UNIVERSITY OF CAPE COAST

TEACHERS' TESTING PRACTICES OF ACHIEVEMENT TEST IN JUNIOR HIGH SCHOOLS IN THE SISSALA EAST MUNICIPALITY, GHANA

JALLU ZAKARIYA

2020

UNIVERSITY OF CAPE COAST

TEACHERS' TESTING PRACTICES OF ACHIEVEMENT TEST IN JUNIOR HIGH SCHOOLS IN THE SISSALA EAST MUNICIPALITY, GHANA

BY

JALLU ZAKARIYA

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

JULY 2020

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: ……………….…………………………….…………………………

Supervisors' Declaration

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

Principal Supervisor's Signature ……………………… Date ……………

Name: ……………….…………………………….…………………………

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

Name: ……………….…………………………….…………………………

ABSTRACT

The study sought to investigate achievement test practices of teachers in Junior High Schools in the Sissala East Municipality. Descriptive survey was used to investigate the practice of achievement testing of teachers in Junior High Schools in Sissala East Municipality. The study employed the multistage sampling techniques (purposive, stratified and simple random sampling technique). Questionnaire was used to collect data from 248 Junior High School teachers in the Sissala East Municipality. The results show that, Junior High Schools teachers in the Sissala East Municipality averagely adhere to most principles of test construction. The results again showed that, majority of the teachers in the Sissala East Municipality averagely adhere to test administration principles in their achievement test. The results gave evidence that most Junior High Schools teachers in the Sissala East Municipality have no good scoring skills and this always affect the achievement test scores. Furthermore, it was evident that most of the achievement test strategies were not used among Junior High Schools teachers in the Sissala East Municipality. The results also showed that there are numerous challenges that confront the use of achievement test among Junior High Schools teachers in the Sissala East Municipality. Large class size is one of the problems that most teachers complained of. The study concluded that teachers in the Sissala East Municipality were not well equipped with test construction, administration and scoring skills. It was recommended that more workshops and in-service training should be organized to teachers in Junior High Schools with respect to their testing practices (construction, administration and scoring of tests).

ACKNOWLEDGEMENTS

I wish to acknowledge the patience, commitment and constructive suggestions of my supervisors, Dr. Kenneth Asamoah-Gyimah, and Dr. Andrews Cobbinah who dedicated their precious time out of their tight schedule and provided intellectual support, guidance and mentorship throughout the conduct of this study. Their zeal and over all supervision of the thesis is highly appreciated. I wish to acknowledge the invaluable support of Mr. Ishaku Tontine, and Mr. Yahaya Basuglo who availed themselves as field assistants and facilitated the data collection process. I am thankful for the support and encouragement I receive from Mr. Baluwie Nalwie Salifu and Mr. Godfred Bavero Kanton.

I am highly grateful to the authorities of the Tumu Municipal Education Office in Sissala East Municipality for the support they offered me in the conduct of the study. Finally I want to thank all my respondents for their time, patience and tolerance without which this study wouldn’t have been executed.

However, I am entirely responsible for any errors and omissions that might be found in this thesis.

DEDICATION

To my late father, Salifu Lutiyeh Bacheneh, my mother Alima Salifu and daughter, Hibbah Zelezomo Jallu.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

In education, it is undeniable fact that, testing has a colossal impact on the practice of education, and it materializes large in the minds of countless families as they decide the next step of action after they have the glimpse of performance of their wards. Classroom teacher made testing seems reassuringly straightforward and common tool used to assess students in almost all levels of education in Ghana. Precisely, because of the importance given to test scores in our society, any mistake that may emerge from the test can have serious consequences in educational decision making.

Background to the Study

Assessment can be seen as an umbrella term which includes the use of various strategies and methods to determine the extent to which students are achieving the predetermined learning objectives and outcomes of a lesson (Mussawy, 2009). There are different types of assessment which can be used to test students' knowledge and see their current levels in specific subjects. Two major types of assessment widely used are traditional types of assessment and performance-based assessment (Birenaum, & Feldman, 1998). Birenaum, and Feldman, (1998) argue that traditional types of assessment tools are generally knowledge-based and include conventional types of tests such as multiple-choice questions, short answer essays or constructed responses and standardized tests whereas in performance based assessments, students are required to perform a task rather than select from options provided and students are assessed according to their performance outcomes and the extent to which those outcomes are in relation to the rubrics or feedback tools. An achievement assessment test requires students to exhibit the extent of their learning through a demonstration of mastery (Poikela, 2004).

It is absolutely impossible for anybody to study in an entire educational system without being exposed to a wide range of educational and psychological assessment procedures. This is because constantly in an educational system, decisions have to be made about students, curricula and programmes, and educational policies. According to Nitko (1996), decisions about students include managing classroom instruction, placing students into different types of programmes, assigning them to appropriate categories, guiding and counselling them, selecting them for educational opportunities and credentialing and certifying their competence. Decisions about curricula and programmes include decisions about their effectiveness (summative assessment) and about ways to improve them (formative assessment). In Ghana, decisions about educational policies are made at the national level. It is worth knowing, however, that educational assessments, of which in the Ghanaian educational system, tests predominate, provide some of the needed information for these types of decisions.

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). However, it must be noted that the psychological attributes of an individual cannot be measured directly as can height or weight. The existence of such psychological construct can never be absolutely confirmed. The degree to which any attribute characterises an individual can only be inferred from observation of his or her behaviour. It becomes more prudent if one can quantitatively relate the subjective judgments of individuals about the estimated amount of construct or trait that exist in a person by establishing standards for such measurement.

Test is an essential tool that helps to quantify such constructs which helps one to make a value judgment about the degree to which such constructs might probably exist in an individual. A large number of assessment techniques may be used to collect information about students. These include formal and informal observation of students, paper-and-pencil test, a student's performance on homework, laboratory work, and projects during oral questioning and analysis of students' records.

Teachers in the educational setting would want to estimate the degree to which their students are characterize by the knowledge they have imparted to them within a given period. All the domain of such construct might not be known by a single test. Nevertheless, a well-constructed test could sample to a large extent a reasonable amount of the construct on which value judgment could be made from. Educators and teachers must also be aware that a test itself is subject to errors which adversely could affect its use in making decision about students. According to Daniel (2008), "a test score is just one indicator of what a student has learned 3an exceptionally useful one in many ways, but nonetheless one that is unavoidably incomplete and somewhat error prone" (p. 10). Tom and Gary (2003) further asserted that:

1. tests are only tools, and tools can be appropriately used, unintentionally misused, and intentionally abused.
2. tests, like other tools, can be well designed or poorly designed.
3. both poorly designed tools and well-designed tools in the hands of ill-trained or inexperienced users can be dangerous. (p. 1).

They went further to state that test misuse and abuse can occur when users of test results are unaware of the factors that can influence the usefulness of the test scores. Among the major factors are the technical adequacy of a test and its validity and reliability. The technical inadequacies might emerge from factors such as, test appropriateness for the purpose of testing, the content validity evidence, the appropriateness of the reading level, language proficiency and cultural characteristics of students and teachers and pupils' factors that may have affected administration procedure and scoring of the test, among others. It must also be noted that even when a test is technically adequate, misuse and abuse can occur because technical adequacy does not ensure that test scores are accurate or meaningful.

When students' achievement levels are not properly measured and interpreted, the teachers and school administrators will not be able to provide the right educational opportunities and support each individual student needs. Testing provides feedback on which educational decisions are made. These decisions may be the ones that require information about the success of learning programmes or about students who have reached particular levels of skill and knowledge (Izard, 2005). Accurate and valid information about student achievement is widely understood to be essential for effective instruction, as it enables teachers to give appropriate feedback and adapt their instruction to match student needs. However, there is much less agreement about the relative merit of different measurement methods used to obtain this information. Previous research has often found substantial positive correlations between teacher judgments of student achievement and the scores the students obtain on standardized tests. However, the strength of this association has been asserted to be varying considerably across subjects, grades, and teachers (Hoge & Coladarci 1989; Perry & Meisels, 1996).

Tests are indispensable tools in every educational system. Tests and teaching are interwoven. Quaigrain (1992) has stated that-tests provide needed information for evaluation. Without evaluation there cannot be feedback and knowledge of results. Without knowledge of results there cannot be any systematic improvement in learning. In the Ghanaian educational system, standardised achievement, aptitude, and intelligence tests that are found in the developed countries such as the United States of America (USA), Canada and Great Britain are to a large extent non-existent. The tests that are conducted by the West African Examinations Council (WAEC) at the terminal points of the educational system cannot be said to be standardised since they do not meet all the standard characteristics of standardised achievement tests. Examples of the WAEC conducted tests are the Basic Education Certificate Examination (BECE) and the Senior Secondary School Certificate Examination (SSSCE).

Statement of the Problem

Testing at the basic schools assumes that, most teachers have had a course or training in "testing" as part of the assessment process at their various colleges of education. Previous research has indicated that most of the teachers in the second cycle institutions in Ghana lacked the basic test construction skills. This was justified by the findings that not all teachers in the Secondary Schools in Ghana have undergone professional training in testing techniques (Amedahe, 1989).

The studies by Amedahe (1989) and Quagrain (1992) revealed that most Ghanaian teachers had limited skills for constructing the objective and essay type tests, which are the most frequently used instruments in our schools. The study of Amedahe (1989) showed that, to a great extent, secondary school teachers in the Central Region did not follow the basic prescribed principles of classroom test construction. Quagrain replicated the study of Amedahe in 1992 and confirmed the report of Amedahe. This is because most initial teacher training programmes do not make adequate provision for a course in testing. Amedahe (2000) stated that "teacher –made tests may be made of a number of factors, notably among them are, training in assessment techniques, class size and a particular school's policy in assessment with implications on validity and reliability of the assessment results" (p. 112-113).

On contrary to those previous studies, Oduro (2000) concluded in his study that to a great extent, teachers followed the basic principles in test construction, administration and scoring. The findings of the study of Boakye (2016) also revealed that teachers to some extent adhered to the basic principles of test practices. Could it be seen that because these two studies Oduro (2000) and Boakye (2016) were conducted at Ashanti region. In contrast, Sasu (2017) repeated the same study in Central region and found out that teachers, to some extent, have little knowledge in test construction which was in support of the studies by Amedahe (1989) and Quagrain (1992).

Gleaning from the literature, it was evident that the numerous studies in assessment practices have focused attention on teachers in the southern part of Ghana, however, in the case of those in the northern part of the country, it appears much have not been documented. Interestingly, as I went around and interacted with some headmasters and Directors of Education in some districts in the northern part of the country, it appeared that most teachers within some of the districts in the Northern part of Ghana are not professionally trained teachers. Amasingly, these teachers construct, administer, score, and interpret results of their students. The question that really comes to mind is the soundness and appropriateness of results from these assessment results. It is pertinent to examine the achievement testing practices among teachers in the Sissala East, since this would bring to bear the extent to which junior high school teachers' practices, and this would help identify the lapses in assessment practices in order to provide an antidote to the situation, Hence, the need to conduct this study in a different region to help throw more light to teachers test practices in Ghana.

Purpose of the Study

The purpose of the study was to investigate achievement test practices of teachers in Junior High Schools in the Sissala East Municipality. Specifically, the study sought to:

1. asses how Junior High Schools teachers in the Sissala East Municipality adhere to principles of test construction, administration and scoring.

2. find out the kinds of achievement test strategies Junior High Schools teachers in the Sissala East Municipality use to assess their students' learning outcomes.
3. investigate the challenges Junior High Schools teachers in the Sissala East Municipality encounter in the use of achievement test
4. assess difference among the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality with respect to how they adhere to test construction

Research Questions

In order to achieve the purpose of the study, the following research questions were posed.

1. How do Junior High Schools teachers in the Sissala East Municipality adhere to the following principles of test:

a. construction

b. administration

c. scoring

2. What kinds of achievement test strategies do Junior High Schools teachers in the Sissala East Municipality use to assess their students' learning outcomes?
3. What challenges do Junior High Schools teachers in the Sissala East Municipality encounter in the use of achievement test?

Research Hypothesis

Based on the last objective of the study, this research hypothesis was formulated.

H_0:1 there is no statistically significant difference among the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality with respect to how they adhere to test construction.

H_A:1 There is a statistically significant difference among the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality with respect to how they adhere to test construction.

Significance of the Study

The results that were gathered from the study would help stakeholders to determine the state of affairs with respect to achievement testing in the Ghanaian educational system. This, it is believed, will help teachers who received instruction in assessment in education to be up and doing and put their acquired knowledge into practice since testing principles will be related to practice throughout the study. Positive suggestions would be offered as a means of addressing these flaws. It is hoped that these suggestions will help all teachers to improve on their testing practices.

The results of the study will help to enlighten the Junior High Schools teachers in the Sissala East Municipality on their knowledge of assessment in general and achievement test in particular. The findings of this study would help curriculum developers, educators and teachers to understand the impact of teacher's perceptions of achievement tests on instructional practices, student's performance and the goal of education.

Specifically, the findings of this study will inform teachers about the value and impact of achievement test tasks on their instruction. The results of this study would provide insight for curriculum developers, educators and teachers regarding the challenges impeding the effective use of achievement test for appropriate intervention.

Delimitation

The study was confined to only the JHS teachers in the Sissala East Municipality. It focused on only the public JHSs in the Municipality. Also, it was delimited to only teachers teaching the four (4) core subjects (Mathematics, English Language, Integrated Science and Social studies).

More so, the study focused on only teacher-made tests/classroom achievement tests. Finally. the study focused on only three aspects of classroom assessment; construction, administration and scoring.

Limitations

A questionnaire was used for the data collection. Therefore, the possibility of respondents providing responses to some of the questions, perhaps, without correct understanding of the questions was high. Hence, the tendency of introducing errors into the findings of the study. Another limitation of the study was the tendency of respondents giving socially desirable responses to the questions on the questionnaire, and that therefore, could affect the results of the study as well as the interpretations and uses therein.

Definition of Key Terms

For the purpose of this study, certain terms used are explained below:

Assessment: A process of gathering evidence of what a student can do, and provide feedback on a student's learning to encourage further development.

Achievement tests: They are generally teacher-made tests

Continuum: It is a continuous sequence in which adjacent e lement are not perceptibly different from each other, but the extremes are quite distinct.

Perception*:* Views or opinions held by an individual resulting from experience and external factors acting on the individual.

Organization of the Study

The study was organized into five chapters. Chapter one consists of an introduction to the study; the background of the study, statement of the problem, the purpose of the study and objectives of the study. In addition, the research questions, significance of the study, delimitation, limitations, definition of terms as pertains to the study as well as organization of the study, are described. Chapter two dealt with the review of related literature to the study from documents published and unpublished, including books, journals, newspapers, the internet and other materials that were relevant to the study.

Chapter Three dealt with the research methods used in the study. Contents of this chapter include the research design, study area, population, sampling procedure, data collection instruments, data collection procedure as well as the data processing and analysis plan. Chapter Four focused on the results of the study and discussions. Chapter five dealt with the summary, conclusions drawn from the study, recommendations and suggestions for further research studies.

CHAPTER TWO

LITERATURE REVIEW

Introduction

The main drive of the study was to investigate achievement test practices of teachers in Junior High Schools in the Sissala East Municipality. This chapter reviewed at the literature related to the topic. The chapter consists of the conceptual review, theoretical review and empirical review

Theoretical Review

Constructivist Learning Theory

Constructivist learning theory says that all knowledge is constructed from a base of prior knowledge (Davis, 1991). According to Vigosky (cited in Davis 1991), children are not blank slate and knowledge cannot be imparted without the child making sense of it according to their current conceptions; therefore, children learn best when they are allowed to construct a personal understanding based on experiencing things and reflecting on those experiences. Davis (1991) again states that learners are the makers of meaning and knowledge and constructivist teaching fosters critical thinking, and creates motivated and independent learners. This theoretical framework holds that learning always builds upon knowledge that a student already has; this prior knowledge is called a schema (Davis, 1991). He then explains that because all learning is filtered through pre-existing schemata, constructivists suggest that learning is more effective when a student is actively engaged in the learning process rather than attempting to receive knowledge passively.

James and Pedder (2006) also state that the focus of constructivists is on how people construct meaning and make sense of the world through organizing structures, concepts and principles in schema (mental models). According to James and Pedder (2006), prior knowledge is regarded as a powerful determinant of a pupil's capacity to learn new material. He then indicates that cognitive constructivists emphasize 'understanding,' thus problem solving is seen as the context for knowledge construction. Davis (1991), again argues that processing strategies, such as deductive reasoning from principles and inductive reasoning from evidence, are important and as a result, differences between experts and novices are marked by the way in which experts organize knowledge structures and their competence in processing strategies.

Torrance and Pryor (2001), point out that the interaction between teacher-pupil goes further than just finding out whether the pupil has reached the target behaviour, as in behaviourism. Teacher-pupil interaction in a test situation goes beyond the communication of test results, the judgments of progress and the provision of additional instruction, to include a role for the teacher in assisting the pupil to comprehend and engage with new ideas and problems (Torrance & Pryor, 2001). To them, the process of assessment itself is seen as having an impact on the pupil, as well as the product or the result.

Harlen (2006) stated that the constructivists' view of learning focuses attention on the processes of learning and the role of learners. Teachers engage pupils in self-assessment and use their own assessment to try to identify their current understanding and levels of skills.

