journal*, 8(8), 431-439.
- Agyedu, G. O., Donkor, F., & Obeng, S. (1999). Teach yourself research methods. *Kumasi, Ghana: Unpublished UEW Course Book.*
- Ajayi, O. S. (2009). *Effective Teaching of Physics*. A paper presented at a seminar on effective teaching of science in Ekiti State organized by Ekiti State Ministry of Science and Technology, Ado-Ekiti.
- Alkharusi, H. (2015). An Evaluation of the Measurement of Perceived Classroom Assessment Environment. *International Journal of Instruction*, 8(2), 45-54.
- Alkharusi, H., (2011). Teachers' classroom assessment skills: influence of gender, subject area, grade level, teaching experience and in-service assessment training. *Journal of Turkish Sciences Education*, 8 (2)

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*.
Monterey, CA: Brooks.
- Allen, M. J., & Yen, W. M. (2002). *Introduction to Measurement
Theory. Prospect heights, Ill Waveland: Waveland Press Inc, 1.*
- Allen, M. J., & Yen, W. M. (2012). *Introduction to measurement theory*.
USA: Waveland Press Inc.
- Amedahe, F. K. & Asamoah- Gyimah, K. (2003). *Measurement and
evaluation, Cape*
- Amedahe, F. K. (1989). Testing practices in secondary schools in the Central
Region of Ghana. *Unpublished Master's thesis, University of Cape
Coast, Cape Coast.*
- Amedahe, F. K. (2004). Research methods notes for teaching. *Unpublished
Manuscript.*
- Amedahe, F. K., & Asamoah-Gyimah, K. (2015). *Introduction to educational
research*. Cape Coast: UCC Printing Press.
- American Psychological Association, American Educational Research
Association & National Council on Measurement in Education (ncme).
(2014). *Technical recommendations for psychological test and
diagnostic techniques*. Washington DC: American Psychological
Association.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation.
Journal of educational psychology, 84(3), 261.

- Amoako, I., Asamoah, D., & Bortey, J. (2019). Knowledge of formative assessment practices among senior high school mathematics teachers in Ghana. *African Journal of Humanities and Social Sciences Research (AJHSSR)*, 3 (3), 8-13.
- Anastasi, A. (1982). *Psychological testing*. New York: Macmillan Publishing Company.
- Anhwere, Y. M. (2009). *Assessment practices of teacher training colleges tutors in Ghana*. Unpublished Master's thesis. University of Cape Coast. Cape Coast Ghana.
- Ary, D., Jacobs, L., C., & Sorenen, C., (Eds.). (2010). *Introduction to Research in Education* (8th ed.). Canada: Wadsworth Cengage Learning.
- Asamoah, D., Songnalle, S., Sundeme, B., & Derkye, C. (2019). Gender difference in formative assessment knowledge of senior high school teachers in the upper west region of Ghana. *Journal of Education and Practice*, 10(6).
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational measurement: Issues and practice*, 17(1), 10-17.
- Benzehaf, B. (2017). Exploring teachers' assessment practices and skills. *International Journal of Assessment Tools in Education*, 4(1), 1-18.
- Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, 40(1), 90-98.

- Birnbaum, A., Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. *Some latent trait models and their use in inferring an examinee's ability.*
- Bratton, M., & Mattes, R. (2005). Gyimah-Boadi. *Public Opinion, Democracy and Market Reform in Africa*, 246-8.
- Brookhart, S. M. (1997). Effects of the classroom assessment environment on mathematics and science achievement. *The Journal of Educational Research*, 90(6), 323-330.
- Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers college record.*
- Brookhart, S. M., & Bronowicz, D. L. (2003). 'I Don't Like Writing. It Makes My Fingers Hurt': Students talk about their classroom assessments. *Assessment in Education: Principles, Policy & Practice*, 10(2), 221-242.
- Brookhart, S. M., & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education*, 16(1), 27-54.
- Chalachew, A. A., & Terefe, A. (2020). Teachers' Self-perceived Skills as the function of Gender and Teaching Experiences in the Classroom Assessment: A study in High Schools of South West Shewa Zone, Ethiopia. *International Journal of Progressive Education*, 16(1), 11-24.

- Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of educational psychology*, 93(1), 43. Coast, Ghana: Centre for Continuing Education
- Cohen, L., Mansion, I., Morrison, K. (2007). *Research methods in education* (6th ed.). New York: Routledge.
- Conlan, T. J., Posner, P. L., Conlon, T. J., & Posner, P. L. (2008). Intergovernmental management and the challenges ahead. *Intergovernmental management for the 21st century*, 1-9.
- Cormack, D. (2013). *The research process in nursing*. John Wiley & Sons.
- Couchman, W., & Dawson, J. (1995). Nursing and healthcare: a practical guide. *Balliere Tindall in Association with the Royal College of Nursing, London*.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.
- Crocker, L., & Algina, J. (2008). Introduction to classical and modern test theory Mason. *OH: Cengage Learning*.
- Cunningham, G. K. (1986). *Educational and psychological measurement*. New York: Macmillan Publishing Company.
- Dambudzo, I. I. (2005). *The relationship between learner self-concept and achievement in secondary schools in Zimbabwe* (Doctoral dissertation).

- DeVellis, R. F. (1991). Guidelines in scale development. *Scale Development: Theory and Applications*. Newbury Park, Calif: Sage, 5191.
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and perspectives*, 38(1), 105.
- DuBois, P. H. (1970). *A history of psychological testing*. Allyn and Bacon.
- Ebel, R. L., & Frisbie, D. A. (1972). *Essentials of educational measurement* (pp. 492-494). Englewood Cliffs, NJ: Prentice-Hall.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of Educational Measurement*. 5th edn. Englewoods Cliffs.
- Etsey, Y. K. A. (2004). *Educational measurement and evaluation*. Lecture notes on EPS 203. Unpublished document, University of Cape Coast, Ghana.
- Etsey, Y. K. A. (2012). *Assessment in education*. University of Cape Coast: Unpublished.
- Fajonyomi, A. A. (2007). Reviving higher education relevance in Nigeria. A paper presented at the *3rd Faculty of Week and Annual National Conference, Faculty of Education*. University of Maiduguri on the theme Higher education reforms in Nigeria from 26th to 27th September.
- Fauth, G. C. (2016). Women educational administration: A research profile. *The Education Forum*, 59, 61-67.
- Feed the Future Ghana District Profile series (2017). Sissala East Municipality.
- Field, A. (2011). *Discovering statistics using spss (3rd ed.)*. London, United Kindom: Sage Publication Ltd.