Constructivists' assessment

Traditionally, assessment in the classrooms is based on testing thus it is important for the student to produce the correct answers (Davis, 1991). However, he further posits that in constructivist teaching, the process of gaining knowledge is viewed as being just as important as the product. Thus, assessment is based not only on tests, but also on observation of the student, the student's work, and the student's points of view (Davis, 1991). According to Davis (1991), some constructivists' assessment strategies include:

1. Oral discussions. The teacher presents students with a "focus" question and allows an open discussion on the topic.
2. What we know, what we want to know, what we have learned, how we know it (KWL-H) Chart. This technique can be used throughout the course of study for a particular topic, but is also a good assessment technique as it shows the teacher the progress of the student throughout the course of study.
3. Mind Mapping. In this activity, students list and categorize the concepts and ideas relating to a topic.

Examples of Constructivist Activities

The constructivist classroom, students work primarily in groups and learning and knowledge are interactive and dynamic (Harlen, 2006). Davis (1991) states that with the constructivist classroom, there is a great focus and emphasis on social and communication skills, as well as collaboration and exchange of ideas which is contrary to the traditional classroom in which students work primarily alone, learning is achieved through repetition. He further argues that the subjects are strictly adhered to and are guided by a textbook. According to Gielen, Dochy and Dierick (2003), some activities encouraged in constructivist classrooms are:

1. Experimentation: Students individually perform an experiment and then come together as a class to discuss the results.
2. Research projects: Students research a topic and can present their findings to the class.
3. Field trips. This allows students to put the concepts and ideas discussed in class in a real-world context. Field trips would often be followed by class discussions.
4. Films. These provide visual context and thus bring another sense into the learning experience.
5. Class discussions. This technique is used in all of the methods described above. It is one of the most important distinctions of constructivist teaching methods.

Classical True Score Theory

A test theory or test model is a symbolic representation of the factors influencing observed test scores and is described by its assumption. Classical true score theory is a simple, model that describes how errors of measurement can influence observed score. Classical true score theory states that an observed score (X) is equal to the sum of a true score, or true underlying ability (T), and the measurement error (E) associated with estimating observed scores, or $X = T + E$. It is believed that when students take a particular test measuring a construct twice in a succession, it is unlikely that their scores will be identical. This is due to the effect of some factors such as, fatigue, guessing, careless marking, or miss scoring. A different form of test would also result in a change in scores because of variation in content. These inconsistencies in individual scores due to the sampling of tasks or occasions must be regarded as measurement error, (Crocker & Algina, 2008). According to Crocker and Algina, the "True Score" can be interpreted as the average of the observed scores obtained over an infinite number of repeated testing of the same test. In the classroom setting, the "true score" is the score a teacher would obtain if he is to take the average score from an infinite number of test administrations. Of course, in practice, one cannot administer a test an infinite number of times, and as noted previously, the vast majority of the time we get only one chance. Therefore, we use reliability coefficients to estimate both true and error variance associated with our observed test scores (Crocker & Algina, 2008).

Several assumptions are made about the relationship among these three components (True Score, Observed Score and Error Score). Most of the standard procedures for creating and evaluating classroom teacher made test are based on a set of assumptions on the Classical true-score theory. The model assumes certain conditions to be true; if these assumptions are reasonable, then the conclusions derived from the model are reasonable. However, if the conditions are not reasonable, then the use of the model leads to faulty conclusions.

Assumptions of the Classical True Score Theory

1. X=T+E states that, the observed score "X" is the sum of the True score "T" and the error of measurement "E"

2. ε(X)=T. This states that the expected value (population mean “ε”) of “X” is “T”. This assumption is the definition of T: T is the mean of the theoretical distribution of X scores that will be found in repeated independent testing of the same person with the same test
3. ρ_{ET} = o. This assumption implies that examinees with high true score do not have systematically more positive or negative error of measurement than examinee with low true score. This assumption will be violated if for example, one administration of a college entrance exams, students with low true scores copied answers from those with high true scores; this situation will create a negative correlation between true score and error score
4. ρ_{E1E2}= O, where E_1 is the error score for Test 1and E_2 is the error score for Test2.This assumption states that, the error scores of two different tests are uncorrelated. That is if a person has a positive error score in Test 1, he or she is not more likely to have a positive or negative error score in test 2. This assumption is not reasonable if the test scores are greatly affected by factors such as fatigue, practice effect, the examinee’s mood, or effects of the environment
5. ρ_{E1T2}= O; this assumption states that, the error scores on one test (E_1) are uncorrelated with the true scores on another test (T_2). This assumption would be violated if Test 2 measures personality trait or ability dimension that influences error on Test 1. The assumption would also be violated if students with low true scores copied answers from those with high true scores

6. If two tests have observed score X and another X' that satisfies assumption 1 through 5, and if, for every population of examinees T = T' and varies of $\sigma\frac{2}{E}$ = $\sigma\frac{2\prime}{E}$ then the test are called parallel test. For ,$\sigma\frac{2}{E}$ equal to $\sigma\frac{2\prime}{E}$ the condition leading to error of measurement, such as mood, and environmental effect, must vary in the same way for the two tests.
7. If two tests have observed scores X_1 and X_2 that satisfies assumption 1 through 5, and if, for every population of examinees, $T_1 = T_2 + C_{12}$, where C is a constant, then the test are called τ -equivalent test (Allen, & Yen,2012, pp. 56-59).

The implication of this theory therefore, means that in order to achieve the reliability and validity of classroom teacher made test, the principles of the theory needs to apply. The premise of the theory rest on validity and reliability

Conceptual Review

The Concept of Classroom Achievement Tests

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 and Dorr-Bremme and Stiggins and Bridgeford (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 (Ebel & Frisbie, 1991; Etsey, 2004; Gronlund, 1988; Kubiszyn & Borich, 1984; 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 contribute 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. The teacher-made test classification that is most popular with testing experts is the classification based on the type of item format used, which classifies tests into objective-type tests and the essay-type tests (Cunningham, 2001; Etsey, 2004; Gronlund, 2012; Nunnally, 1964; Tamakloe et al, 1986). The aforementioned testing experts have contended that essay-type tests can either be the extended or the restricted response types while objective-type tests can take the form of the short-answer, true-false, matching or multiple-choice.

Testing in educational institutions is designed to assess either curriculum based (classroom instructional) achievement or a variety of student 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, assessing almost any student performance deriving or related to the classroom curriculum, including achievement testing could be an example of curriculum-based assessment (CBA). It must be emphasized that achievement testing is concerned with assessing students based on the domain of content areas they have studied, which are drawn 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). An extensive review of the literature posits two main types of achievement tests. These are teacher-made tests and external tests (Nitko, 2001). Assessment made by teachers of students" attainment, knowledge and understanding is called variously as teacher-made tests. Teachers construct these tests to assess the amount of learning done by students (Amedahe, 1989).

External tests or "extra-classroom assessments" (Nitko, 2001, p. 43), on the other hand, include assessment instruments that are developed and/or graded by people who are not associated with the schools providing the students' learning (Lissitz & Schafer, 2002). Commercial test publishers, departments of education, and local school jurisdictions, usually develop external test (Reeves, 2003). According to the National Association of School Psychologists (NASP, 2002), external tests are usually mandated by core components of standard based reform, which includes (1) content and performance standards set for all students, (2) development of tools to measure the progress of all students toward the standards, and (3) accountability systems that require continuous improvement of student achievement. External test can take the form of textbook accompaniments, survey tests and mandated tests (Munson & Parton, 2013; Nitko, 2001; Zucker, 2004).

Construction of Classroom Achievement Tests

The basic principles for the construction of teacher-made tests have been developed over the years by a number of educational measurement experts (Amedahe, 1989). While some of the test construction 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 available literature, the test construction principles that the researcher judged as most comprehensive and practicable in the classroom testing situation were those postulated by Tamakloe, Atta and Amedahe (1996) and Etsey (2004). These are in eight steps. The steps are:

a) define the purpose of the test,

b) determine the item format to use,

c) determine what is to be tested,

d) write the individual items,

e) review the items,

f) prepare the scoring key,

g) write directions, and

h) evaluate the test.

According to Gronlund (1988), the key to effective achievement testing is careful planning. It is during the planning stage that the purpose of the test must be determined. As already pointed out in the literature, tests can be used for a number of purposes. It is worthy of note, however, that each type of test use typically requires some modification of the test design and thereby determines the type of item format to be used.

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 in classroom achievement testing are the essay- and the objective-types. According to Etsey (2004), it is sometimes necessary to use more than one item format in a single test. This is because depending on the purpose of the test, one item format cannot be used exclusively to measure all learning outcomes. According to Mehrens and Lehmann (1991), the choice of an appropriate item format depends on factors such as the purpose of the test, the time available to prepare and score the test, the number of students to be tested, the skills to be tested, the difficulty level desired, the physical facilities available for reproducing the test, the age of the students and the teacher‘s skill in writing the different types of items.

The final step of the planning stage is the determination of what is to be tested or measured. According to Etsey (2004), the teacher at this point should determine the chapters or units of the course content that the test should cover as well as the knowledge, skills or attitudes to be measured. Instructional objectives need to be defined in terms of student behaviours and linked to what has been stressed in class. A test plan made up of a table of specifications should be made. The table of specifications matches the course content with the instructional objectives (Etsey, 2004). With the total number of items on the test in mind, the specification table helps to avoid overlapping in the construction of the test items, helps to determine the weighting of learning outcomes with respect to content areas, and makes sure that justice is done to all aspects of the course, thereby helping to ensure the content validity of the test.

After the planning stage, 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 continually referred to as the items are written.
2. The test items must be related to and match the instructional objectives.
3. Well-defined items that are not vague and ambiguous must be formulated. Grammar and spelling errors must be checked. Textbook or stereotyped language must be avoided.
4. Excessive verbiage and complex sentences must be avoided.
5. The test items must be based on information that students should know.
6. More items than are actually needed in the test must be prepared in the initial draft. Mehrens and Lehmann (1991) suggested that the initial

number of items should be 25% more while Hanna (as cited in Amedahe, 1989) has suggested 10% more items than are actually needed in the test.

7. Items of varying levels of difficulty must be used. This, however, depends on the purpose of the test.
8. The items and the scoring keys must be written as early as possible after the material has been taught.
9. The test items must be written in advance (at least two weeks) of the testing date to permit reviews and editing.

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) has suggested that the items must be critically examined at least a week after writing them. He has emphasised that where possible, fellow teachers or colleagues in the same subject area should review the test items. Reviewing and editing the items are for the purpose of removing or rewording poorly constructed items, checking difficulty level of items, checking the length of the test, and the discrimination level of the items (items must discriminate between low- and high-achievers). All test items should be checked for technical errors and irrelevant clues.

After reviews and editing, the test items can now be assembled. In assembling test items, the following points must be considered (Etsey, 2004; Kubiszyn & Borich, 1984; 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. According to Hambleton and Traub (cited in Mehrens & Lehmann, 1991), ordering items in ascending order of difficulty leads to better performance than either a random or hard-to-easy ordering. Lafitte (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 (cited in Mehrens & Lehmann, 1991), Mehrens and Lehmann (1991) and other testing experts have recommended that for lengthy or timed tests, items should progress from the easy to the difficult-if for no other reason than to instill confidence in the examinee, especially at the beginning.

It should be noted however, that, the use of statistical item difficulty or item difficulty indexes by the classroom teacher seems impracticable to a large extent (Kubiszyn & Borich, 1984; 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, Gronlund (2012) and Etsey (2004) have recommended that for objective-type tests, the options must be written vertically below the stem rather than across the page. Further, Etsey (2004) has suggested that test items can also be arranged according to the order in which they were taught in class or the order in which the content 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). The marking scheme according to Etsey (2004) and Amedahe and Gyimah (2003), must be prepared when the items are still fresh in the teacher‘s mind and always before the administration of the test. 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) have pointed out that if the teacher considers it prudent to have differential weighting for different essay questions, then factors such as the time needed to respond, the complexity of the question, and emphasis placed on that content area during the instructional phase must 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. As argued by Nunnally (1964), and Ebel and Frisbie (1991), a good working rule is to try to set a time limit such that about 90 percent of the students will feel that they have enough time to complete the test. Directions according to Etsey (2004), must include penalties for undesirable writings, number of items to respond to, where and how the answer should be written, credits for orderly presentation of material (where necessary), and mode 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. Clarity refers to how simply and clearly the items are written vis-à-vis the ability level of the testees and the material the test is measuring. It also refers to the kinds of knowledge the test is measuring and how adequately the test items relate to the content and course objectives (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996).

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 as a result of the possible uses to which the test scores are put (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.

History of Testing and its Development

The historical development and up-bringing of testing in Africa and in Europe has been interwoven with the development of psychology as a scientific discipline. Test theory evolved from testing to three major areas of development: civil-service examination, school examination, and the study of individual differences. Civil service testing began in China about 3000 years ago when an Emperor decided to assess the competency of his officials. Later, government positions were filled by persons who scored well on examinations that covered topics such as music, horsemanship, civil law, writing, etc. Such examinations were eliminated in 1905 and were replaced by formal educational requirements. Paradoxically, as the Chinese were phasing out their examinations, civil-service exams were being the efforts of psychologists in Europe and in Africa. Du Bois (1970) attributed the increase in the use of tests in Britain and the United States as a fair way of selecting among job applicants for government jobs. Early evidence of the effectiveness of the examination was anecdotal in nature, but the examinations were popular because they removed decisions from the biases of political judgments.

Students in European schools were giving civic examination until well after the 20th century, when paper began replacing parchment and papyrus. In the 16th century the Jesuits started using tests for the evaluation and placement of their students. In France, Binet (1905) developed the first individual tests of intelligence as part of his work on the study of individual differences. A German, William (1928), developed the intelligent quotient (IQ), which he defined as the ratio of mental (measured) age to chronological (actual) age. Charles Spearman a British, followed the footsteps of Galton and Pearson, and his work led to the modern concepts of test reliability and factor analysis. Most of early tests were designed for administration to only one individual at a time. Although work had begun on tests that could be given to many examinees at once, group-administered tests did not become widely used or accepted until after their introduction by the United States Army in World War 1.

In the West, in England, civil service ability testing was adopted during the middle portion of the 19th century (Cunningham, 2001; Flanagan et al., 1997). Cunningham (2001) continued by noting that the Chinese method of selecting government employees was used as a basis for the establishment of the Indian civil service. He concluded that the first British civil service commission was set up in 1850. In the USA, testing began in the later part of the 19th century DuBois cited in (Cunningham, 2001; Flanagan et al., 1997). Dubois pointed out that following the successful use in England of the Chinese method of selecting government employees, the method was adopted in the USA. He pointed out that the first civil service was established in 1883. Formal testing in schools (paper and pencil tests) began with the introduction of paper in the 12th century Dubois (as cited in Cunningham, 2001).

According to Cunningham (2001), assessment by means of written tests was first used by the Jesuits at St Ignatio. He noted that the development of academic tests was pioneered in Britain, particularly in the University of London. Under its initial charter, testing and awarding of degrees were recognised as a legitimate basis for decision making. It is worth noting however, that, prior to this period, academic testing (oral testing) in USA schools had already begun. As stated by DuBois (cited in Anastasi, 1982), among the ancient Greeks, testing was an established adjunct to the educational process where tests were used to assess the physical as well as intellectual skills. Anastasi (1982) pointed out that the Socratic method of teaching with its interweaving of testing and teaching has much in common with today’s programmed learning.

On the account of Ebel (as cited in Amedahe, 1989) and Anastasi (1982), from their beginnings in the Middle Ages, European universities relied on formal examinations in awarding degrees and honours. These examinations, however, were largely oral. Test development, like many other aspects within psychology and education, is a product of many contributors and disciplines throughout history.

Importance of Testing

Educational Importance

Educational uses of tests have been classified under instructional management decisions, selection decisions, classification decisions, placement decisions, counselling and guidance decisions, and credentialing and certification decisions (Nitko, 2001; Amedahe & Gyimah, 2003). The instructional management decisions refer to all the classroom decisions taken by the teacher on the basis of the assessment results of students. Firstly, tests provide useful information for instructional diagnosis and remediation. The classroom teacher constantly needs to diagnose his instruction and remediate the aspects which have been defective (Amedahe & Gyimah, 2003). This is made possible through feedback from students to the teacher. In instructional diagnosis and remediation, the teacher engages in diagnostic testing to identify which students need remedial help or special attention. According to Nitko (2001), diagnosis involves identifying both the appropriate content and the features of the learning activities in which a student should be engaged to attain the learning target.

Tests are used in the modelling of learning targets. According to Nitko (2001), “assessments define for students what the teacher wants them to learn”. (p. 9). He continued by noting that students can always compare their current performance on the learning targets with the desired performance. The teacher can then teach his students to detect the ways in which their performance is matching the criterion and the ways in which it is deficient. In this way, the teacher can direct his teaching on the remediation of any identified deficiency and students are also able to know what is important to learn once they are able to evaluate their own performance vis-à-vis the desired learning targets.

Tests are needed for the provision of motivation for students, rewarding those who have prepared well in advance and providing negative consequences for those who have not prepared well. The frequency of an individual behaviour is increased by reinforcement. Hence, it can be reasonably concluded that tests cause students to study more in the sense that the motivation derived from tests as a result of performing well can activate and direct their learning by sustaining their interest (Cunningham, 2001; Ebel & Frisbie, 1991; Gronlund, 2008; Nitko, 2001).

Tests are used for the assignment of grades to students. The grades or symbols (A, B, C) that the classroom teacher reports, represent his /her formal evaluation or judgement of the quality or worth of his/her students ‘achievement of the important learning objectives (Amedahe & Gyimah, 2003; AERA/APA/NCME, 2014; Nitko, 2001). It is worth noting that assessment results of which tests constitute the most important part as it is in the Ghanaian educational system provide the basis for the assignment of grades. AERA/APA/NCME (2014) have cautioned here that to serve effectively the purpose of stimulating, directing and rewarding students ‘effort to learn, grades must be valid. To achieve this, the highest grades must go to those students who have demonstrated the highest level of achievement with respect to the course objectives.

On the issue of selection decisions, sometimes, an institution decides whether some persons are acceptable for specific programmes while others are not. Those not acceptable are rejected and are no longer the concern of the institution (Amedahe & Gyimah, 2003; Cronbach, 1960; Nitko, 2001). An educational institution often uses test results to provide part of the information on which selection decisions are based. Typical examples are the selection of candidates for admission into Senior High Schools (SHS) in Ghana which is based on the test scores of students at the end of the Junior High School and university admissions in Ghana which are based on the test scores of students at the end of the SHS.

Tests provide the basis for the grouping of children with reference to their ability to profit from different types of school instruction and the identification of the intellectually retarded and the gifted (Cunningham, 2001). Nitko (2001) has pointed out that sometimes, based on test results, a decision is made that result in a person being assigned to one of several different but unordered categories of programmes. According to Cronbach and Glaser (as cited in Nitko, 2001), these types of decisions are called classification decisions. These decisions result in either assigning students in the same classroom to different groups for effective instruction or assigning students to special education classes. Cunningham (2001) however cautioned test users about the over reliance on test results in assigning students to special education classes by pointing out that intelligence tests are only one component of the assessment of students referred for possible placement in special classes.

On the issue of placement decisions, Cronbach (1960); Kubiszyn and Borich (1984) and Nitko (2001) have pointed out that placement decisions are made after an individual has been accepted into an educational programme. Cronbach et al., (2001), continued by noting that placement decisions basically involve using assessment results or test data to determine where in a programme an individual is best suited to begin work. Such decisions are characterised by assigning individuals to different levels of the same general type of instruction or education based on their ability, with no one rejected by the institution Cronbach and Glaser cited in (Nitko, 2001). Promotion in Ghanaian schools from one class or form to another which in most cases is based on the performance in tests of the previous class is an example of a placement decision.

Counselling and guidance decisions involve using assessment results, with test data inclusive; to help students in exploring and choosing careers and in directing them to prepare for the careers they select (Anastasi, 1982; Amedahe & Gyimah, 2003; Kubiszyn & Borich, 1984; Nitko, 2001). Amedahe and Gyimah (2003) have explained that guidance is one of the students' personnel services provided in a non-instructional setting to cater for the needs of students including educational, emotional, and moral and adjustment needs. Nitko (2001) and, Amedahe and Gyimah (2003) have agreed with the fact and argued that due to the complexities involved in guidance and counselling decisions, test data must always be combined with other assessments such as interviews, interest inventories, various aptitude tests and personality questionnaire together with additional background information on students and discussed with students in a series of counselling sessions in order to help students make good decisions.

On credentialing and certification decisions, Nitko (2001) and Amedahe and Gyimah (2003) explained that they are concerned with assuring that a student has attained a certain standard of learning. Credentialing and certification may be mandated by state legislation as in the USA and executed by an external examining body at the state level. In Ghana, certification and credentialing of students is done by the WAEC. With the introduction of the practice of continuous assessment as a result of the educational reforms in 1987, Ghanaian classroom teachers contribute 30% of the total marks for certification of students at the JHS and SHS levels (Amedahe, 2000; Pecku, 2000).

Non-Educational uses of Tests

George (2002) noted that one of the first problems that stimulated the development of psychological tests was the identification of the mentally retarded. Over the centuries the uses of tests have been quite diverse with various non-educational applications. George (2002) again pointed out that non-educational uses of tests include clinical applications in the area of the examination of the emotionally disturbed, the delinquent and other types of behaviour deviants. According to Gielen, Dochy and Dierick (2003), clinical uses of tests are mainly found in the diagnosis and classification of mental patients to determine the type of treatment suitable for them.

The selection and classification of industrial personnel represent another major non-educational application of tests (George, 2002). Gielen, Dochy and Dierick (2003), claimed that from the assembly-line operator to top management, tests have proved helpful in such matters as hiring, job assignment, transfer, promotion or termination. According to Cronbach (1960) and Anastasi (1982), testing constitutes an important part of the total personnel programme. A typical example is the application of psychological testing in the selection and classification of military personnel worldwide. George (2002) argued that from simple beginnings in World War I, the scope and variety of psychological tests employed in military circumstances underwent a phenomenal increase during World War II. Jackson and Davis (2000), however, asserted that where people are assigned to different levels of work, rather than to distinctly different types of work, the decision becomes a placement decision. This is exemplified in a case of choosing officer candidates from among enlisted men where men, not chosen as officers, remain in the army and are assigned other duties. This is a placement decision.

Types of Classroom Teacher-Made Tests

Assessments made by tutors of student's attainment, knowledge and understanding is called variously as teacher-made or classroom made test and school-based assessment (Amedahe, 1989). The rationale of teacher-made tests is linked with the constructivist model of learning. In this model, it is important to understand what the student knows and how he/she articulates it in order to develop his/her knowledge of understanding. In this model, it is learning with understanding which counts and to this end, information about existing ideas and skills is essential. Work in psychology and learning portrays similarly that for effective learning, the task must be matched to the student's current level of understanding Gipps (1992), and either pitched at the level to provide practice or slightly higher in order to extend and develop the student's skills. For content of a course to be adequate and ensure that it is relevant as well, the content should match the understanding level of a particular student. Salvia and Yesseldyke (2001) asserted that, teacher made tests are better when used to evaluate students because they are curriculum matched. If the new task is much too easy, the students can become bored, and if much too difficult, the student can become de-motivated (Gipps, 1999).

Essentially, there are two main forms of teacher or classroom-made test; formal and informal tests. Tutors may pose questions, observe activities, and evaluate students' work in a planned and systematic or ad hoc way (Gipps, McCallum, McAlister & Brown, 1995). Classroom tests are basically teacher-made tests. Teachers have the responsibility to provide their students with the best instruction possible. This implies that they must have some relevant content procedures or method whereby they can reliably and validly evaluate how effectively their students have learnt what has been taught them (Mehrens & Lehmann, 2009). The pencil and paper or teacher-made test is one such tool. Classroom teacher-made tests mostly prevail in subjects –matter like Science and Social Studies. Classroom tests can also, be tailored to fit a teachers' particular instructional objectives, essentially, when one wishes to provide for optimal learning on the part of the pupil and optimal teaching on the part of the teacher (Bejar, 1984). Here, without classroom tests, the objectives that are unique to a particular school or teacher might not be evaluated. The emphasis on the desirability and importance of the classroom teachers being able to construct their own personal, unique and relevant tests is based on the principles of assessment in education.

A survey conducted by Stiggins and Bridgeford (1985) on the uses of various types of tests reported that the tests are

1. For assigning grades and evaluating the effectiveness of an instructional treatment.
2. For diagnosis
3. For remedial teaching
4. To motivate students to learn to improve in their work
5. To provide the basis for guidance in selection and placement in the world.
6. For certification.

Despite the aforementioned importance of teacher-made tests, a study conducted in the United States of America revealed some deficiencies in teacher-made tests, in the sense that, teachers were only trained to teach but not to assess their students (Gullickson, 2001).

To begin with, ambiguous questions is when a statement or word have two or more meanings, one has ambiguity. For example, in essay tests, words such as discuss or explain may be ambiguous in that different pupils may interpret these words differently.

Again, excessive wording contributes to difficulty in teacher-made test. Too often teachers think that the more wording there is in a question, the clearer it will be to the student. This does not always happen. The more precise and clear-cut the wording, the greater the probability that the student will not be disorganised. Mostly, teacher-made tests do not cover the objectives stressed and taught by the teacher and do not reflect proportionally the teacher's judgement as to the importance of those objectives. Teacher-made achievement tests are mostly heavily loaded with items that only test the students' ability to recall specific facts and information (Fleming & Chambers, 1983).

Use of inappropriate item formats also contributes to deficiency in teacher-made tests. Some teacher uses different item formats like true-false or essay solely because they feel that change or diversity is desirable. But the need for diversity should not govern the type of item to be used therefore; teachers should be selective and choose the format that is most effective for measuring a particular objective.

According to Nitko (2001), assessment content is relevant when teacher-made or classroom test comprises choice formats such as (multiple choice, true or false, matching exercise and other formats like greater - less same items), short answers and completion format and essay format (restricted responses and extended responses). Some educators argue that essay tests are more susceptible in scoring than the objective tests. However, classroom teachers exclusively use both since one cannot be used exclusively to measure all learning outcomes. According to Bartels (2003), with regard to the objective type tests, the multiple choice, short-answer/fill-in-the blanks, matching and true or false types are the major ones used by tutors in the teacher colleges of education in Ghana.

Objective -Type Tests

The objective-type item was developed in response to the criticism levelled against the essay type tests. Some of the criticisms were, poor content sampling, unreliable scoring, time-consuming to grade, and encouragement of bluffing. The objective test-items normally consist of a large number of items and the responses are scored objectively, to the extent that competent observers can agree on how responses should be scored (Amedahe & Etsey, 2003).

Objective-type item formats are put into two groups; the supply type and the selection type. The supply type format consists of completion type, fill-in-the blanks and short answer. The selection type consists of true-false, matching, and multiple-choice item type. According to Amedahe and Etsey (2003) objective type test items are most useful when class sizes are very large and when there is limited time to submit the results of the test. The short-answer and completion format consist of one or more blanks in which the student writes his answers to the question with a word or, phrase. This type of objective test is also known as constructed – response type. It consists of a statement or question and the respondent is required to complete it with a short answer usually not more than one line (Etsey, 2012). It is used for testing knowledge of facts or recall of specific facts (example, "knowledge objective" in Bloom's taxonomy of educational objectives). Short-answer and completion format can be used to assess higher-level abilities like, to make simple interpretations of data and applications of rules, to solve numerical problems in science and mathematics, and to manipulate mathematical symbols and balance mathematical and chemical equations.

A true or false test consists of a statement to be marked true or false. Here their utilities are placed primarily in assessing knowledge of factual information. True or false items are difficult to prepare (Salvia & Ysseldyke, 2001). True or false test items are made up of four types; simple true or false, (here only two choices; true or false), complex true or false (comprises three choices; true or false and opinion), compound true or false (consists of two choices, true or false plus a conditional completion response) and finally multiple true or false (consist of a stem with three, four or five options and the respondent indicates if the options are true or false (Etsey, 2012).

One of the limitations in constructing the true or false test items is that, the probability of getting right answer by guessing is high. It can be used to assess only a few numbers of educational objectives, and can be used to evaluate definitions, facts, meaning of the true or false, recognition, and interpretation of charts/graphs. An advantage of true or false test item is that, they can cover a wide range of content within a relatively short period of time.

Matching test format is another choice format item which presents respondents with three things; (a) Directions for matching (b) A list of premises (c) A list of responses. The simple matching exercise requires simple matching based on association that a student must remember. This is basically done to assess respondents' comprehension of concepts and principles. One of advantages of matching test format is that, matching test format use pictorial materials to assess student's abilities to match words and phrases with pictures of objects or with locations on maps and diagrams.

A multiple-choice item consists of a stem followed by a list of two or more proposed alternatives; here the respondents are expected to select the correct option from the alternatives. Normally, only one of the options is the correct or best answer to the question one poses. This is called the keyed alternative, keyed answer or basically the key whiles the remaining incorrect options are called foils or distractors. The purpose is to allow students to demonstrate their knowledge and understanding of the learning targets. There are three types of multiple-choice tests. These are the single correct type and the "multiple responses" type. The "single correct" type consists of a stem followed by three or more responses and the respondent is to select only one option to complete the stem. The "multiple responses" type consists of a stem followed by several true or false statements or words. The respondent is to select which statement could complete the stem. Multiple-choice tests format does not require students to write out and elaborate their answers and minimize the opportunity for less knowledgeable students to "bluff" or "dress-up" their answers (Wood, 2007).

According to Etsey (2012), he outlined the following the Strengths and weaknesses of objectives items

Strengths

1. Scoring is easy and objective
2. They allow an extensive coverage of subject content.
3. They do not provide opportunities for bluffing.
4. They are best suited for measuring lower-level behaviours like knowledge and comprehension.
5. They provide economy of time in scoring
6. Student writing is minimized. Premium is not placed on writing.
7. They are amenable to item and statistical analysis
8. Scores are not affected by extraneous factors such as the likes and dislikes of the scorer.

Weaknesses

1. They are relatively difficult to construct.
2. Item writing is time consuming.

3. They are susceptible to guessing.
4. Higher-order mental processes like analysis, synthesis and evaluation are difficult to measure.

Essay-Type Tests

According to Amedahe and Etsey (2003), essay test items consist of relatively few items, but each require an extended response. Essay test items provide respondents with the freedom to organize their own ideas and respond with limited restriction. Here respondents are asked to speak to a particular issue and for that reason they could not just write a single word as an answer than to express themselves in terms of what they know about the items. The ability of the respondents to express themselves clearly and fluently and with content required tells the instructor that they have actually mastered the content of the subject. Essay questions are most useful in assessing instructional objectives prepared at a comprehension level or higher order thinking (Salvia & Yesseldyke, 2001). Nitco (2001) noted that "what is perhaps unique about the essay format is that it offers students opportunity to display their abilities to write about, to organize, to express and to explain interrelationship among ideas" (p.187).

The essay test has two major types; extended and restricted response depending on the amount of scope or freedom given the student to organize ideas and write answers. Extended-response type of essay questions has no bounds placed on the student as to the point(s) to discuss and the type of organization to use. This type of question permits the student to demonstrate the ability to:

1. call on factual knowledge

2. evaluate factual knowledge

3. organize ideas

4. present ideas in a logical, coherent written fashion

The extended response makes the greatest contributions at the levels of synthesis and evaluation of writing skills (style, quality).

Under the restricted-response essay questions, the student is more limited in the form and scope of the answer because it tells specifically the context that the answer is to take. This type of question is of greatest value for measuring learning outcomes at the comprehension, application, and analysis level, and its use is best reserved for these purposes.

According to Etsey (2012), he outlined the following the Strengths and weaknesses of Essay Test items

Strengths

1. They provide the respondent with freedom to organize his own ideas and respond within unrestricted limits.
2. They are easy to prepare.
3. They eliminate guessing on the part of the respondents.
4. Skills such as the ability to organize material and ability to write and arrive at conclusions are improved.
5. They encourage good study habits as respondents learn materials in wholes.
6. They are best suited for testing higher-order behaviours and mental processes such as analysis, synthesis and evaluation
7. Little time is required to write the test Items.
8. They are practical for testing a small number of students.

Weaknesses

1. They are difficult to score objectively. Starch and Elliott (1912, 1913) reported that inter-rater variability could be as high as 68.
2. They provide opportunities for bluffing where students write irrelevant and unnecessary material.
3. Limited aspects of student's knowledge are measured as students respond to few items only.
4. The items are an inadequate sample of subject content. Several content areas are omitted.
5. A premium is placed on writing. Students who write faster, all things being equal are expected to score higher marks.
6. They are time-consuming to both the teacher who scores the responses and the student who writes the responses.
7. They are susceptible to the halo effect where the scoring is influenced by extraneous factors such as the relationship between scorer and respondent.
8. A critical reader as well as a competent scorer can only effectively score responses.

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). 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 student's results must be supported with many evidences. The results must be devoid of errors and therefore, the soundness of the results. The focus here is not necessarily on scores or items, but rather interpretations made from the instrument. That is, the behavioural interpretations that one can deduct from test scores is of paramount concern. “In order to be valid, the inferences made from scores need to be appropriate, meaningful, and useful” (Gregory, 1992, p. 117).

Validity is an integrated evaluative judgment on the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment (Messick, 2003). The validity of classroom assessment depends on, analysing the intended learning and all its embedded elements, having a good match among the assessment approaches, the intended learning, and the decisions that teachers and learners make about learning, ensuring that the assessment adequately covers the targeted learning outcomes including content, thinking processes, skills and attitudes (Northern Canadian Protocol for Collaboration in Education, 2006). Validity ensures the central question; does assessment measure what it purports to measure (Winograd & Perkins, 1997)? There are three deferent types of validity evidence namely, criterion validity, construct validity and content validity.

A measure itself is neither valid nor invalid; rather, the issue of validity concerns the interpretations and uses of a measure’s scores. The interpretations and uses of one’s assessment results are also valid only when the values implied by them are appropriate. Essentially, the interpretations and uses one make of one’s assessment results are also valid when the consequences of these interpretations and uses are consistent with appropriate values. Here, when the values of the assessment are not in accordance with the consequence of the assessment then this principle is violated. A second important implication of the definition of validity is that validity is a matter of degree; it is not an “all-or-none” issue. That is, the validity of a test interpretation should be conceived in terms of strong versus weak instead of simply valid or invalid. For test users, validity should be a deciding factor in their choice of psychological tests. Although such choices are based on a number of practical, theoretical, and psychometric factors, a test should be selected only if there is strong enough evidence supporting the intended interpretation and use. A third important facet of validity is that the validity of a test’s interpretation is based on evidence and theory. For a test user to be confident in an interpretation and use of test scores there must be empirical evidence supporting the interpretation and use. In addition, contemporary views on validity emphasize the importance of grounding the interpretation and use of a test in a defensible psychological theory.