- Fincham, J. E. (2008). Response rates and responsiveness for surveys, standards, and the *Journal American journal of pharmaceutical education*, 72(2).
- Fink, A. (2015). *How to conduct surveys: A step-by-step guide*. Sage Publications.
- Flanagan, D, Genshaft, J. L. & Harrison, P. L. (1997). *Intellectual assessment, tests, and issues*. New York: The Guilford Press.
- Fraenkel, J. R., & Wallen, W. E. (2000). How to design and evaluate educational research.
- Frey, B. B., & Schmitt, V. L. (2010). Teachers' classroom assessment practices. *Middle Grades Research Journal*, 5(3).
- Gemeda, T., & Getachew, A., (2019). Continuous assessment perception of Madda Walabu University instructors, South East Ethiopia. *Journal of Education Practice* 10 (4).
- Ghana Education Service (2020). *Statistics on Junior High School teachers: Sissala East Municipal Office*. Tumu, Ghana.
- Gronlund, N. E. (1988). *How to construct achievement tests*. Prentice Hall. Pearson.
- Groth-Marnat, G. (1997). The Wechsler intelligence scales. *Specific Learning disabilities and difficulties in children and adolescent Psychological assessment and evaluation*, 29-51.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, 95(7-8), 798-812.
- Isard, J. (2005). Overview of test construction. *Quantitative Research Methods in Educational Planning içinde*, 6.

- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *science*, 319(5865), 966-968.
- Kini, T., & Podolsky, A. (2016). Does teaching experience increase teacher effectiveness. *A Review of the Research*, 1-72.
- Koloi-Keaitkitse, S. (2012). *Classroom assessment practices: a survey of Botswana primary and secondary school teachers*. Unpublished doctoral dissertation, Ball State University.
- Krejcie, R.V., & Morgan, D.W. (1970). *Determining sample size for research activities*
- Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, 12(2), 241-279.
- Lissitz, R. W., & Schafer, W. D. (2002). *Assessment in Educational Reform: Both Means and Ends*. College Division, Allyn & Bacon, Inc., 75 Arlington Street, Suite 300, Boston. MA 02116.
- LoBiondo-Wood, G., & Haber, J. (1998). *Nursing Research: Methods. Critical appraisal and Utilization.*, Mosby, St Louis.
- Lorna M. E., & Manitoba. School Programs Division. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Manitoba Education, Citizenship and Youth.
- Mall, A. L. (2014). Development and validation of indicators of secondary mathematics teachers' positive dispositions toward problem solving.
- Maree, K. (2007). *First steps in research*. Van Schaik Publishers.

- Marriot, P., & Lau, A. (2008). The use of on-line summative assessment in undergraduate financial accounting course. *Journal of Accounting Education, 26*,773-90.
- McDaniel, E. (1994). *Understanding educational measurement*. Madison, Wisconsin: WCB Brown & Benchmark.
- Mehrens, W. A. & Lehman, J. J. (1991). *Measurement and evaluation in education and psychology*. New York: Harcourt Brace College Publishers.
- Messick, S. (1989). Test validity: A matter of consequence. *Social Indicators Research, 45*(1-3), 35-44.
- Messick, S. (2003). The psychology of educational measurement. *Journal of Educational Measurement, 21*, 215-237.
- Ministry of Education (2018). *National Pre-Tertiary Education Curriculum Framework*, Cantonments, Accra, Ghana.
- Mohuidin, G., (2015). A study on educational assessment of secondary and higher secondary teachers of Bangladesh. *World Scientific News, 16*, 95-115.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of educational Research, 62*(3), 229-258.
- Mugenda, O. M., & Mugenda, M. AG (2003). *Research Methods: Quantitative and Qualitative Approaches*, Nairobi: African Centre Technology Studies press (ACTS).

- Murphy, E. J. (2009). The far-infrared-radio correlation at high redshifts: physical considerations and prospects for the square kilometer array. *The Astrophysical Journal*, 706(1), 482.
- National Council on Measurement in Education (NCME). (2014). Standards for educational and psychological testing. Washington, D. C: Author
- Nenty, H. J. (1992). Attributional theory and performance in mathematics. *Obudu Journal of Educational Studies.*, 1(1), 130-138.
- Nitko, A. J. (1996). *Educational assessment of students*. Prentice-Hall Order Processing Center, PO Box 11071, Des Moines, IA 50336-1071.
- Nitko, A. J. (2001). *Educational assessment of students* (3rd. ed.). Upper Saddle River, NJ: Merrill.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3 (1), 1-18.
- Oduro-Okyireh, G. (2008). *Testing practices of senior secondary school teachers in the Ashanti region of Ghana* (Doctoral dissertation, University of Cape Coast).
- Ogunkola, B. J. (2013). Instructional assessment practices of science teachers in Barbados: pattern, techniques and challenges. *Academic Journal of Interdisciplinary Studies*, 2(1), 57-65.
- Ololube, N. P. (2008). Evaluation competencies of professional and non-professional teachers in Nigeria. *Studies in Educational Evaluation*, 34(1), 44-51.
- Onyechere, I. (2000). New face of examination malpractice among Nigerian youths. *The Guardian Newspaper July*, 16.

- Opara, I. M., (2018). Teachers' characteristics as determinants of their attitude towards practices. *American Journal of Educational Research*. 6 (100), 1351-1355.
- Opara, I. M., Onyekuru, B. U., & Njoku, J. U. (2015). Predictive Power of School Based Assessment Scores on Students' Achievement in Junior Secondary Certificate Examination (JSCE) in English and Mathematics. *Journal of Education and Practice*, 6(9), 112-116.
- Owusu-Mensah, J. (2019). Teacher characteristics as correlate of classroom assessment practices at Atwima Nwanbiagya South District. Unpublished Master's thesis. University of Cape Coast. Cape Coast Ghana.
- Pavet, W., Diener, E., Colvin, C. R., & Sandvick, E. (1991). Further validation of satisfaction with life scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment*. 57, 149-161.
- Polit, D. F., & Beck, C. T. (2004). *Nursing research: Principles and methods*. Lippincott Williams & Wilkins.
- Quaigrain, A. K. (1992). Teacher competence in the use of essay tests: A study of secondary schools in the western region of Ghana. *Unpublished Master's thesis, University of Cape Coast, Ghana*.
- Quansah, F., & Amoako, I. (2018). Attitude of Senior High School (SHS) teachers towards test construction: Developing and validating a standardised instrument. *Research on Humanities and Social Sciences*, 8(1), 25-30.

- Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' Test Construction Skills in Senior High Schools in Ghana: Document Analysis. *International Journal of Assessment Tools in Education*, 6(1), 1-8.
- Reynolds, C., R., Livingston, R. B., & Willson, V. (2009). *Measurement and assessment in education* (2nd. ed.). Ohio: Pearson.
- Rowntree, D. (1987). *Assessing students: How shall we know them?* Taylor & Francis.
- Saefurrohman & Balinas, E. S., (2016). English teachers' classroom assessment practices. *International Journal of Evaluation and Research in Education*. 5 (1),82-92.
- Sahinkarakas, S., (2012). The role of teaching experience on teachers' perception of language assessment. *Procedure- Social and Behavioral Sciences*. 47, 17871792.
- Sarantakos, S. (2005). *Social research*. New York: Palgrave Macmillan.
- Seweje, R. O., & Jegede, S. A. (2005). *Science Education and Science Teaching Methods*. Lagos: Atlantic Association.
- Sofu, S., Ocansey, R. T., Nabie, M. J., & Asola, E. F. (2013). Assessment practices among Secondary Physical Education Teachers in Ghana. *International Online Journal of Educational Sciences*, 5(2), 274-281.
- Stainback, S. B. E., & Stainback, W. C. (1996). *Inclusion: A guide for educators*. Paul H Brookes Publishing.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Suny Press.

- Susuwele-Banda, W. J. (2005). *Classroom assessment in Malawi: teachers' perception and practices in mathematics*. Unpublished doctoral dissertation, faculty of the Virginia Polytechnic Institute and State University. Retrieved from <http://hdl.handle.net/10919/26269>.
- Tamakloe, T. K. Atta, & Amedahe, F. (1996). *Principles and methods of teaching*.
- Tinkelman, S. N. (1971). Planning the objective test. *Educational measurement*, 46-80.
- Trochim, W. (2000). *The research methods knowledge base* (2nd ed.). Cincinnati, O. H.: Atomic Dog Publishing.
- Trochim, W. M. (2006). Idea of construct validity. *Access April*, 23, 2009.
- Van de Watering, G., Gijbels, D., Dochy, F., & Van der Rijt, J. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education*, 56(6), 645.
- Wiliam, D. (1993). Validity, dependability and reliability in National Curriculum assessment. *The Curriculum Journal*, 4(3), 335-350.
- Zhang, Z. R., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers 'self-perceived assessment skills. *Applied Measurement in Education*, 16, 323-342.





APPENDIX C

QUESTIONNAIRE

UNIVERSITY OF CAPE COAST

DEPARTMENT OF EDUCATION AND PSYCHOLOGY

The purpose of this questionnaire is to collect information on teachers' assessment practices. It would therefore, be appreciated if you could please, provide honest answers to the questionnaire items. You are assured of complete confidentiality and anonymity of every information provided

SECTION A

DIRECTIONS: Please tick (√) the box that best describes your response(s) where applicable.

1. **Gender:**

a) Male []

b) Female []

2. **Number of years in teaching service**

a) 5 years -less []

b) 6 – 10 years []

c) 11-15 years []

d) 16-20 years []

e) 21-more []

3. **Educational Qualification:**

a) Teachers' Certificate A []

b) Diploma with Education []

c) Bachelors with Education []

d) Bachelors without Education []

e) Masters with Education []

f) Masters without Education []

4. Which of the following core subject areas do you teach?

a) English language []

b) Mathematics []

c) Integrated Science []

d) Social Studies []

SECTION B

DIRECTIONS: Please tick (√) the cell that indicates closely how frequently you practice the following **test construction principles** using the scale:

Always-A; Very Often-VO; Sometimes-ST; Rarely- R; Never-N

In defining the purpose of my test, I					
	A	VO	ST	R	N
5. relate the instructional objectives of the subject matter to the test.					
6. inform students in advance before the test.					
7. inform students about how the scores will be used, immediately after the test is written.					
In determining the question format to use, I					
8. select the test format suitable for testing stated objectives.					
9. select the test format that is convenient for scoring the test.					
In determining what is to be tested, I					
10. prepare and use a table of specifications or test blueprint in selecting the questions on the test.					
11. inform students what chapters or units the test will cover.					

In writing the individual questions, I					
12. prepare more questions than needed in the test.					
13. write the test questions in advance (at least two weeks) of the test date to permit reviews and editing.					
14. write the test questions a day before the test to avoid leakage of questions.					
In reviewing the test questions, I					
15. review the test questions (at least one week) after they have been written, by reading over them.					
16. ensure that the test questions are challenging to the high achievers who think they are very intelligent.					
In preparing the scoring key, I					
17. prepare the marking scheme immediately after the test is written.					
In writing the directions / instructions for the test, I					
18. write clear and concise directions for the entire test and sections of the test.					
In evaluating the test questions, I					
19. ensure that the test questions are simple and clear.					
20. ensure that the test is a representative sampling of the material taught/ the content covered.					
21. ensure that students have enough time to complete the test.					
22. ensure that test is the best instrument to assess the desired knowledge / skills taught.					
23. prepare the students adequately for the test.					