Validity Evidence

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

Content – Related Validity

Content- related Validity is often defined as the extent to which the sample of items, tasks, or questions on a test is representative of the domain of content (Moss, 1992). Bollen (1989) defined content validity as a qualitative type of validity where the domain of the concept is made clear and the analyst judges whether the measures fully represent the domain (p.185). But, Wiliam (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). Here, Wiliam is advocating for content-related evidence to extend to include the behaviour elicited actually corresponding to the intentions of the assessment task. He explains with an example, a test claiming to assess students' understanding of forces "would be invalidated if it turned out that the reading requirements of the test were so demanding that students with poor reading ability, but a sound understanding of forces, obtained low marks" (p. 4). On the other hand, if a student possesses an understanding of an issue demanded by a test, but fails to show it for reasons of linguistic difficulty then, the results of that test would be invalid.

Wiliam takes this idea from Ackerman and Smith (1988). Ackerman et al., points out that a test would be considered *biased* and invalid, if it makes different impact on the people who take it because of interfering factors which prevent the appropriate response from being demonstrated. Content-related evidence is therefore, not only demonstrated by the degree to which samples of assessment tasks are representative of some domain of content. It is important for the behaviour elicited by the test item not to have been influenced by factors that conceal the true ability or potential of the student. This could be an argument in support of school-based teacher assessment as the conditions of assessment can be arranged to provide ecological validity; that means relating the assessments as closely as possible to the learning experiences of the student. As Crooks (2001) point out, "the circumstances under which student performances are obtained can have major implications for the validity of the interpretations from an assessment" (p. 270). Issues such as low motivation, assessment anxiety, and inappropriate assessment conditions can all be threats to the valid of students' assessment results.

Content validity is a general property of a test. Test author who defines the content domain and writes items to represent the domain succeeds to some degree in attaining their goal. In addition to content validity, is the face validity which answers the question: "Does the assessment look as if it will mean what it is, supposed to mean?" (William, 1993, p. 5). In other words, it answers the question; does the assessment appear to be measuring the sort of tasks required of a particular subject domain? In the teacher's context, the crucial face validity question would be whether the assessments appear to measure the kind of things expected of teaching. Hoste and Bloomfield (1975) put it in another way: "does the assessment procedure appear to test the aims of the course adequately?". Such questions are important since they have implications for what can be assessed as well as how it should be assessed.

According to Miller, Mclntire and Lovler (2011), 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 body of knowledge or behaviour that a test presents. They further asserted that, the step involves reviewing other instruments that measure the same construct, interviewing experts who are familiar with the construct. The purpose is to ensure that you clearly understand and can clearly define the construct you will be measuring. According to Groth-Marnat (1997), evidence of validity bases on test content requires that the test cover 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 test's content. The specification delineates, the thinking process 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)

Gipps (1994) points out that performance assessment does tend to have good face validity. As Patton (1990) explains this is because performance measurement calls for examinees to demonstrate their capabilities directly, by creating some product or engaging in some activity that relates to the ultimate task. Similarly, Delandshere (1996) has indicated that new teacher assessment methods, such as portfolios, reflective essays and practical tasks, appear to have more face validity than written tests. William (1993) notes that for assessment to command a good measure of confidence among users such as teachers in the senior high schools in Ghana, it is important that it possesses high face validity.

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. When the criterion exists at the same time as the measure, we talk about concurrent validity. Concurrent ability 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 to be scored. When the criterion occurs in the future, we talk about predictive validity. Predictive validity evidence refers to 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 differ 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 - Related 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.) Moss (1992) 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 operationalized and syntactically defined in order to measure it effectively (Benson, 1998; Crocker & Algina, 1986; Gregory, 1992). 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; Gregory, 1992).

If a relationship is causal, what are the particular cause and effect behaviours or constructs involved in the relationship? Construct validity refers to how well you translated or transformed a concept, idea, or behaviour that is a construct into a functioning and operating reality, the operationalization (Trochim, 2006). Here, the study will lay emphasis on how teachers measure the constructs of students when assessing them.

Reliability of Test Items

This is about consistency. Reliability is defined as the degree of consistency between two measures of the same thing (Ebel & Frisbie, 2001). Assessment results would be similar under slightly different measurement conditions to the degree. For instance, if one assesses a student twice, one hope that, he would obtain almost the same score if one assesses the student a day later. Here, if one measures a person's level of achievement, one hope that the scores will be similar under different administrators, using different scorers, with similar but not identical items.

A reliable assessment is one that consistently achieves the same results with the same (or similar) cohort of learners. Gay (2006) defined reliability as "the degree to which a test (or qualitative research data) consistently measures whatever it measures" p. 5. If the assessment process is reliable, the inferences about a learner's learning outcome should be similar when they are measured by different teachers, when learning is assessed using various methods or when learners demonstrate their learning at different times (Northern

Canadian Protocol for Collaboration in Education, 2006). According to William (2008) a reliable test is one in which scores that a learner gets on different occasions or with a slightly different set of questions on the test, or when someone else does the marking, does not change very much. Various factors affect reliability including ambiguous questions, too many options within a question paper, vague marking instructions and poorly trained markers. Decisions are based on data. These data may come from classroom and standardised test scores, classroom observations, parental reports and many other sources. In using the data for decision making, one should know something about the quality of the data. Here, high-quality data should be weighted more deeply in one's decision than poor-quality data. In principle, data should be reliable, and the inferences one draws from the data should be valid. Reliability is paramount in assessing individuals.

In physical measurement, one can ordinarily obtain very reliable measures (Ebel & Frisbie, 2001). This is true mainly for three basic reasons;

1. Physical characteristics can usually be measured directly rather than indirectly.
2. The instruments used to obtain the measures are quite precise.
3. The traits or characteristics being measured are relatively stable

Reliability operates at two levels as follows; that of the individual assessed, and that of a number of assessors (Freeman & Lewis, 2008). Reliable assessors make the same decision on a particular assessment whenever they mark it. When more than one assessor is concerned, reliability is achieved if presented with work of the same standard; all assessors make the same judgment. Reliable assessment ensures accurate and consistent comparisons, whether between the performances of different pupils or between a learner's performance and the criteria for success (Freeman & Lewis, 2008). Maizan (2005) contends that there are three types of reliability that are most relevant to classroom tests. These are internal consistency, inter-scorer and intra-scorer reliability. In the view of Maizan (2005) internal consistency refers to the consistency of objectives among the items of a test while inter-score reliability refers to the consistency between marks given by different teachers. On the other hand, intra-scorer reliability refers to marks given by the same teacher on different occasions. According to Brown (2007), "the major threat to reliability is the lack of consistency of an individual marker" (p. 78). However, intra –ratter reliability might not in fact be a major concern when ratter is supported by rubrics (Jonson & Svingby, 2007). Consistent grading is essential in order to ensure reliability of test scores.

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 textural 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 (2009) outlined the following stages and steps as being important to 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, Adamolekun (2012) 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 (2001), 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, 2001) 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, 2008; Mehrens & Lehmann, 2009). 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 contribute 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 (2001), 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, 2012; 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 is emphasised by the JCSEPT (1999) by stating that, -reasonable effort should be made to assure the integrity of the test scores by eliminating opportunities for test takers to attain scores by fraudulent means (p. 64). This calls for ensuring a congenial psycho-physical atmosphere for test taking (Tamakloe et al., 1996, p. 214). This was also emphasised by Airasian (as 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; Lindquist, 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 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 student‘s 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 effective test administration. This facilitates students' reading of instructions and test items without straining their eyes, thereby working faster (Gronlund, 2012). ―Good ventilation and comfortable temperature should be assured since their absence could create unrest or uneasiness in testees 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-minutes 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 at minimising 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 (2005), in administering test items, classroom teachers are to consider that, the following information are essential maximising students' performance.

1. Students must be made aware of the rules and regulations covering the conduct of the test. Penalties for malpractice such as cheating should be clearly spelt out and clearly adhered to.
2. 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.
3. Avoid giving tests when students would normally be doing something pleasant e.g. having lunch etc.
4. The sitting arrangement must allow enough space so that pupils will not copy each other's work.
5. Adequate ventilation and lighting is expected in the testing room.
6. Provision must be made for extra answer sheets and writing materials.

7. Pupils should start the test promptly and stop on time.

Scoring of Classroom Achievement Tests (Essay 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, 2001). This is known as a marking scheme, a scoring rubric or a scoring key. (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 2006).

The Holistic scoring rubrics require 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 synthesize and create and when no definite correct answer can be prespecified" (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 (2001), 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 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 (2001); 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 ratter 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 be on students ‘weaknesses and strengths in 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 (cited in Mehrens and

Lehmann, 2001) 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 (2001) 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.

Assessment Standards

Assessments depend on professional judgment. "Testing standards, guidelines, and codes of practices are developed by large committees or testing publishers to provide guidance on fairness practices for the broader educational communities" (Xiaomei, 2014, p. 51). Standards, guidelines, and codes of practices identify issues to consider in exercising professional judgment and in striving for the fair and equitable assessment of all students (JCTP, 2004).

However, not all of such documents are useful and relevant to all testing purposes. Gipps and Stobart (2009) noted that fairness considerations in large-scale high-stakes testing might be different from fairness considerations in classroom teacher-made testing. Therefore, for the purposes of usefulness and relevance, I considered only standards, guidelines and codes that pertain to large-scale testing, and these include:

1. The Standards for Educational and Psychological Tests (AERA et al., 1999; 2014), which is geared primarily for test developers, researchers, and psychometricians.
2. Responsibilities of Users of Standardized Test (JCTP, 2000), which provides a concise statement useful in the ethical practice of testing.
3. ETS Standards for Quality and Fairness (ETS, 2014), which helps to design, develop, and deliver technically sound, fair, accessible, and useful products and services.
4. The Principles (Joint Advisory Committee on Testing Practices, 1993), which was developed primarily in response to inappropriate use of large-scale assessment results in Canada.
5. Code of Professional Responsibilities in Educational Measurement (NCME, 1995), which serves as a statement of professional responsibilities for stakeholders in testing.

Newman and Wehlage (1993) noted that achievement tests tasks need to be organized and structured well so that they are contextualized, integrative, related to the curriculum taught, flexible (requires multiple applications of knowledge and skill), open to self-assessment and peer-assessment, contain specified standards and criteria. They again emphasize that authentic assessment task must consider the following standards:

Organization of information: The task asks students to organize, synthesize, interpret, explain, or evaluate complete information in addressing a concept, problem, or issue. Consideration of alternatives: The task asks students to consider alternative solutions, strategies, perspectives, or points of view in addressing a concept, problem, or issue. Disciplinary content: The task asks students to show understanding and/or use of ideas, theories, or perspectives considered central to an academic or professional discipline. Disciplinary process: The task asks students to use methods of inquiry, research, or communication characteristic of an academic or professional.

Teachers' Perceptions

Researchers have attempted to investigate teachers' perceptions of assessment in many different ways (Chester & Quilter, 1998). Chester and Quilter believed that studying teachers' perceptions of authentic assessment is important in the sense that it provides an indication of how different forms of authentic assessment are being used or misused and what could be done to improve the situation. More critical also is the fact that perceptions affect behavior (Atweh, Bleicker & Cooper, 1998; Calderhead, 1996; Cillessen & Lafontana, 2002).

Creswell (2012) engaged 25 teacher-volunteers to participate in a study representing six secondary rural schools from New South Wales, Australia. The researchers used the Structure of Observed Learning Outcome (SOLO); a cognitive structural model which provided "a basis for both assessing students' understanding and identifying ways of enhancing students' learning" (Creswell, 2012, p. 420). Three two-day workshops were conducted at the University for these teachers, focusing around the SOLO model assessment tasks and teaching strategies of the 25 teacher-volunteers by Creswell (2012). The researchers primarily used two sources of data: "students' scripts coded using the SOLO model" and interviews with teachers. They inquired from the teachers their experiences with the new approach to teaching i.e. (SOLO) and assessment practices to enhance students' learning. The researchers found out that all teachers who participated in this project represented a change in their perception enabling them use collaborative effort to engage students' understanding in their classrooms.

According to Creswell (2007), the project helped teachers recognize that restricting the type and style of questions in their teaching and assessment provide limited scope for students to demonstrate their conceptual understanding" (p. 431). Overall, the researchers found out that teachers reported a shift in their perceptions of learning demonstrated in their teaching and assessment practices which was noticed by students and other teachers as well (Creswell, 2012).

Chester and Quilter (1998) in their study on in-service teachers' perceptions of classroom assessment; standardized testing, and alternative assessment methods in Debre Markos University in Ethiopia concluded that teachers' perceptions of classroom assessment affected their classroom assessment practices. They found out that teachers that attached less value to classroom assessment used standardized tests most of the times in their classrooms. Chester and Quilter went further to say that teachers with negative experiences in alternative assessment and standardized testing are least likely to see the value in various forms of assessment for their classroom. They recommended, therefore, that in-service training should focus on helping teachers see the value of other assessment methods rather than "how to" do assessment.

An interview with a fifth-grade teacher at Deerfield Elementary school in Lexington, USA by Kentucky Department of Education (1991) confirms that teachers are aware of the limitations of standardized tests. They further revealed that the teacher indicated that curriculum must emphasize subjects for which the state accountability test measures proficiency: math, reading, social studies and science. The teacher argued further that test scores do not truly reflect her students' abilities and are too vague to help her pinpoint individual needs (Kentucky Department of Education, 1991). The teacher asserted that she longs for an assessment that relies on more than just written problems that could capture the more diverse skills visible in her classroom and valued in the workplace, such as artistic talent, computer survey, and the know-how to diagnose and fix problems with mechanical devices (Kentucky Department of Education, 1991).

Empirical Review

Assessment tasks and strategies

Fox and Soller (2001), in their study on authentic assessment strategies and tools employed by teachers in Malawi found out that students in lower classes prefer working collaboratively using projects, computer-based simulation task, storytelling and demonstrations while students in upper classes also demonstrated high level performance in working competitively using writing samples, performance products, and graphic organizers. It was also revealed in the study that education systems that emphasize tests and examinations put some student at a disadvantage (Mbano, 2003; Nampota & Wella, 1999).

Fook and Sidhu (2010) conducted a study in Malaysia to investigate the different types of authentic assessment used in higher education in Malaysia. The study employed a qualitative research method and involved the use of instruments such as interview, document analysis and classroom observations to collect relevant data in the classroom.

The researchers identified that different types of authentic assessment were used. The study revealed that teachers employed the following assessment tools; portfolio (10%), article review (10%) performance product (20%), project (40%) and test (20%). The findings indicated that alternative and authentic assessment have more acceptances from students and should, therefore, be viewed as an alternative to traditional standardized assessment. The study again revealed that assessment practices in some subject areas like Mathematics, Science and Social Studies indicated favourable emphasis being given to formative assessment because 80% of the total marks have been allocated to on-going assessment and 20% was for the test. Moreover, students interviewed also agreed that project and portfolio assignment given were to a great extent real and authentic tasks that they could relate to their future workplace.

Beckmann, Senk and Thompson (1997) studied the assessment and grading practices of 19 high school mathematics teachers in the United States.

Their study revealed that the most frequently used assessment tools were tests and quizzes and these determined about 77% of students' grades. Twelve of the nineteen teachers used other forms of assessment such as written projects, experiments, demonstrations or interviews with students. The study also revealed that teachers recorded a very high level of student participation in the written projects, experiments.

Challenges of using Achievement Test

Eshun et al. (2014) conducted a study to investigate the influence of achievement test on classroom practices of teachers and the challenges they encounter in the Social Studies classroom in Ghana. The study used a descriptive case study design and it involved 10 senior high schools and twenty teachers randomly sampled from fifty-seven (57) senior high schools in the Central Region of Ghana. Semi-structured interview guide was the main instrument used for data collection. The research found out that the forms of achievement test some teachers used in their classrooms were limited due to examination policies, time, resources and assessment methods employed by their schools. Furthermore, they revealed that most teachers they observed were not using assessment techniques that involved students in the teaching and learning process. Again, they indicated that some teachers revealed that using the achievement test would delay them in completing topics in their syllabuses given to them. Beckmann, Senk and Thompson (1997) in their study conducted in USA identified three reasons why teachers do not use multiple assessment methods. First, some teachers had limited knowledge of different forms of assessment. Second, teachers felt they had no time to create/develop authentic assessment.

Chapter Summary

Studies in United States and, England revealed that teachers lacked competences in their test construction. In the case of Ghana, studies have shown discrepancies with respect to particular testing principles that teachers adhere to. Findings from all the studies gave ample evidence to conclude that, in terms of test administration, teachers possess some potential. However, with respect to test construction and scoring, studies have shown that teachers lacked appreciable competence.

CHAPTER THREE

RESEARCH METHODS

Introduction

The study sought to investigate achievement test practices of teachers in Junior High Schools in the Sissala East Municipality. It is generally accepted that, the quality of any research project hinges on gathering relevant information that would be used to solve a stated problem. The quality of these processes determines the validity and reliability of data collection and the results obtained (Willington, 2000). This chapter discussed the methodology that was employed in carrying out the study. The methods and approaches as described in the chapter were under nine sub-sections. These were the Research Design, Population, Sampling Procedure, Data Collection Instruments, Pre-testing Procedure, Validity and Reliability of the Instruments, Ethical Consideration, Data Collection Procedure and Data Processing and Analysis.