SECTION C

DIRECTIONS: Please tick (√) the cell that closely indicates how frequently you practice the following **test administration principles using the scale:**

Always-A; Very Often-VO; Sometimes-ST; Rarely- R; Never-N

In the administration of the test, I					
	A	VO	ST	R	N
24. give tests immediately before a long vacation, or other important events.					
25. give tests immediately after a long vacation, or other important events.					
26. ensure that the sitting arrangement allows enough space so that students will not copy from each other.					
27. ensure that there is adequate ventilation and lighting in the testing room.					
28. use —'Do Not Disturb' test in Progress sign when students are taking tests.					
29. expect and cater for all possible emergencies e.g. shortage of question papers, answer booklets etc.					
30. announce the remaining time (time left to complete test) at regular intervals.					
31. stand where I can view all students and move among the students once a while, to check on malpractices.					
32. keep all remarks in the testing room to a minimum and make sure they are related to the test.					
33. ask students to work faster during the time of testing in order to finish on time.					
34. tell students the dire consequences of failure in the test they are taking.					

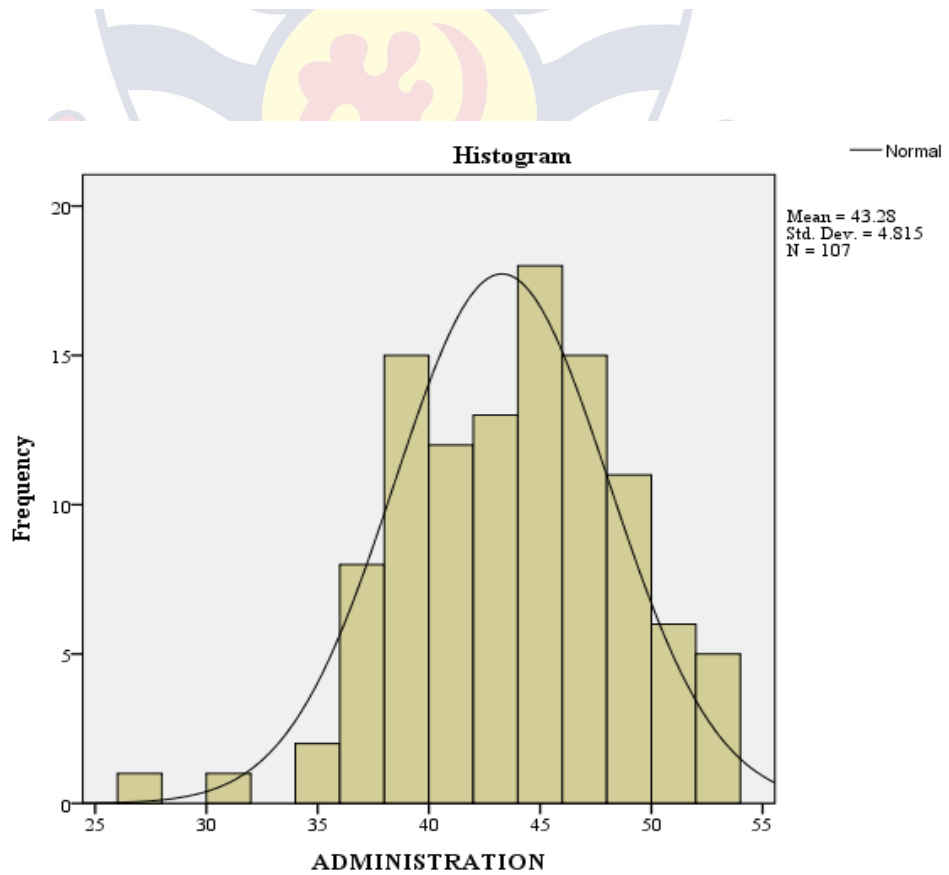
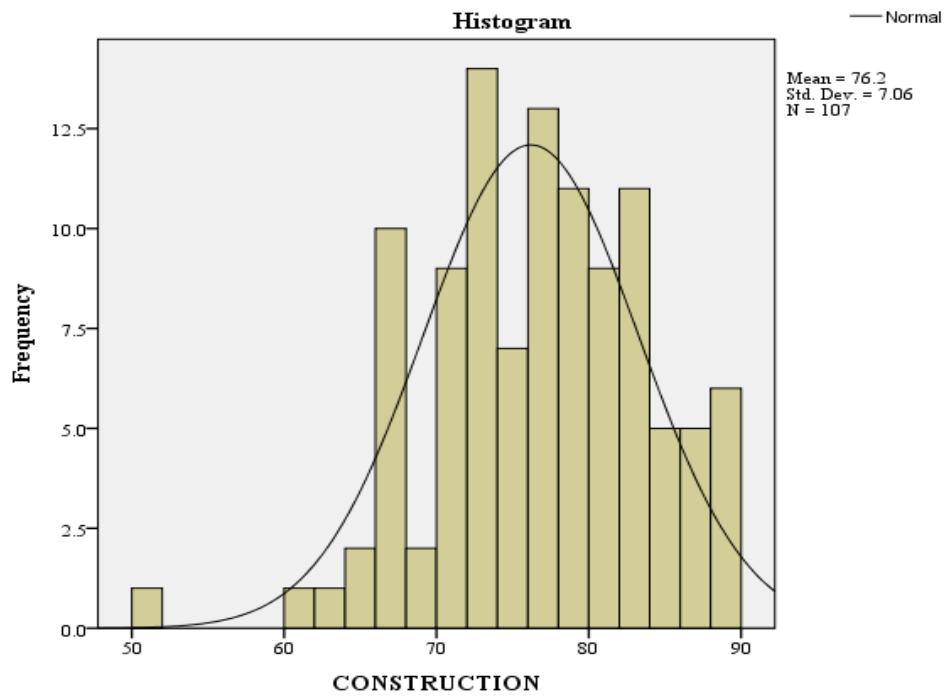
SECTION D

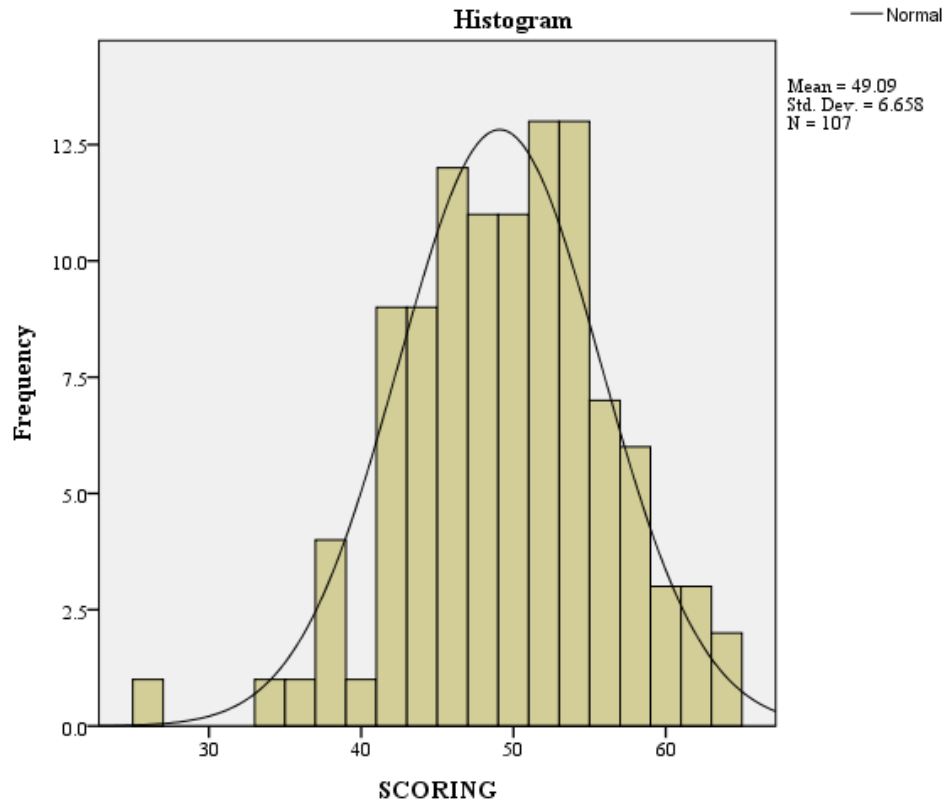
DIRECTIONS: Please tick (√) the cell that closely indicates how frequently you practice the following **test scoring principles using the scale: Always-A; Very Often-VO; Sometimes-ST; Rarely- R; Never-N**

In the scoring of the test, I					
	A	VO	ST	R	N
35. mark papers just after the test is taken.					
36. prepare scoring guide and use it.					
37. make sure test takers are kept anonymous.					
38. score the essay type responses, question by question.					
39. keep scores of previous questions out of sight.					
40. periodically rescore previously scored questions.					
41. shuffle scripts before scoring.					
42. score essay test when I am physically sound in a sound environment.					
43. score essay test when I am mentally alert in a sound environment.					
44. constantly follow scoring guide.					
45. am influenced by scores of the first few papers read when scoring test items.					
46. score a particular question on all papers at a sitting, but break when fatigue sets in.					
47. provide comments and correct errors on scripts.					
48. give extra marks to students based on handwriting, grammar etc. but which are not desired by the test.					