Research Design

The research design chosen for the study was the descriptive sample survey. According to Amedahe (2004), "descriptive research is research which specifies the nature of a given phenomenon" (p. 50). Gay (cited in Amedahe, 2004), explains that descriptive research involves the collection of data in order to test hypotheses or answer research questions concerning the current status of the subjects of the study. Dawson (2002) posits that a research design is the conceptual structure within which research would be conducted.

In this regard, this study would adopt the descriptive research design. Descriptive research was used because; the data collected was used to investigate achievement test practices of teachers in Junior High Schools in the Sissala East Municipality. Surveys was assisted in gaining a better understanding of achievement test practices of teachers in Junior High Schools in the Sissala East Municipality.

Again, the researcher chose this approach because he was interested in learning about the practice of achievement test from JHS teachers' perspective in the Sissala East Municipality. According to Murphy (2009), the major advantage that goes with this type of design is that, the data collection techniques present several advantages as they provide a multifaceted approach for data collection. 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 thus it provides more accurate picture of events and seeks to explain people's perceptions and behaviour on the basis of data gathered at a point in time (Murphy, 2009).

However, the design has some weaknesses. Confidentiality is the primary weakness of descriptive research (Murphy, 2009). According to Murphy (2009), respondents are often not truthful as they feel the need to tell the researcher what they think the researcher wants to hear and also participants may refuse to provide responses they view to be too personal. Another weakness of this design, according to Murphy (2009) is that it presents the possibility for error and subjectivity. However, the design was used despite its weaknesses because it seeks to explain people's perceptions and behaviour on the basis of data gathered at a point in time and can provide statistics about an event while also illustrating how people experience that event thus providing a multifaceted approach for data collection.

Population

According to Diamantopoulos (2004), a population is a group of items that a sample will be drawn from. Alan (2000), also defined a population as a set of all measurements that is of interest and possesses at least one common characteristic. A target population can be viewed as a group with things in common, which distinguishes them from other groups. In the view of (Neumann, 2006) a target population is made of group of cases from which a researcher studies a sample and then generalizations are made from the results of the sample. The population for this study comprises all Junior High School teachers in the Sissala East Municipality. The population is 700 Junior High School teachers in the Sissala East Municipality.

Sampling Procedures

Sarantakos (2005) postulated that a sample consists of a carefully selected unit that comprises all the categories of the population. Sarantakos (2005) indicates that estimation of the sample size varies significantly, with some researchers showing interest in pure quantity, others in quality and yet others combining in what is called triangulation of sources, data and methods. However, a wise rule is that the sample size must be as large as necessary, and as small as possible. An estimated sample size of 248 junior high teachers selected for the study using Krejcie and Morgan (1970) sampling table. According to the table, a population of 700 gives a sample size of 248. Fraenkel and Wallen (2009) have also indicated that for descriptive studies, a larger sample size produces desirable results to generalise over the population. Therefore, a sample size of 248 for this study was considered large enough to produce the desired results and allowed for generalisation of the findings over the population.

The study employed the multistage sampling techniques (purposive, stratified and simple random sampling procedure). Per the nature of the study population, purposive, stratified and simple random sampling procedure were used to select cases in the public Junior High Schools in nine (9) educational circuits in the Sissala East Municipality. Purposive sampling was used because the researcher selectively chose to study only Junior High School teachers teaching the four core subjects: English Language, Mathematics, Integrated Science, and Social Studies. According to Crossman (2017), purposive sampling is a non-probability sample that is selected based on characteristics of a population and the objective of the study. Purposive sampling is also known as judgmental, selective, or subjective sampling. This type of sampling can be very useful in situations when you need to reach a targeted sample quickly, and where sampling for proportionality is not the main concern (Crossman, 2017).

Stratified random was used because the population comprised of different circuits within the Sissala East Municipality. Stratified sampling was used to select equivalent number of Junior High Schools from the nine circuits. According to Van Dalen (2012), stratified sampling is a procedure for selecting a sample that includes identified subgroups from the population in the proportion that they exist in the population. Van Dalen posited that, stratified sampling can be used to select equal numbers from each of the identified subgroups if comparisons between subgroups are important.

According to Cohen, Manion and Morrison (2011), the quality of any research not only stands or falls by the appropriateness of methodology and instrumentation but also by the suitability of the sampling strategy that is adopted. Therefore, the researcher used stratified sampling to guarantee the desired distribution among the selected subgroups of the population and to aid equivalent selections of Junior High Schools from the nine different circuits.

At the last stage, the researcher used simple random (lottery method) to select the Junior High Schools in the nine (9) educational circuits in the Sissala East Municipality. The simple random technique was used in order to give Junior High Schools equal chance of being selected and it helped to avoid biases in selecting the Junior High Schools. This is to help improve the representativeness of the sample by reducing sampling error (Saunders et al., 2007).

Data Collection Instrument

The questionnaire was the main source of collecting data for the study. The instrument was developed by the researcher from literature. A thorough literature reviewed on research related to achievement test was performed prior to the development of the questionnaire. This instrument was used as the main tool for data collection as it affords greater assurance of confidentiality and anonymity to respondents (Sarantakos, 2005). Questionnaire was used for the study because it offered the researcher the opportunity to sample the perceptions of a larger population. The items on the questionnaire were prepared based on the objectives of the study to elicit the needed information.

Saunders (2007) reiterates that a questionnaire is an ideal tool when collecting a lot of information over a short period of time. Again, the researcher deemed it ideal to use questionnaire because his respondents were literate. The questionnaire was closed ended type. The questionnaire was developed using four- point Likert-type scale ranging from “Strongly Disagree to Strongly Agree”. The research instrument was organised into six sections (A, B, C, D, E and F). Section ‘A’ comprises the background information of the students. The Section, ‘B’, constitutes the basic principles of items construction of achievement test by teachers in Junior High Schools in the Sissala East Municipality. “Section C” constitutes the basic principles of test administration of Junior High Schools teachers in the Sissala East Municipality. “Section D” was made up of how Junior High Schools teachers in the Sissala East Municipality followed the basic principles of tests scoring. “Section E” was based on the kinds of achievement test strategies Junior High Schools teachers in the Sissala East Municipality use to assess their students’ learning outcomes and finally “Section F” sought to elicit information on the challenges Junior High Schools teachers in the Sissala East Municipality encounter in the use of achievement test.

The questionnaire was a four-point Likert -type scale which requires participants to indicate their level of agreement or disagreement to the items using strongly agree, agree, disagree or strongly disagree. The responses were scored as follows: Strongly Agree = 4; Agree = 3; Disagree = 2; Strongly Disagree =1.

Validity and Reliability of the Instrument

According to Dambudzo (2009), the idea of validity hinges on the extent to which research data and the methods of obtaining the data are deemed 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 attributes targeted in the proposed research. Validity ensures that inferences based on collected data are accurate and meaningful. It is necessary to have experts examine the instrument items and judge their representativeness (McMillan & Schumacher, 2001). To ensure the validity of the construct, the self-developed questionnaire was evaluated by my supervisors in the Department of Education and Psychology. Based upon this, some changes were made to the questionnaire prior to the pre-testing.

Subsequently, to achieve the reliability of the instrument, Cronbach alpha 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. The research instruments were improved by revising or deleting items. For reliability of the instruments, a pre-test of the instrument were carried out on Junior High School teachers in the Sissala West Municipality to check the reliability of the instruments. The aim of the pre-testing was to improve the reliability of the instruments.

The respondents were given draft copies of the questionnaire. The respondents were told to discuss verbally and frankly with me any ambiguity, incoherence or incomprehension that they would experience about any aspect of the draft questionnaire. The necessary corrections were effected after the trial testing. The pre-test results were used to determine the reliability of the instruments with the Cronbach's Alpha measure of internal consistency. A reliability coefficient of .81 was attained.

Ethical Considerations

McNabb (2004) points out that there are four stages in research ethics, namely: planning, data gathering, processing and interpretation of data as well as the dissemination of results. At the data collection stage, in conducting administering questionnaires, ethical guidelines were followed. The teachers were given the opportunity to fill their questionnaires privately, in order to ensure confidentiality. In dissemination of results, measures were taken to ensure privacy, anonymity and confidentiality of all participants. This means that the names of the participants were not used or revealed throughout the research project (Maree, 2007). The discussions of the findings were based on the trends that emerged from the data and not from any preconceived ideas.

Data Collection Procedures

A letter of introduction was collected from the Department of Education and Psychology, University of Cape Coast, to seek for permission from the Head teachers in the Sissala East Municipality. The questionnaires were administered by the researcher to two-hundred and forty-eight teachers (248) in the Sissala East Municipality. The researcher had also established the necessary contacts with the head teachers of the selected schools to seek permission to administer the questionnaire. A brief self-introduction was made by the researcher to explain the purpose of the study to the respondents before the questionnaires were distributed to them. The researcher stayed with them and had interactions with them. The researcher has appealed to all the respondents to take their time to read the questionnaire and respond to it appropriately.

Data Processing and Analysis

In every research, data collected becomes meaningful only when it is organized and summarized. The analysis focused on descriptive statistics that involved computing of frequencies, percentages, means and standard deviations. The hypothesis was analysed using One Way Analysis of Variance (ANOVA). This was tested using .05 significance level.

Chapter Summary

The research was quantitatively motivated. The design employed for the study was the descriptive survey design. Data were analysed using Inferential (ANOVA) and descriptive statistics (means and standard deviations).

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

This chapter presented an analysis of the data gathered from the field in relation to achievement test practices of teachers in Junior High Schools in the Sissala East Municipality. The study aimed at finding out whether Junior High School teachers in the Sissala East Municipality practise the basic principles of construction, administration and scoring of classroom achievement tests. The data was analysed using descriptive statistics (frequencies, percentages, means and standard deviations) and inferential statistics (ANOVA). The results were presented with discussions. The results on the demographics data was presented in Table 1.

Table 1: *Results on the Demographics of the Respondents*

Demographics Variables	Sub-scales	Freq.(No)	Percent. (%)
Gender	Male	146	58.8
	Female	102	41.2
Number of years in teaching service	Under 5 years	92	37.1
	5 – 10 years	109	43.9
	Above 10 years	47	18.9
Professional Qualification	Teachers' Certificate A	02	0.81
	Diploma with Education	06	2.41
	Bachelors with Education	185	74.5
	Masters with Education	48	19.4
	Masters without Education	07	2.82
	Others	00	0.00

Source: Field Data, 2019 n=248

From Table 1, 146 representing 58.8% of teachers were males while 102 of them representing 41.2% were females. With respect to the number of years in teaching service, the results showed that most of the teachers that is (109) representing 43. 9% had taught for 5-10 years. Few of them representing 18.9% had taught above 10 years. On the last aspect of the demographic characteristics of the teachers, the results indicated that most of the teachers (n=185, 74.5%) hold Bachelors with Education. Those with Masters with Education followed (n=48, 19.4%).

Research Question One

This research question sought to find out the kind of principles that Junior High School teachers use in the construction of their achievement tests. In addressing this research questions (**Q1abc-Q4**), means and standard deviations were used for the analysis. The teachers were given a four-point Likert scale items on teachers use in the construction of their achievement tests to respond to. The scoring of items was based on the four-point Likert scale of measurement ranging from "Strongly Agree" (scored 4) to "Strongly Disagree" (scored 1). In the analysis, means provides the summary of the responses from teachers and the standard deviation indicates whether teachers' responses were clustered to the mean score or dispersed.

The criterion value (CV) of 2.50 was established for the scale. To obtain the criterion value (CV=2.50), the scores were added together and divided by the number of the scale (4+3+2+1= 10/4=2.50) (Green & Neil, 2014). To understand and interpret the mean scores, any items/statements that scored a mean of 2.50 and above indicate respondents' positive perception of the variables under study while a mean of 2.49 and below indicates a negative perception towards variables under study. The findings are presented as below:

Research Question 1a: How do Junior High Schools teachers in the Sissala East Municipality adhere to the construction of test items?

In the quest of achieving the purpose of the study, I assessed how Junior High Schools teachers in the Sissala East Municipality adhere to the principles of construction of test items. In achieving this, the responses from the teachers were analysed using Means and Standard Deviations. The results are presented in Table 2.

Table 2: *Results on how Junior High Schools teachers in the Sissala East Municipality adhere to test construction*

When constructing test, I…….	N	Mean	SD
	Test Value=2.50		
Evaluate items given to the students	248	3.87	1.13
Set questions from past questions	248	3.57	1.02
Consider the time individual will spend on a question	248	3.53	1.09
Provide clear and simple instructions on how test is to be answered	248	3.45	1.35
Consider students' language proficiency	248	3.34	1.82
Follow the principles of test construction for each format	248	3.25	1.92
Write items at least two weeks before time	248	2.98	1.17
Consider meaning of wording against different ethnic background	248	2.92	1.26

Table 2 Continue

Prepare marking scheme after students have answered the question	248	2.13	1.52
Use a test specification table	248	2.22	1.46
Consider variation of students with respect to physical disability	248	2.23	1.43
Match learning outcomes to the items	248	2.45	1.97
Construct test when it is time to assess	248	2.35	1.14
Write more items than needed	248	2.32	1.96
Specify the construct to be measured	248	2.23	1.19
Ask any other colleagues to help me construct items	248	2.15	1.14
Use questions directly from text books	248	2.12	1.76
State the purpose of the test	248	2.11	1.28
Try solving the questions myself to determine the time required	248	2.02	1.13
Average Mean/SD	248	2.46	1.44

Source: Field Data, 2019 Cut-off Mean value=2.50

Key-M= Mean, SD =Standard Deviation, n=Sample Size

Table 2 presents results on how Junior High Schools teachers in the Sissala East Municipality reported that they adhere to the principles of construction of test items. The results showed that Junior High Schools teachers in the Sissala East Municipality reported that did not adhere to most principles of test construction (MM=2.46, SD=1.44). Some of the test construction principles Junior High School teachers in the Sissala East Municipality reported that they adhere to include the following:

a. The teachers confirmed that they evaluate test items given to their students (M=3.87, SD=1.13, n=248)

b. They further indicated that they consider the time individual will spend on a question (M=3.53, SD=1.09, n=248)

c. They agreed that they provide clear and simple instructions on the test paper as how the test should be answered (M=3.45, SD=1.35, n=248)

d. Another construction principle Junior High School teachers in the Sissala East Municipality adhere to, was that they consider their students' language proficiency (M=3.34, SD=1.82).

e. The fifth construction principle Junior High Schools teachers in the Sissala East Municipality adhere to, was that they followed the principles of test constructions for each format (M=3.25, SD=1.92, n=248).

f. Junior High Schools teachers in the Sissala East Municipality confirmed that they write test items at least two weeks before time (M=2.98, SD=1.17, n=248).

g. Junior High Schools teachers in the Sissala East Municipality also consider the meaning of wording against different ethnic background when constructing test items (M=2.92, SD=1.26, n=248).

h. It was confirmed that few Junior High Schools teachers in the Sissala East Municipality least averagely prepare marking scheme after students have answered questions (M=2.13, SD=1.52, n=248)

Some of the test constructions principles Junior High Schools teachers in the Sissala East Municipality did not adhere to are the following;

a. Most Junior High Schools teachers in the Sissala East Municipality were below average in their use of test specification table (M=2.22, SD=1.46, n=248).

b. Again, below average of the Junior High Schools teachers in the Sissala East Municipality were considering variation of students with respect to physical disability (M=2.23, SD=1.52, n=248).

c. In a similar result, below average of the teachers pointed out that they match learning outcomes to the items (M=2.45, SD=1.97, n=248).

d. Also, below average of the Junior High Schools teachers in the Sissala East Municipality were writing more items than needed (M=2.32, SD=1.96, n=248).

e. Some few teachers in the Sissala East Municipality again pointed out that they ask any other colleagues' teacher to go through their constructed test items (M=2.15, SD=1.14, n=248).

f. Some few teachers in the Sissala East Municipality were of the view that they use questions directly from text books (M=2.12, SD=1.76, n=248).

The findings from the present study disagree with the assertion of Tom and Gary (2003), who indicated that, when teachers fail to consider meaning of words against different ethnic background in constructing test items, the interpretation made from the test may lead to faulty conclusions. The possible cause of this finding may be the limited time and excessive workload on teachers which may lead them to pay less attention to such important principles.

The study further revealed that teachers often ask other colleague who are not in the subject area to help them construct test items. This attitude might have a great deal of implication to validity of test results. This is because the teacher assessing the students might not appropriately measure the real competence of the students since he/she might not know the detail of the content coverage and the thinking process to assess on a particular topic. The result from the study also revealed that, teachers do not often review their test items before administering them. This confirms the findings of Quaigrain (1992) who indicated that some teachers do not review their test.

The accumulated findings on how Junior High Schools teachers in the Sissala East Municipality adhere to the construction of test items supports the assertion of Wiliam (2008), who stated that, to increases the validity of a test, teachers must consider the student's language proficiency. He further stipulated that test would be invalidated if it turned out that the reading requirements of the test were so demanding that students with poor reading ability, but a sound understanding obtained low marks. On the other hand, if a student possesses an understanding of an issue demanded by a test, but fails to show it for reasons of linguistic difficulty then, the results of that test would be invalid.

Research Question 1b: How do Junior High Schools teachers in the Sissala East Municipality adhere to the Administration of test items?

Test administration serves as one of the key components of achievement test in the classroom. In achieving this, the responses from the teachers were calculated using Means and Standard Deviations to show how they adhere to test administration. The results are presented in Table 3.

Table 3: *Results on how High Schools teachers in the Sissala East Municipality adhere to test Administration*

When administering test, I…….	N	Mean	SD
	Test Value=2.50		
Prepare classroom a day before test is taken	248	1.77	1.78
Inform student about the test format	248	3.63	1.75
Give more instructions during the time the students are taking the test	248	2.98	1.27
Proof read all test items before administration	248	2.76	0.96
Inform students in advance areas for the test	248	2.72	1.76
Make provision for extra sheets and writing materials	248	2.20	1.74
Make students aware of the rule and regulation covering the test	248	2.17	1.22
Make provision for emergencies during the time the test is taken	248	2.12	1.95
Students start and stop test on time	248	1.85	1.65
Tests are given after a long vacation or important holidays	248	1.72	1.25
Adequate ventilation and lighting	248	1.57	1.14
Use "DO NOT DISTURB SIGN" at the entrance of classroom	248	1.35	1.84
Mean of means /SD	248	2.44	1.66

Source: Field Data, 2019   Cut-off Mean value=2.50

Key-M= Mean, SD =Standard Deviation, **n=Sample Size**

Table 3 gives result on how Junior High Schools teachers in the Sissala East Municipality adhere to the principle of test administration. The results showed that, generally, just below average of the teachers in the Sissala East Municipality adhere to test administration principles in their achievement test.

This was evident after the Mean of Means (MM=2.44, SD=1.66) was less than the Cut-off Mean value of 2.50. The teachers only adhere to some few principles which include:

a. They confirmed that they inform student about the test format (M=3.63, SD=1.75, n=248).
b. Another test administration principle was the fact that most give more instructions in the test paper the time the students are taking test (M=2.98, SD=1.27, n=248).
c. Above average of the Junior High Schools teachers in the Sissala East Municipality indicated that they proof read all test items. (M=2.76, SD=0.96, n=248).

The following are some key principles that below average of the teachers they averagely adhered to which could affect the results of achievement test.

a. Below average of the Junior High Schools teachers in the Sissala East Municipality indicated that they make provision for extra sheets and writing materials (M=2.20, SD=1.74, n=248).
b. In another breath, very few of the teachers make students aware of the rules and regulations covering achievement test (M=2.17, SD=1.22, n=248).
c. Most of the Junior High Schools teachers in the Sissala East Municipality pointed out that they least adhere to the principles; students starting and stopping test on time (M=1.85, SD=1.65, n=248).

d. Majority of the Junior High Schools teachers in the Sissala East Municipality pointed out that they least provided adequate ventilation and lighting (M=1.57, SD=1.14, n=248).

e. Finally, the teachers indicated that they least used the "DO NOT DISTURB SIGN" at the entrance of classroom (M=1.35, SD=1.84, n=248).

From the results in the Table 3 it is evident that most teachers averagely often prepare their students in advance before administering test. This might lead to improper arrangement environment for a test which can affect students' performance. This is because students trying to find a proper place to sit, due to improper arrangement of desks, poor lightening, among other discrepancies may emotionally affect students. Notwithstanding the cause of this practice might be from the fact that, most of the Junior High Schools do not have adequate facilities in terms of classroom and desks to accurately administer tests without interrupting learning process in other classes with respect to space, desks, lighting among others. This finding does not support Anhwere (2009) whose earlier findings suggested that teachers at the Training colleges had adequate facilities and also put in much effort to organise classroom appropriately when administering tests.

The findings further reveal that, teachers averagely control noise when administering tests.This practice is not consistent with the assertion made by Mehrens and Lehmann (2001). According to Mehrens and Lehmann, 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 student's performance. Etsey

(2004) also affirmed 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. The distraction from outside can divert the attention of test takers which could contribute to low performance of students.

The result also indicated that teachers often give tests immediately after a long vacation or an important holiday. This practice does hinder the test construction principles. The practice is inconsistent with the assertion made by Amedahe and Asamoah-Gyimah (2003), and Etsey (2004) who found that 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.

Amedahe and Asamoah-Gyimah (2003) went on to say that 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. Teachers in the field of testing must recognise that the implication from the interpretation made of tests weigh far greater impact on the students more than the teachers' idea of getting a score to represent assessment. Therefore, it would be prudent for teachers to ensure that scores made from students' successive tests yield an appreciable consistency. According to Crocker and Algina (2008), psychological measurement should focus on a way of reducing systematic errors which may result from factors which include "fatigue, boredom, forgetfulness, guessing" among others (p. 6).

Research Question 1c: How do Junior High Schools teachers in the Sissala East Municipality adhere to the scoring of test items?

In achievement test, scoring of test serves as one of the principal components in the classroom that teachers are to adhere to. I therefore assessed how Junior High Schools teachers in the Sissala East Municipality score test items. In achieving this, the responses from the teachers were calculated using Means and Standard Deviations to show how they adhere to test scoring. The results are presented in Table 4.

Table 4: *Results on how Junior High Schools teachers in the Sissala East Municipality adhere to scoring of test items*

When Scoring test, I ……….	n	M	SD
	Test Value=2.50		
mark papers just after the test is taken	248	2.09	1.18
prepare scoring guide	248	2.63	1.65
make sure test takers are kept anonymous	248	1.98	1.97
grade the responses item by item	248	2.96	0.46
keep scores of previous items out of sight	248	1.72	1.86
periodically rescore previously scored items	248	1.90	1.14
shuffle scripts before scoring	248	2.09	1.02
score essay test when I am physically sound and mentally alert in a sound environment	248	1.72	1.58
constantly follow scoring guide	248	2.15	1.75
am influence by the first few papers read when scoring test items	248	3.22	1.58
score a particular item on all papers at a sitting	248	1.57	1.54
provide comments and errors correct on scripts	248	1.35	1.27
give extra marks to students based on Handwriting, Gender etc.	248	1.43	1.58

Source: Field Data, 2019 Cut-off Mean value=2.50

Key-M= Mean, SD =Standard Deviation, **n=Sample Size**

Table 4 depicts results on how Junior High Schools teachers in the Sissala East Municipality score test items. The results give evidence that most Junior High Schools teachers in the Sissala East Municipality have poor scoring principles and this can affect their achievement test. Almost all the pre-coded items were confirmed by the teachers. Few of the scoring principles that the teachers followed were that they:

a. below averagely prepare scoring guide (M=2.63, SD=1.65, n=248).

b. below averagely grade the responses item by item (M=2.96, SD=0.46, n=248).

On a larger scale, Junior High Schools teachers in the Sissala East Municipality who adhere to the Test Scoring Principles were below average. Some of the flaws include the fact that:

a. It was evident that most Junior High Schools teachers in the Sissala East Municipality below averagely mark papers immediately after the test is taken (M=2.09, SD=1.18, n=248).

b. It was again evident that most Junior High Schools teachers in the Sissala East Municipality least prepare scoring guide (M=1.63, SD=1.65, n=248).

c. It was apparent that most Junior High Schools teachers in the Sissala East Municipality least make sure that test takers are kept anonymous (M=1.98, SD=1.97, n=248).

d. In similar results, the teachers least kept scores of previous items out of sight (M=1.72, SD=1.86, n=248).

e. Junior High Schools teachers in the Sissala East Municipality least periodically rescore previously scored items (M=1.90, SD=1.14, n=248).

Teachers indicated that they constantly follow the scoring guide when marking their tests. This process must be hailed to since such attitude would ensure consistency of test scores. This finding supports the assertion that admonish teachers to constantly follow the marking scheme as they score tests items, as this reduces rater drift, which comes from the likelihood of either not paying attention to the scoring guide or interpreting it differently as time passes (Mehrens & Lehmann, 2001; Amedahe & Gyimah, 2003; & Etsey, 2004).

Notwithstanding, the result from the research also indicated that, most teachers do not often consider reshuffling script when scoring their test. The finding opposes the assertion of Mehrens and Lehmann (2001) who asserted that, randomly reshuffling of scripts when beginning to score each set of items will minimise the bias introduced as a result of the position of one's script. Research by Hales and Tokar (as cited in Mehrens & Lehmann, 2001) 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 (2001) 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.

Another finding of the research indicated that, teachers do not often score a particular item on all papers at a sitting. This practice has been chastised by Mehrens and Lehmann (2001); Amedahe and Gyimah (2003);

and Etsey (2004), who agreeably asserted that responses of item should be scored item by item rather than script by script. This principle is to minimise the carryover effect on the scores and thereby ensure consistency. However, this finding do not support the findings of Amedahe (1989), who recounted that teachers in the schools used mainly the analytic method in scoring their essay-type tests. He further asserted that, teachers in the schools scored their essay-type tests either item by item or script by script. On the part of Quaigrain (1992), he found that majority of teachers in the schools used the analytic method in scoring their essay-type tests.

With regards to scoring, teachers also indicated that they give extra marks to students based on handwriting, gender etc. This practice has been elaborated by Amedahe and Gyimah (2003), and Etsey (2004), who indicated that, 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. When teachers are influenced by factors other than the subject matter, the marks awarded would represents construct irrelevant or construct mis-representativeness. This simply means higher scores on tests might not reflect the ability of students on the subject matter but rather discriminate students in proficiencies they have over other students.

The results also indicated that, anonymity is not ensured when teachers score their test. This finding flouts the assertion of Etsey (2004) who indicated that scripts must be scored anonymously. He suggested 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.

Research Question Two: What kinds of achievement test strategies do Junior High Schools teachers in the Sissala East Municipality use to assess their students' learning outcomes?

To obtain a comprehensive result, I assessed the kinds of achievement test strategies that Junior High Schools teachers in the Sissala East Municipality use to assess their students' learning outcomes. In accomplishing this, the responses from the teachers were compiled and ranked using Means and Standard Deviations. The results are presented in Table 5.

Table 5: *Results on the kinds of achievement test strategies Junior High Schools teachers in the Sissala East Municipality use*

Kinds of achievement test strategies	N	M	SD	Remarks
Writing Samples	248	3.19	1.65	S
Assessing work samples	248	3.09	1.78	S
Experiments/Demonstrations	248	2.96	0.98	S
Presentations	248	2.16	1.72	NS
Computer simulation task	248	1.95	1.49	NS
Exhibitions	248	1.86	1.54	NS
Projects	248	1.72	1.75	NS
Constructed-Response Items	248	1.67	1.12	NS
Report writing	248	1.66	1.59	NS
Role-play	248	1.63	1.54	NS
Drama	248	1.42	1.53	NS
Story Telling	248	1.09	1.57	NS

Source: Field Data, 2019 Cut-off Mean value=2.50

Key-M= Mean, SD =Standard Deviation, n=Sample Size, S=Strategy, NS=Not a Strategy

Table 5 presents on the kinds of achievement test strategies Junior High Schools teachers in the Sissala East Municipality use to assess their students’ learning outcomes. From the results, it is evident that few of the achievement test strategies are used. Some of the strategies include: writing samples (M=3.19, SD=1.65, n=248); assessing work samples (M=3.09, SD=1.78, n=248); experiments/demonstrations (M=2.96, SD=0.98, n=248).

Some of the kinds of achievement test strategies Junior High Schools teachers in the Sissala East Municipality averagely use to assess their students’ include presentations (M=2.16, SD=1.72, n=248); computer simulation task (M=1.95, SD=1.49, n=248); exhibitions (M=1.86, SD=1.54, n=248); projects (M=1.72, SD=1.75, n=248); constructed-response items (M=1.67, SD=1.12, n=248); report writing (M=1.66, SD=1.59); role-play (M=1.63, SD=1.54. n=248); drama (M=1.42, SD=1.53) and storytelling (M=1.09, SD=1.57, n=248).

Research Question Three: What challenges do Junior High Schools teachers in the Sissala East Municipality encounter in the use of achievement test?

I assessed challenges that Junior High Schools teachers in the Sissala East Municipality encounter in the use of achievement test. In achieving this, the responses from the teachers were compiled and ranked using Means and Standard Deviations. The results are presented in Table 6.

Table 6: *Results on the challenges Junior High Schools teachers in the Sissala East Municipality encounter in the use of achievement testing*

Challenges	N	M	SD	Remark
		Test Value=2.50		
Inadequate time to prepare in terms of gathering information and materials to be used for achievement testing.	248	3.72	1.23	A challenge
Large class size makes it difficult to assess students using achievement testing.	248	3.67	1.74	A challenge
Inadequate time allotted on the timetable for various subjects does not permit the use of achievement testing.	248	3.66	1.64	A challenge
Developing achievement testing task is difficult	248	3.65	1.46	A challenge
Lack of support from the school authorities in terms of logistics and facilities	248	3.56	1.78	A challenge
Lack of funds to embark on some activities and projects	248	3.52	1.48	A challenge
The school assessment system makes it difficult to use achievement testing	248	3.49	1.40	A challenge
Some topics are difficult to assessed using achievement testing	248	2.85	1.29	A challenge
Mean of Means/SD	248	3.48	1.47	

Source: Field Data, 2019 Cut-off Mean value=2.50

Key-M= Mean, SD =Standard Deviation, **n=Sample Size**

Table 6 present results on the challenges do Junior High Schools teachers in the Sissala East Municipality encounter in the use of achievement test. The results showed that there are numerous challenges that confront the use of achievement test. Some of the challenges included inadequate time to prepare in terms of gathering information and materials to be used for achievement testing (M=3.72, SD=1.23, n=248). Another challenge was the fact that large class size makes it difficult to assess students using achievement testing (M=3.67, SD=1.74, n=248)

In another results, inadequate time allotted on the timetable for various subjects does not permit the use of achievement testing (M=3.67, SD=1.74, n=248). Developing achievement testing task is difficult also served as another challenge (M=3.65, SD=1.64, n=248). Another challenge was lack of support from the school authorities in terms of logistics and facilities (M=3.56, SD=1.78, n=248).

Aside the above, Junior High Schools teachers in the Sissala East Municipality confirmed that lack of funds to embark on some activities and projects (M=3.52, SD=1.48, n=248). Junior High Schools teachers in the Sissala East Municipality pointed out that the school assessment system makes it difficult to use achievement testing (M=3.49, SD=1.40). Moreover, the teachers agreed that Lack of motivation from school authorities pose a challenge (M=3.16, SD=1.22). Finally, some topics are difficult to be assessed using achievement testing (M=2.85, SD=1.29, n=248).

The results are in line with the study of Eshun et al. (2014) conducted a study to investigate the influence of achievement test on classroom practices of teachers and the challenges they encounter in the Social Studies classroom in Ghana. The study used a descriptive case study design and it involved 10 senior high schools and twenty teachers randomly sampled from fifty-seven (57) senior high schools in the Central Region of Ghana. Semi-structured interview guide was the main instrument used for data collection. The research found out that the forms of achievement test some teachers used in their classrooms were limited due to examination policies, time, resources and assessment methods employed by their schools. Furthermore, they revealed that most teachers they observed were not using assessment techniques that involved students in the teaching and learning process. Again, they indicated that some teachers revealed that using the achievement test would delay them in completing topics in their syllabuses given to them

Research Hypothesis

H_0:1 there is no statistically significant difference among the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality with respect to how they adhere to test construction

H_A:1 There is a statistically significant difference among the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality with respect to how they adhere to test construction.

At an alpha level of .05 confidence, the hypothesis was tested to find out whether the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality will differ in terms of how they adhere to test construction. To achieve this, between-groups one-way analysis of variance (ANOVA) was deemed appropriate for the analysis. To obtain the scores for the analysis, the responses on how they adhere to test construction transformed into a single variable. The data on questionnaire was made up of independent variable that is the years of working experience which is categorical (nominal) and dependent variable was test construction which was measured on continuous scale. The between-groups one-way analysis of variance (ANOVA) was conducted to determine whether there are any statistically significant differences among the means of the independent groups (years of working experience) and test construction. ANOVA assumptions of normality and homogeneity of variances of the data distribution was checked. Figure 1 and 2 present the Test of Normality and Linearity.

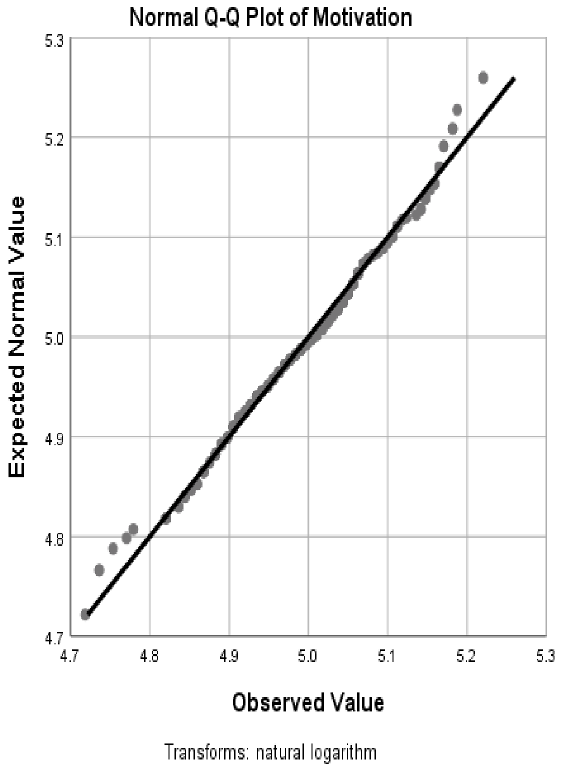


Figure 1: Diagnostic Test of Normality and Linearity

Source: Field survey (2019)

According to Pallant (2007), a straight normal probability plot is an indication of normality and linearity. Pallant noted that when ANOVA assumptions are met, it produces a reliable result. From Figures 1, 2 and 3 a reasonable straight line could be seen from the plot demonstrating normality and linearity of the data among the two variables (Years of teaching experience and test construction). This therefore, means that conducting between-groups one-way analysis of variance (ANOVA) test was justified.