APPENDIX D

TEST FOR NORMALITY

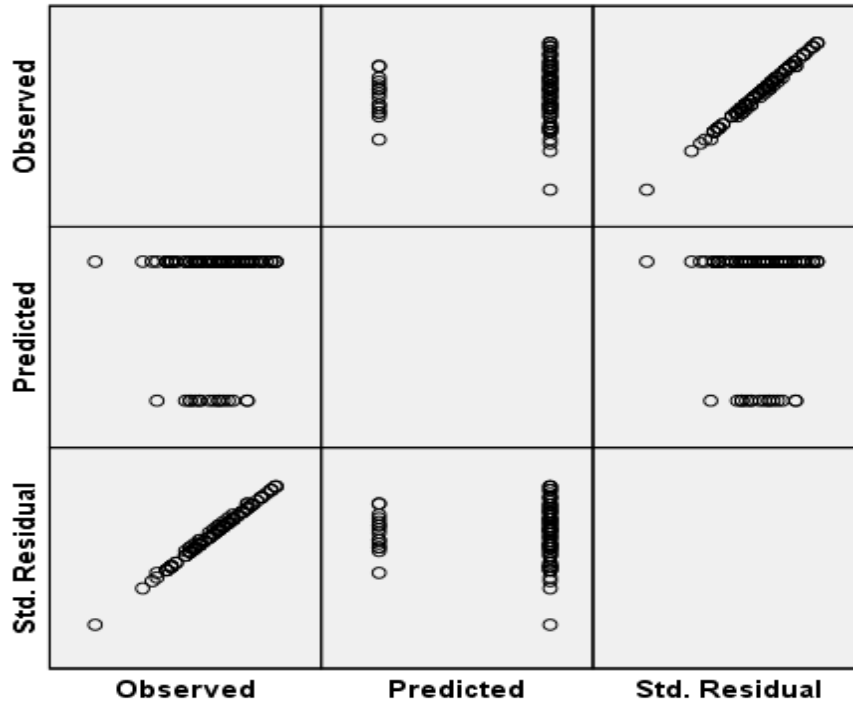




APPENDIX E

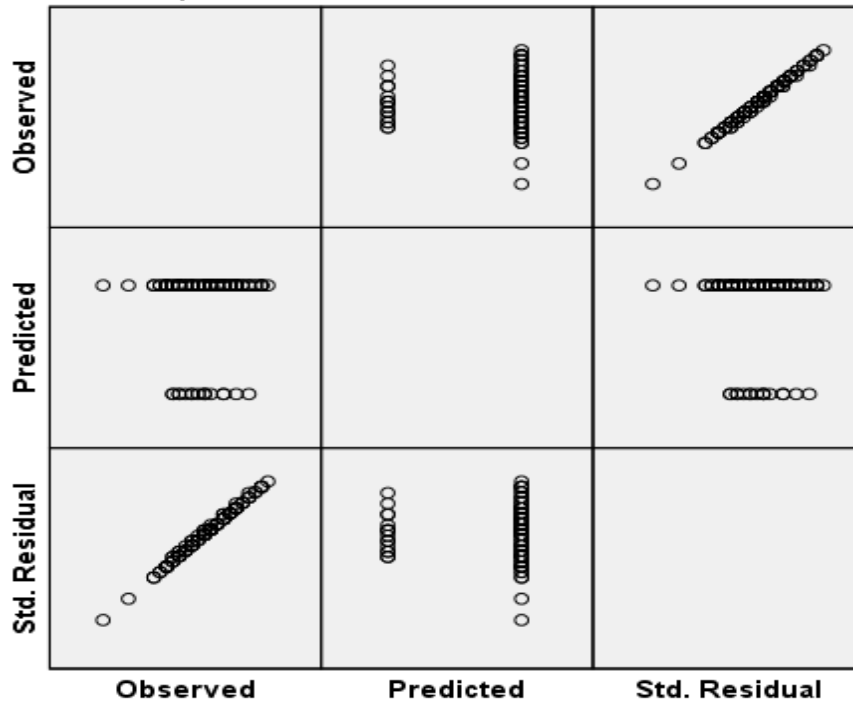
TEST FOR MULTIVARIATE NORMALITY

Dependent Variable: CONSTRUCTION

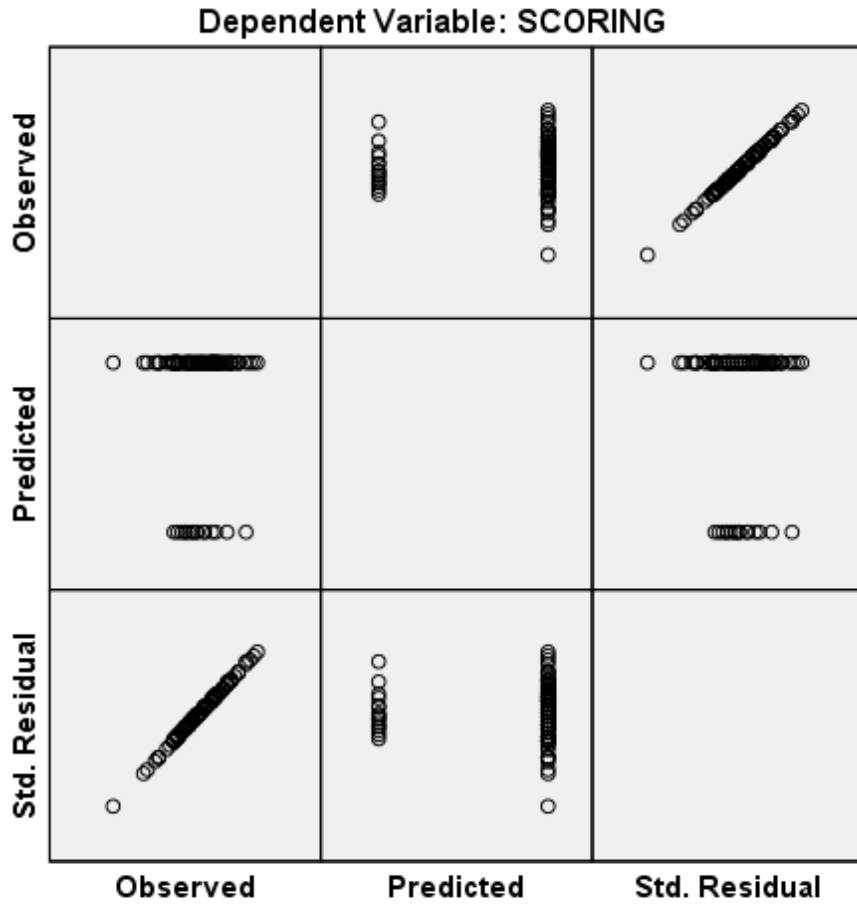


Model: Intercept + Female

Dependent Variable: ADMINISTRATION



Model: Intercept + Female



Model: Intercept + Female



APPENDIX F
RESULTS OUTPUT FOR HYPOTHESIS ONE
Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.991	3957.833 ^b	3.000	103.000	.000	.991
	Wilks' Lambda	.009	3957.833 ^b	3.000	103.000	.000	.991
	Hotelling's Trace	115.277	3957.833 ^b	3.000	103.000	.000	.991
	Roy's Largest Root	115.277	3957.833 ^b	3.000	103.000	.000	.991
Female	Pillai's Trace	.005	.189 ^b	3.000	103.000	.904	.005
	Wilks' Lambda	.995	.189 ^b	3.000	103.000	.904	.005
	Hotelling's Trace	.005	.189 ^b	3.000	103.000	.904	.005
	Roy's Largest Root	.005	.189 ^b	3.000	103.000	.904	.005

a. Design: Intercept + Female

b. Exact statistic

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	CONSTRUCTION	21.590 ^a	1	21.590	.431	.513	.004
	ADMINISTRATION	9.695 ^b	1	9.695	.416	.520	.004
	SCORING	5.304 ^c	1	5.304	.119	.731	.001
Intercept	CONSTRUCTION	530949.462	1	530949.462	10596.205	.000	.990
	ADMINISTRATION	171456.044	1	171456.044	7354.439	.000	.986
	SCORING	220160.176	1	220160.176	4925.009	.000	.979
Female	CONSTRUCTION	21.590	1	21.590	.431	.513	.004
	ADMINISTRATION	9.695	1	9.695	.416	.520	.004
	SCORING	5.304	1	5.304	.119	.731	.001
Error	CONSTRUCTION	5261.288	105	50.108			
	ADMINISTRATION	2447.894	105	23.313			
	SCORING	4693.762	105	44.702			
Total	CONSTRUCTION	626511.000	107				
	ADMINISTRATION	202889.000	107				
	SCORING	262587.000	107				
Corrected Total	CONSTRUCTION	5282.879	106				
	ADMINISTRATION	2457.589	106				
	SCORING	4699.065	106				

a. R Squared = .004 (Adjusted R Squared = -.005)

b. R Squared = .004 (Adjusted R Squared = -.006)

c. R Squared = .001 (Adjusted R Squared = -.008)