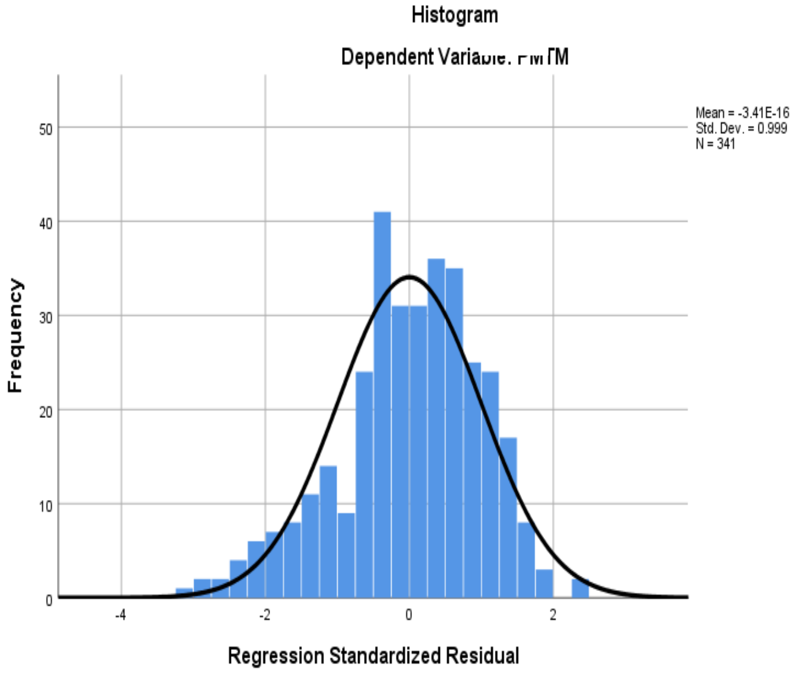


Figure 2: Histogram Test of Normality and Linearity

Source: Field survey (2019)

The Histogram plot of standardised predicted values verses standardised residuals showed that the data met the assumptions of normality of variance and linearity and the residuals were approximately normally distributed.

Table 7: *Normality Test Results of the Variables*

Years of Working Exp.	Kolmogorov-Smirnov[a] Statistic	Df	Sig.	Shapiro-Wilk Statistic	Df	Sig.
Under 5 years	.063	140	.200*	.991	20	.545
6 – 10 years	.167	20	.145	.895	41	.053
Above 11 years	.051	168	.200*	.993	168	.592

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Source: Field Survey (2019) *Significant difference exist at p$\leq$0.05

From Table 7, Kolmogorov-Smirnov was reported based on the assumptions that it uses a sample size greater than 50 (n >50). The results indicated that the dependent variable (test construction) was normally distributed among years of working experience. For example, under 5 years scored a Kolmogorov-Smirnov indicating that it was normal *(KS = .063, df=140, p=.200, n=248,* teachers who have worked for 6 – 10 years recorded a Kolmogorov-Smirnov indicating that it was normal *(KS = .167, df=20, p=.145, n=248)*. Finally, teachers who have worked for Above 11 years also recorded a Kolmogorov-Smirnov indicating that it was normal *(KS = .051, df=168, p=.200, n=248).*

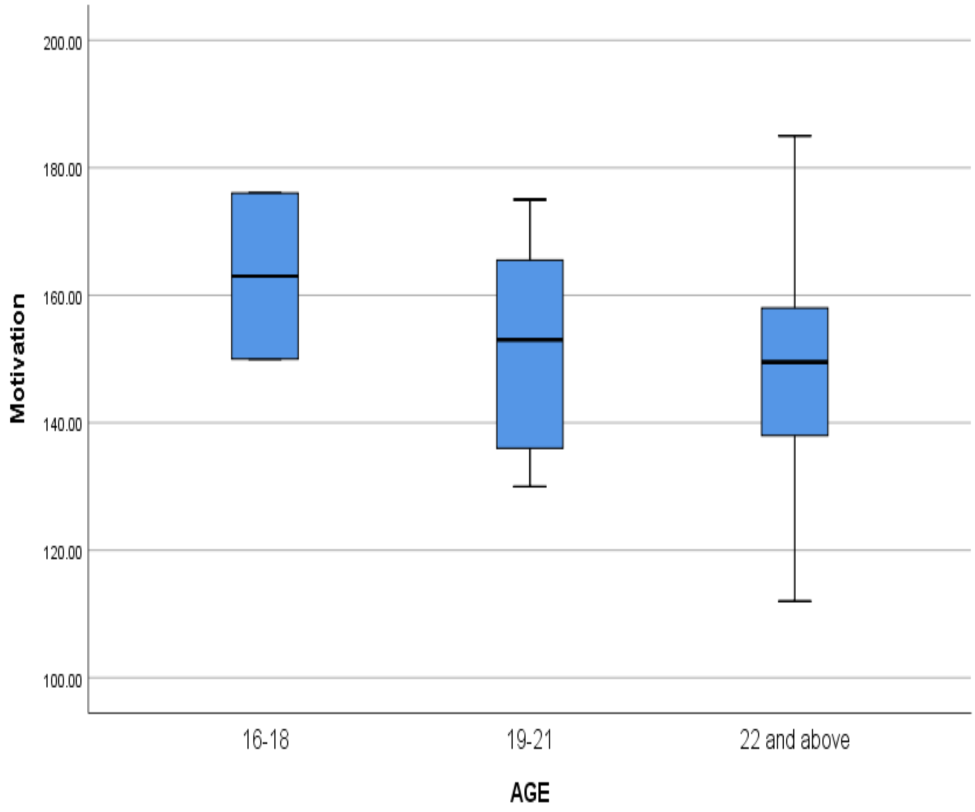


Figure 3: Normality of Test of Study Variables

Source: Field survey (201**9)**

Having tested for the normality, we progressed to check whether the data were homogeneous. The results are presented in Table 5.

Table 8: *Results of Homogeneity of Variances Test*

Variables		Levene Statistic	df1	df2	Sig.
Motivation	Based on Mean	.773	2	187	.463
	Based on Median	.770	2	187	.464
	Based on Median and with adjusted df	.770	2	179.992	.464
	Based on trimmed mean	.774	2	187	.463

Source: Field Data (2019) *Significant difference exists at P≤0.05, n=248

Table 8 depicts the test of homogeneity of variances of the study variables. The homogeneity of variances test results indicated that, assumption of homogeneity has not been violated. [t (df1=2, df2=187) = .773, p<0.05, Sig. = .463, 2-tailed)]. Performing of ANOVA test was therefore justifiable. Table 9 presents results on the descriptive statistics of the test.

Table 9: *Descriptive Statistics of the Test*

	Mean	Std. D	Std. Error	95% Confidence Interval for Mean Lower Bound	Upper Bound	Min	Max
Under 5 years	163.00	18.384	13.00	-2.1807	328.1807	150.0	176.0
6 – 10 years	152.05	16.086	3.597	144.52	159.57	130.0	175.0
Above 11 years	148.85	15.041	1.160	146.56	151.14	112.0	185.0
Total	149.34	15.187	1.101	147.16	151.51	112.0	185.0

Source: Field Data (2019) n=248

The descriptive statistics as in Table 9 demonstrates that, the differences existed in the mean scores. For example teachers from under 5 years was the highest (M= 163.00, SD= 18.384) indicating that descriptively, teachers under 5 years' experience construct test well. This was followed by those from 6 – 10 years (M=152.05, SD= 16.086). The descriptive statistics further indicated that those from above 11 years were least constructors of test (M= 148.85, SD= 15.041). Nevertheless, the one-way analysis of variance (ANOVA) was conducted to establish more statistical evidence on whether the observed difference was by chance. Figure 6 presents an easy way to compare the mean scores of the variables.

Table 10: *Summary of One-way Analysis of Variance (ANOVA) Results*

Sources	Sum of Squares	Df	Mean Square	F	Sig.	Rks
Between Groups	559.242	2	279.621	1.215	.299	No Diff.
Within Groups	43037.521	172	230.147			
Total	43596.763	175				

Source: Field Data (2019) *Significant difference exists at p≤0.05, n=175

A one-way Analysis of variance (ANOVA) was conducted to compare mean scores of the study variable. From the one-way ANOVA in Table 10, the results show that there was no statistically significant difference in the years of working experience of the teachers and test construction, *F (df1=2, df2=267) =.1.215, p = .299, 2-tailed).* This gives statistical evidence to the effect that there were no significant differences in mean scores of the tested variable. The tested hypothesis means non-significant difference existed among the years of working experience and test construction among the teachers. Hence, null hypothesis which states that, "*There is no statistically significant difference among the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality with respect to how they adhere to test construction"* was upheld. Since the differences were non-significant, post-hoc test/follow up test was not applicable.

This result is consistent with previous study conducted by Anwhere (2009) using the Tutors in the Teacher Colleges of Education in Ghana. In this study, no statistically significant difference was found among teachers' test construction practices with respect to years of teaching among Tutors in Colleges of Education. It is therefore possible that teachers irrespective of years of teaching follow similar practices when constructing test items. However, the finding here is also at discrepancy with the finding of Amedahe (1989) who found that a moderate relationship exists between number of years of teaching and the accuracy with which teachers constructed their classroom achievement tests among teachers in Senior High Schools in Cape Coast.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction

The last chapter of this study starts with a summary of the objectives of the study, its methodology and data analyses techniques. It proceeds with a summary of the key findings pertaining to each objective and the conclusions drawn from them. Specific recommendations from the findings and conclusions are made to stakeholders for decision making.

Summary

Overview of the Study

The study sought to find out if Junior High School teachers in the Sissala East Municipality adhere to the basic prescribed principles in the area of construction, administration and scoring of classroom achievement tests. The study was guided by the following research objectives:

1. Assess how Junior High Schools teachers in the Sissala East Municipality adhere to principles of test:

a. construction

b. administration

c. scoring

2. Find out the kinds of achievement test strategies Junior High Schools teachers in the Sissala East Municipality use to assess their students' learning outcomes.

3. Investigate the challenges Junior High Schools teachers in the Sissala East Municipality encounter in the use of achievement test.
4. Assess difference among the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality with respect to how they adhere to test construction.

A descriptive sample survey was conducted in Sissala East Municipality using questionnaire as the data collection instrument. Stratified proportionate sampling, random sampling and purposive sampling were used to select two hundred and forty-eight (248) Junior High School teachers from the Sissala East Municipality for the study. The analysis focused on descriptive statistics that involved computing of frequencies, percentages, means and standard deviations. The hypothesis was analysed using One Way Analysis of Variance (ANOVA).

Key Findings

The results show that generally, Junior High Schools teachers in the Sissala East Municipality averagely adhere to most principles of test construction (MM=2.46, SD=1.44). This may be as a result of some teachers relying on past questions instead of constructing the items on themselves. Some teachers copy test items directly from text books. Yet these problems may happen due to inadequate knowledge of teachers in test constructions. Because if teachers have knowledge in the principles of test constructions they will know that it is not ideal to use already constructed items to assess their students.

The results show that generally, majority of the teachers in the Sissala East Municipality averagely adhere to test administration principles in their achievement test. The researcher found out from the study that some teachers do not stay with students in the classroom when they are writing test.

The results gave evidence that most Junior High Schools teachers in the Sissala East Municipality have low average scoring abilities and this always affect the achievement test. It was revealed that most teachers look for answers to items after students take test. It was also revealed that most teachers administered test to students immediately after a lesson do not prepare the keys.

From the results, it was evident that most of the achievement test strategies were not used among Junior High Schools teachers in the Sissala East Municipality.

The study revealed that there are numerous challenges that confront the use of achievement test among Junior High Schools teachers in the Sissala East Municipality. Large class size is one of the problems that most teachers complain of. Because scoring of these test needs more energy.

From the hypothesis, the tested hypothesis suggested non-significant difference existed among the years of working experience and test construction among the teachers. Hence, null hypothesis which states that, "There is no statistically significant difference among the years of teaching experience of Junior High Schools teachers in the Sissala East Municipality with respect to how they adhere to test construction" was upheld.

Conclusions

It was evident from the findings of the study that teachers in the Sissala East Municipality were not well equipped with test construction, administration and scoring skills. Teachers having such a sensitive responsibility of assessing and making decision concerning students' academic progress are expected to be professional in the process of achievement testing strategies. However, teachers engaging in some negative test practices when constructing test items, administering as well as scoring the test items maybe that they are comfortable with such practices without recognising the impact of their practices on issues of validity and reliability.

Recommendations

With respect to the findings resulting from the study, the following recommendations are made for the improvement of testing practices among Junior High Schools teachers in the Sissala East Municipality:

1. I suggest, regular workshops and in-service training should be organised by the Ghana Education Service for teachers in Junior High Schools on how to plan achievement test (especially test construction, administration and scoring) effectively. This could be achieved through the collaboration of the ministry of education, the institute of education and other stakeholders of education.
2. Since it was evidence from the findings that teachers use paper and pen as the only strategy to assess their students, teachers are encouraged to use of other equally important assessment strategies such as observation, drama, storytelling, exhibition and presentations. This will help to assess the students as a whole.
3. There should be an intensive monitoring by headteachers and other supervisors of education on how teachers practice achievement test. Headteachers should ensure that teachers provide test specification

table for the test items they construct and make sure that other test practice principles are adhere to.

Suggestions for Future Research

The following are suggested for future research:

1. A study could be carried out to look into testing practices in terms of item analyses of objective type test of teachers.
2. A study could also be carried out to check on the interpretation of test and their consequences.
3. A study also needs to be carried out to look at the perception of teachers in testing practices and its effect on their practices.
4. The study can further be replicated to cover a wide range of population to establish the extent to which teachers in Ghana follow the basic principles of test construction, administration and scoring.

REFERENCES

Ackerman, T. A., & Smith, P. L. (1988). A comparison of the information provided by essay, multiple-choice and free response writing test. *Applied Psychological Measurement, 12*, 117-128.

Adamolekun, J. (2012). Effective test construction: A case for the junior senior secondary school mathematics curricula. *Journal of the School of Pure Science, 1*(3), 197-224.

Alan, P. (2000). *Human resource management in a business context*. Halifax, Canada: Canale & Co. Publishers.

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, Ghana.

Amedahe, F. K. (2001). *Fundamentals of educational research methods*. (Unpublished class notes). University of Cape Coast

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

American Educational Research Association [AERA], (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Anastasi, A. (1982). *Psychological testing*. New York, NY: Macmillan Publishing Company.

Anhwere, Y. M. (2009). *Assessment practices of teacher training college tutors in Ghana.* (Unpublished Master's thesis). University of Cape Coast. Cape Coast Ghana.

Archbald, D. A. (1991). Authentic assessment: Principles, practices, and issues. *School Psychology Quarterly*, *6*(4), 279-287.

AREA/APA/NCME (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Bartels, E. K. (2003). The practice of continuous assessment in teacher training colleges in Ghana. *Journal of Educational Development and Practice*, *1*(1), 59-72.

Bejar, I. (1984). Educational diagnostic assessment. *Journal of Educational Measurement*, *2*(1), 175-189.

Benson, J. (1998). Developing a strong program of construct validation: a test anxiety example. *Educational Measurement: Issues and Practice, 17,* 10-17.

Bernstan, S. (1953). Assessment of teaching. In W. R. Houston (Ed.), *Handbook of research on teacher education: A project of the Association of Teacher Educators* (pp. 569 – 598). New York, NY: Macmillan.

Binet, A. (1905). New methods for the diagnosis of the intellectual level of subnormals. L'Annee Psychologique, 12, 191–244. In E. S. Kite (Ed.), *The development of intelligence in children* (pp. 15-35). Vineland, NJ: Publications of the Training School at Vineland.

Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, *40*(1), 90-98.

Boakye, M. E. (2016). *Fair testing practices in district-mandated testing programme in the Ashanti region of Ghana*. (Unpublished master's thesis). University of Cape Coast, Cape Coast, Ghana.

Bollen, P. J. (1989). Formative and summative assessment by teachers. *Studies in Science Education*, *21*(3), 49 – 97.

Brown, J. (2007). Feedback: the learner perspective. *Research in Post-Compulsory Education*, *12*(1), 33 – 51.

Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (p. 709–725). Macmillan Library Reference USA; Prentice Hall International.

Chester, C. & Quilter, S. M. (1998). Inservice teachers' perceptions of educational assessment. *Journal for Research in mathematics Education*, 33(*2*), 210-236

Cohen, L., Manion, L. & Morrison, (2011). *Research methods in education* (5th ed.). London, UK: Routledge Falmer.

Creswell, J. W. (2012). *Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson Education Inc.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Ohio, US: Language learning.

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

Cronbach, L. J., Corno, L., Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W. & Talbert, J. E. (2001). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Mahwah, N.J: Erlbaum.