**APPENDIX G
RESULTS OUTPUT FOR HYPOTHESIS TWO
Multivariate Tests^a**

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.981	1718.535 ^b	3.000	100.000	.000	.981
	Wilks' Lambda	.019	1718.535 ^b	3.000	100.000	.000	.981
	Hotelling's Trace	51.556	1718.535 ^b	3.000	100.000	.000	.981
	Roy's Largest Root	51.556	1718.535 ^b	3.000	100.000	.000	.981
Six	Pillai's Trace	.004	.123 ^b	3.000	100.000	.946	.004
	Wilks' Lambda	.996	.123 ^b	3.000	100.000	.946	.004
	Hotelling's Trace	.004	.123 ^b	3.000	100.000	.946	.004
	Roy's Largest Root	.004	.123 ^b	3.000	100.000	.946	.004
Eleven	Pillai's Trace	.018	.617 ^b	3.000	100.000	.606	.018
	Wilks' Lambda	.982	.617 ^b	3.000	100.000	.606	.018
	Hotelling's Trace	.019	.617 ^b	3.000	100.000	.606	.018
	Roy's Largest Root	.019	.617 ^b	3.000	100.000	.606	.018
Sixteen	Pillai's Trace	.057	1.998 ^b	3.000	100.000	.119	.057
	Wilks' Lambda	.943	1.998 ^b	3.000	100.000	.119	.057
	Hotelling's Trace	.060	1.998 ^b	3.000	100.000	.119	.057
	Roy's Largest Root	.060	1.998 ^b	3.000	100.000	.119	.057
Twentyone	Pillai's Trace	.049	1.721 ^b	3.000	100.000	.168	.049
	Wilks' Lambda	.951	1.721 ^b	3.000	100.000	.168	.049
	Hotelling's Trace	.052	1.721 ^b	3.000	100.000	.168	.049
	Roy's Largest Root	.052	1.721 ^b	3.000	100.000	.168	.049

a. Design: Intercept + Six + Eleven + Sixteen + Twentyone

b. Exact statistic

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	CONSTRUCTION	558.748 ^a	4	139.687	3.016	.021	.106
	ADMINISTRATION	95.955 ^b	4	23.989	1.036	.392	.039
	SCORING	51.165 ^c	4	12.791	.281	.890	.011
Intercept	CONSTRUCTION	216005.921	1	216005.921	4663.843	.000	.979
	ADMINISTRATION	69404.632	1	69404.632	2997.616	.000	.967
	SCORING	90748.658	1	90748.658	1991.515	.000	.951
Six	CONSTRUCTION	11.842	1	11.842	.256	.614	.003
	ADMINISTRATION	.053	1	.053	.002	.962	.000
	SCORING	1.895	1	1.895	.042	.839	.000
Eleven	CONSTRUCTION	72.882	1	72.882	1.574	.213	.015
	ADMINISTRATION	22.232	1	22.232	.960	.329	.009
	SCORING	10.278	1	10.278	.226	.636	.002
Sixteen	CONSTRUCTION	267.992	1	267.992	5.786	.018	.054
	ADMINISTRATION	22.858	1	22.858	.987	.323	.010
	SCORING	25.063	1	25.063	.550	.460	.005
Twentyone	CONSTRUCTION	170.140	1	170.140	3.674	.058	.035

	ADMINISTRATION	59.292	1	59.292	2.561	.113	.024
	SCORING	7.835	1	7.835	.172	.679	.002
	CONSTRUCTION	4724.131	102	46.315			
Error	ADMINISTRATION	2361.634	102	23.153			
	SCORING	4647.900	102	45.568			
	CONSTRUCTION	626511.000	107				
Total	ADMINISTRATION	202889.000	107				
	SCORING	262587.000	107				
	CONSTRUCTION	5282.879	106				
Corrected Total	ADMINISTRATION	2457.589	106				
	SCORING	4699.065	106				

a. R Squared = .106 (Adjusted R Squared = .071)

b. R Squared = .039 (Adjusted R Squared = .001)

c. R Squared = .011 (Adjusted R Squared = -.028)



APPENDIX H
RESULTS OUTPUT FOR HYPOTHESIS THREE
Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.985	2194.978 ^b	3.000	100.000	.000	.985
	Wilks' Lambda	.015	2194.978 ^b	3.000	100.000	.000	.985
	Hotelling's Trace	65.849	2194.978 ^b	3.000	100.000	.000	.985
	Roy's Largest Root	65.849	2194.978 ^b	3.000	100.000	.000	.985
Certificate	Pillai's Trace	.024	.812 ^b	3.000	100.000	.490	.024
	Wilks' Lambda	.976	.812 ^b	3.000	100.000	.490	.024
	Hotelling's Trace	.024	.812 ^b	3.000	100.000	.490	.024
	Roy's Largest Root	.024	.812 ^b	3.000	100.000	.490	.024
Bachelorseducation	Pillai's Trace	.013	.445 ^b	3.000	100.000	.722	.013
	Wilks' Lambda	.987	.445 ^b	3.000	100.000	.722	.013
	Hotelling's Trace	.013	.445 ^b	3.000	100.000	.722	.013
	Roy's Largest Root	.013	.445 ^b	3.000	100.000	.722	.013
Bachelorswithout	Pillai's Trace	.040	1.406 ^b	3.000	100.000	.246	.040
	Wilks' Lambda	.960	1.406 ^b	3.000	100.000	.246	.040
	Hotelling's Trace	.042	1.406 ^b	3.000	100.000	.246	.040
	Roy's Largest Root	.042	1.406 ^b	3.000	100.000	.246	.040
MastersEducation	Pillai's Trace	.019	.644 ^b	3.000	100.000	.589	.019
	Wilks' Lambda	.981	.644 ^b	3.000	100.000	.589	.019
	Hotelling's Trace	.019	.644 ^b	3.000	100.000	.589	.019
	Roy's Largest Root	.019	.644 ^b	3.000	100.000	.589	.019
MastersWithout	Pillai's Trace	.000	. ^b	.000	.000	.	.
	Wilks' Lambda	1.000	. ^b	.000	101.000	.	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.	.
	Roy's Largest Root	.000	.000 ^b	3.000	99.000	1.000	.000

a. Design: Intercept + Certificate + Bachelorseducation + Bachelorswithout + MastersEducation + MastersWithout

b. Exact statistic

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	CONSTRUCTION	304.579 ^a	4	76.145	1.560	.191	.058
	ADMINISTRATION	88.363 ^b	4	22.091	.951	.438	.036
	SCORING	231.969 ^c	4	57.992	1.324	.266	.049
Intercept	CONSTRUCTION	284559.680	1	284559.680	5830.322	.000	.983
	ADMINISTRATION	93398.420	1	93398.420	4020.992	.000	.975
	SCORING	120442.320	1	120442.320	2750.134	.000	.964
Certificate	CONSTRUCTION	109.327	1	109.327	2.240	.138	.021
	ADMINISTRATION	3.106	1	3.106	.134	.715	.001
	SCORING	8.359	1	8.359	.191	.663	.002
Bachelors education	CONSTRUCTION	26.446	1	26.446	.542	.463	.005
	ADMINISTRATION	.026	1	.026	.001	.973	.000
	SCORING	7.097	1	7.097	.162	.688	.002
Bachelors without	CONSTRUCTION	137.162	1	137.162	2.810	.097	.027
	ADMINISTRATION	36.587	1	36.587	1.575	.212	.015
	SCORING	157.374	1	157.374	3.593	.061	.034
Masters Education	CONSTRUCTION	46.930	1	46.930	.962	.329	.009
	ADMINISTRATION	42.843	1	42.843	1.844	.177	.018

Masters Without	SCORING	32.012	1	32.012	.731	.395	.007
	CONSTRUCTION	.000	0000
	ADMINISTRATION	.000	0000
	SCORING	.000	0000
Error	CONSTRUCTION	4978.299	102	48.807			
	ADMINISTRATION	2369.226	102	23.228			
	SCORING	4467.097	102	43.795			
Total	CONSTRUCTION	626511.000	107				
	ADMINISTRATION	202889.000	107				
	SCORING	262587.000	107				
Corrected Total	CONSTRUCTION	5282.879	106				
	ADMINISTRATION	2457.589	106				
	SCORING	4699.065	106				

- a. R Squared = .058 (Adjusted R Squared = .021)
 b. R Squared = .036 (Adjusted R Squared = -.002)
 c. R Squared = .049 (Adjusted R Squared = .012)

APPENDIX I

RESULTS OUTPUT FOR HYPOTHESIS FOUR

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.976	1375.090 ^b	3.000	101.000	.000	.976
	Wilks' Lambda	.024	1375.090 ^b	3.000	101.000	.000	.976
	Hotelling's Trace	40.844	1375.090 ^b	3.000	101.000	.000	.976
	Roy's Largest Root	40.844	1375.090 ^b	3.000	101.000	.000	.976
English	Pillai's Trace	.097	3.617 ^b	3.000	101.000	.016	.097
	Wilks' Lambda	.903	3.617 ^b	3.000	101.000	.016	.097
	Hotelling's Trace	.107	3.617 ^b	3.000	101.000	.016	.097
	Roy's Largest Root	.107	3.617 ^b	3.000	101.000	.016	.097
Science	Pillai's Trace	.032	1.119 ^b	3.000	101.000	.345	.032
	Wilks' Lambda	.968	1.119 ^b	3.000	101.000	.345	.032
	Hotelling's Trace	.033	1.119 ^b	3.000	101.000	.345	.032
	Roy's Largest Root	.033	1.119 ^b	3.000	101.000	.345	.032
Social	Pillai's Trace	.008	.263 ^b	3.000	101.000	.852	.008
	Wilks' Lambda	.992	.263 ^b	3.000	101.000	.852	.008
	Hotelling's Trace	.008	.263 ^b	3.000	101.000	.852	.008
	Roy's Largest Root	.008	.263 ^b	3.000	101.000	.852	.008

a. Design: Intercept + English + Science + Social

b. Exact statistic

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	CONSTRUCTION	380.570 ^a	3	126.857	2.665	.052	.072
	ADMINISTRATION	6.869 ^b	3	2.290	.096	.962	.003
	SCORING	69.009 ^c	3	23.003	.512	.675	.015
Intercept	CONSTRUCTION	182520.000	1	182520.000	3834.838	.000	.974
	ADMINISTRATION	57116.033	1	57116.033	2400.499	.000	.959
	SCORING	70956.033	1	70956.033	1578.484	.000	.939
English	CONSTRUCTION	248.293	1	248.293	5.217	.024	.048
	ADMINISTRATION	2.681	1	2.681	.113	.738	.001
	SCORING	27.542	1	27.542	.613	.436	.006
Science	CONSTRUCTION	141.204	1	141.204	2.967	.088	.028
	ADMINISTRATION	6.557	1	6.557	.276	.601	.003
	SCORING	7.442	1	7.442	.166	.685	.002
Social	CONSTRUCTION	.175	1	.175	.004	.952	.000
	ADMINISTRATION	1.988	1	1.988	.084	.773	.001
	SCORING	13.076	1	13.076	.291	.591	.003
Error	CONSTRUCTION	4902.309	103	47.595			
	ADMINISTRATION	2450.720	103	23.793			
	SCORING	4630.057	103	44.952			
Total	CONSTRUCTION	626511.000	107				
	ADMINISTRATION	202889.000	107				
	SCORING	262587.000	107				
Corrected Total	CONSTRUCTION	5282.879	106				
	ADMINISTRATION	2457.589	106				
	SCORING	4699.065	106				

- a. R Squared = .072 (Adjusted R Squared = .045)
- b. R Squared = .003 (Adjusted R Squared = -.026)
- c. R Squared = .015 (Adjusted R Squared = -.014)