Crossman, A. (2017). Understanding purposive sampling. *Retrieved July, 31*, 2017.

Cunningham, G. K. (2001). *Educational and psychological measurement.* New York, NY: Macmillan Publishing Company.

Dambudzo A. (2009). Adult learners' performance on standardized and non-standardized test as a function of sex and location. *Nigerian J. Counsel. Dev*, *4*(1), 32-36.

Davis, G. (1991). *Constructivist learning theory*: *Institute for Inquiry*. Retrieved from http://www. exploratorium.edu /ifi/resources/ constructivistlearning.htmlS.

Dawson, A. (2002). *Introduction to research in education* (2nd ed.). Forth Worth, TX: Holt, Rinehart & Winston Inc.

Delandshere, G. (1996). From static and prescribed to dynamic and principled assessment of teaching. *Elementary School Journal*, *97*(2), 106-120.

Diamantopoulos, E. (2004). Hilbert matrix on Bergman spaces. *Illinois Journal of Mathematics*, *48*(3), 1067-1078.

DuBois, P. H. (1970). *A history of psychological testing.* Boston, MA: Allyn and Bacon Inc.

Ebel, L. R., & Frisbie, A. D. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Education and National Education Association (1990). *Standards for teacher competence in education assessment of students.* Washington, DC: National Council on Measurement in Education.

Educational Testing Service (ETS). (2014). *ETS standards for quality and fairness.* Princeton, NJ: Author.

Etsey, Y. K. (2004). Assessing performance in schools: Issues and practice. *Ife* Psychologia *13*(1), 123-135.

Etsey, Y. K. A. (2004). *Educational measurement and evaluation.* Lecture notes on EPS 203. (Unpublished document). University of Cape Coast, Ghana. Department of Educational Foundations, University of Cape Coast, Cape Coast.

Etsey, Y. K. A. (2012). *Assessment in education*. (Unpublished manuscripts). University of Cape Coast, Cape Coast, Ghana.

Flanagan, D, Genshaft, J. L., & Harrison, P. L. (1997). *Intellectual assessment, tests, and issues*. New York, NY: The Guilford Press.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. *New Directions for Testing and Measurement, 19(*3), 29-38.

Fook, C. Y., & Sidhu, G. K. (2010). Authentic assessment and pedagogical strategies in higher education. *Journal of Social Sciences*, *6*(2), 153-161.

Fox, L. H., & Soller, J. F. (2001). Psychosocial dimensions of gender differences in mathematics. *Changing the faces of mathematics: Perspectives on gender*: + En Jacobs, J, 9-24.

Frankel, J. R., & Wallen, N. E. (2009). *Single-subject research*. *How to design and evaluate research in education.* (7th ed.). New York, NY: McGraw-Hill.

Freeman, H. G., & Lewis S. (2008). *Theory and practice of psychological testing*. New Delhi, India: Oxford & Ibh Publishing.

Gay, R. L. (2006). *Educational research: Competencies for analysis and application* (3rd ed). New York, NY: Macmillan Publishing Company.

George, J., P. (2002). *Fundamental statistics in psychology and education.* (7th ed). Tokyo, Japan: Macmillan Publishing Company.

Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. In *Optimising new modes of assessment: In search of qualities and standards* (pp. 37-54). Dordrecht, The Netherlands: Kluwer Academic Publisher.

Gipps, C. (1992a). *National testing at seven: What can it tells us?* London, UK: Hodder and Stoughton Publishing Company.

Gipps, C. (1992b). *What we know about effective primary teaching*. London, UK: Tufnell Press.

Gipps, C., & Stobart, G. (2009). *Fairness in assessment*. In C. Wyatt-Smith & J. Cumming (Eds.), *Educational assessment in 21st century: Connecting theory and practice* (pp. 105-118). Netherlands: Springer Science Business Media.

Gipps, C., J., Brown, M., McCallum, B., & McAlister, S. (1995) *Imitation or evidence? Teachers and national assessment of seven-year-olds.* Washington, DC: American Educational Research Association.

Green, S. B. & Neil, J. S. (2014). *Using spss for windows and macintosh: Analyzing and understanding data* (7th ed.). New York, NY: Pearson

Gregory, R. J. (1992). *Psychological testing: History, principles and applications.* Boston: Allyn and Bacon.

Gronlund, N. E. (1986). *How to construct achievement tests?* (1st ed.) Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gronlund, N. E. (2008). *How to construct achievement tests?* (3rd ed.) Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gronlund, N. E. (2012). *How to construct achievement tests?* (4th ed.) Englewood Cliffs, NJ: Prentice-Hall, Inc.

Groth-Marnat, G. (1997). *Handbook on psychological assessment* (3rd ed.). New York: John Wiley & Sons.

Gullickson, A., & Moen, J. (2001, March). *The use of stochastic methods in local area population forecasts.* Annual meeting of the Population Association of America, Washington DC.

Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. *Assessment and Learning*, *2*, 95-110.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297–313.

Hoste, R. N., & Bloomtreld F. L. (1975). *Staff development implications from a state-wide assessment of classroom teachers' testing skills and practice.* (ED312309). Retrieved from http://eric.ed.gov/ERICWebPortal.

Izard, J. (2005). Overview of test construction. In K. N. Ross (Ed) *Quantitative research methods in educational planning*. Paris, France: UNESCO International Institute for Educational Planning.

James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, *17*(2), 109-138.

Johnson, R. N., & Synyby, F. L. (2007). *A summary of published research: Classroom teachers' knowledge and skills related to the development* *and use of teacher-made tests.* (ED346148). Retrieved from http://eric.ed.gov/ERICWebPortal/home. Online.

Joint Committee on Standards for Educational and Psychological Testing (JCSEPT). (1999). *Standards for educational and psychological testing.* Washington DC: American.

Kankam, B., Bordoh, A., Eshun, I., Bassaw, T. K., & Korang, F. Y. (2014). Teachers' perception of authentic assessment techniques practice in social studies lessons in senior high schools in Ghana. *International Journal of Educational Research and Information Science, 1*(4), 62-68.

Karpicke, J. D., & Roediger, H. L., (2008). The critical importance of retrieval for learning. *Science*, *15*(9), 966 –968.

Krejcie, R.V., & Morgan, D.W. (1970). *Determining sample size for research activities.* Washington D.C: American

Kubiszyn, T. & Borich, G. (1984). *Educational testing and measurement: Classroom application and practice*. Glenview, US: Scott and Foresman Company.

LaFontana, K. M., & Cillessen, A. H. (2002). Children's perceptions of popular and unpopular peers: A multimethod assessment. *Developmental Psychology*, *38*(5), 635.

Liesbet, H. O. O. G. H. E., & Gary, M. (2003). Unraveling the central state, but how? Types of multi-level governance. *American Political Science Review*, *97*(2), 233-243.

Lissitz, R. W., & Schafer, W. D. (2002). *What role will assessment play in school in the future: Assessment in Educational Reform.* Both Means and Ends. Boston, MA: Allyn and Bacon.

Maizan, K. K. (2005). *Districts pare 'electives' for core courses. Education Week.* Retrieved from http://www.edweek.org/ew/articles5stand.h16.

Maree, K. (2007). *First steps in research*. Pretoria: Van Schaik Publishers.

Mbano, N. M. (2003). The effects of cognitive development, age and gender on the performance of secondary school pupils in science and other subjects. *Malawi Journal of Development Education*, *1*(1), 55-76.

McDaniel, E. (1994). *Understanding educational measurement*. Madison, WI: Brown and Benchmark Publishers.

McMillan, J. H., & Sehumocher, L. R. (2001). *Secondary science teachers' classroom assessment and grading practices.* Richmond, VA: (ERIC Document Reproduction Service N. Ed 450 158)

McNabb, D. E. (2004). *Research methods in public administration and nonprofit management*. Routledge.

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

Messick, S. (2003). The psychology of educational measurement. *Journal of* Educational *Measurement*, *21*(4), 215-237.

Miller, A. L, Mclntire, A. S., & Lovler, L. R. (2011). *Foundation of psychological testing* (3rd ed.). Washington, DC: Sage Publications Inc.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, *62*, 229–258.

Mugenda, O. M., & Mugenda, A. G. (2003). *Research methods*. *Quantitative and qualitative approaches*. Nairobi, Kenya: African Center for Technology Studies.

Munson, R., & Parton, C. (2013). *Bias and fairness in state testing*. Retrieved from http://apps.leg.wa.gov/billinfo/summary.aspxl=1450&year=2013

Musphy, C. T. (2004). *Nursing research: Principles and methods*. Philadelphia, PA: Lippincott Williams & Wilkins.

Mussawy, S. A. J. (2009). *Assessment practices: Student's and teachers' perceptions of classroom assessment*. Retrieved from http://apps.leg.wa.gov/billinfo/summary.aspx?bill=1450&year=2013

Namporta, T., & Wella, E. (1999). *Testing in modern classrooms*. London, UK: George Allen and Urwin Ltd.

National Association of School Psychologists (NASP). (2002). *Large-scale assessments and high stakes decisions: Facts, cautions and guidelines.* Bethesda, MD: Author.

National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing.* Washington, D. C: Author.

Neumann, A. (2006). Professing passion: Emotion in the scholarship of professors at research universities. *American Educational Research Journal*, *43*(3), 381-424.

Newmann, F., & Wehlage, G. (1995). *Successful school restructuring: A Report to the Public and Educators by the Center on Organization and Restructuring of Schools.* Madison, WI: University of Wisconsin, Center on Organization and Restructuring Schools.

Nitko, A. J. (2001). *Educational tests and measurements* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Nitko, A. J., & Brookhart, S. M. (1996). *Educational assessment of students*. Englewood Cliffs, NJ: Merrill.

Nunnally, J. C. (1964). *Educational measurement and evaluation.* New York, NY: McGraw-Hill.

Oduro, O. G. (2008). *Testing practices of senior secondary school teachers in the Ashanti Region of Ghana.* (Unpublished master's thesis). University of Cape Coast, Cape Coast, Ghana.

Osuola, E. C. (2001). *Introduction to research methodology*. (3rd ed.). Onitsha, Nigeria: Africana F.E.P Publishers Ltd.

Pallant, J. (2007). SPSS survival manual: A step by step guide to data analysis using SPSS for Windows (Version 10). Sydney, Australia: Allen and Unwin.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). New Delhi, India: Sage.

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

Perry, N. E., & Meisels, S. J. (1996). *How accurate are teacher judgements of students' academic performance? (*Working Paper No. 96-08). Washington, DC: National Center for Educational Statistics.

Poikela, E. (2004). Developing criteria for knowing and learning at work: towards context-based assessment. *Journal of Workplace Learning, 16*(5), 267-274.

Quagrain, A. K. (1992). *Teacher-competence in the use of essay type tests: A study of the secondary schools in the Western Region of Ghana*. (Unpublished thesis). University of Cape Coast, Cape coast Ghana.

Reeves, D. B. (2003). The learning leader/looking deeper into the data. *Educational Leadership, 66*(4), 89-90.

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment in special and remedial education* (3rd ed.). Boston, MA: Houghton Mifflin.

Sarantakos, S. (2005). *Social research* (3rd ed.). Melbourne, Australia: MacMillan Education.

Sasu, O. E. (2017). *Testing practices of junior school teachers in the Cape Coast Metropolis*. (Unpublished master's thesis). University of Cape Coast, Cape Coast, Ghana.

Slijepcevic, A., Tolhurst, K. G., Saunder, G., Whight, S., & Marsden-Smedley, J. B. (2007, September). A prescribed burning risk assessment tool (BRAT). *Proceedings Bushfire Cooperative Research Centre and Australasian Fire Authorities Council Annual Conference. Tassie Fire Conference, Wrest Point Conference Centre, Hobart, Tasmania*.

Stainback, W., & Stainback, S. (1996). *Controversial issues in special education. Divergent perspectives* (2nd ed.). Boston, MA: Ally and Bacon Press.

Starch, D., & Elliot, E. (1912). Reliability in grading high school work in English. *School Review, 20,* 442-457.

Tamakloe, E. K., & Amedahe, F. K. (1996). *Principles and methods of teaching*. Accra, Ghana: Black Mask.

Tamakloe, E. K., Atta, E. T., & Amedahe, F. K. (1986). *Principles and methods of teaching*. Accra, Ghana: Black Mask.

Tinkelman, S. N. (1971). Planning the objective test. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 46-80). Washington, DC: American Council on Education.

Tom, K., & Gary, D. B. (2003). *Educational testing and measurement: Classroom application and Practice.* Hoboken, NJ: John Willey and Sons Inc.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, *27*(5), 615-631.

Trochim, W. M. K. (2006). *Introduction to validity*. *Social research methods*. Retrieved from www.socialresearchmethods.net/kb/introval.php,

Van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism practice*, *6*(5-6), 648-658.

Vigotsky, L. V. (1962). Thought and Language. Cambridge, MA: MIT Press, 1962

Wiliam, D. (1928). Validity, dependability and reliability in national curriculum assessment. *The Curriculum Journal*, *4*(3), 335-350.

Wiliam, D. (2008). Quality in assessment. In S. Swaffield, & M. Williams, (Eds.), *Unlocking assessment: Understanding for reflection and application* (pp. 123-137). London, UK: David Fulton.

Willington, F. L (2000). *Incontinence in the elderly*. Academic Press.

Winograd, P., & Perkins, F. D. (1997). Authentic assessment in the classroom: Principles and practices. *Journal of Education, 9*(94), 1-8.

Wood, J. G. (2007). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of research in education* (pp. 11-16). Washington, DC.: American Educational Research Association.

Xiaomei, S. (2014). *Test fairness in a large-scale high-stakes language test.* (Unpublished doctoral thesis). Queen's University, Ontario, Canada.

Zucker, S. (2004). *Administration practices for standardized assessments: Assessment report*. San Antonio, TX: Pearson Education Inc.

APPENDICES

APPENDIX A

UNIVERSITY OF CAPE COAST

COLLEGE OF EDUCATION STUDIES

DEPARTMENT OF EDUCATION AND PSYCHOLOGY

QUESTIONNAIRE FOR BASIC SCHOOL TEACHERS IN THE SISSALA EAST MUNICIPALITY

Dear Respondent

The study seeks to **Assess Achievement Test Practices of Teachers in Junior High Schools in the Sissala East Municipality**.

Your full input will help make informed decisions about **Achievement Test Practices**. It would therefore be appreciated if you could provide responses to **all** items on the questionnaire, and do it **honestly**. You are assured of complete **confidentiality** and **anonymity** of all information provided. **Nothing** will ever be published or reported that will associate your name and/or school with your responses to the survey questions. Therefore, you **should not** write your name, and/or school name on any part of the instrument. Your participation in this study is **completely voluntary**. Again, questions on this survey instrument have gone through a thorough review by professionals at the University of Cape Coast, and have been declared **ethical** for educational research. You hereby consent to voluntarily participate in this study by providing responses to items of the various sections of this instrument. Thank You.

SECTION A

DEMOGRAPHIC CHARACTERISTICS

1. **Gender**:

    a) Male [ ]

b) Female [ ]

2. Number of years in teaching service

a) Under 5 years [ ]

b) 6 – 10 years [ ]

c) Above 11 years [ ]

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 [ ]

g) Others, specify……………………………………………

SECTION B

TEACHERS KNOWLEDGE ABOUT CONSTRUCTION OF ACHIEVEMENT TEST

Please respond to the following statements on your knowledge about **Construction of Achievement Testing**. Indicate the extent to which you Strongly Agree-SA, Agree-A, Disagree-D and Strongly Disagree-SD to the statements below

Directions: Indicate with a tick [√] your level of knowledge in Construction of Achievement Test. Where: ***SA = Strongly Agree, (4), A = Agree, (3) D = Disagree, (2) and SD = Strongly Disagree (1)***

	Teachers should do the following when	**SA**	**A**	**D**	**SD**

	constructing test items				
4	State the purpose of the test				
5	Specify the construct to be measured				
6	Use a test specification table				
7	Match learning outcomes to the items				
8	Construct test items when it is time to assess				
9	Set questions from past questions				
10	Use questions directly from text books				
11	Ask any other colleagues to help me construct test items				
12	Ask colleagues in subject area to review test items				
13	Prepare marking scheme after students have answered the question(s)				
14	Consider meaning of wording against different ethnic background				
15	Consider students' language proficiency				
16	Consider variation of students with respect to physical disability				
17	Consider the time individual will spend on a question				
18	Try solving the questions myself to determine the time required				
19	Provide clear and simple instructions on how				

	test is to be answered				
20	Evaluate test items given to the students				
21	Write test items at least two weeks before time				
22	Write more test items than needed				
23	Follow the principles of test construction for each format				

SECTION C

TEACHERS KNOWLEDGE ABOUT ADMINISTRATION OF ACHIEVEMENT TEST

Please respond to the following statements on your knowledge about **Administration of Achievement Testing**. Indicate the extent to which you Strongly Agree-SA, Agree-A, Disagree-D and Strongly Disagree-SD to the statements below

Directions: Indicate with a tick [√] your level of knowledge in Administration of Achievement Testing. Where: ***SA = Strongly Agree, (4), A = Agree, (3) D = Disagree, (2) and SD = Strongly Disagree (1)***

	In administration of test items, I ……	SA	A	D	SD
24	make students aware of the rules and regulations covering the test				
25	make room for adequate ventilation and lighting				
26	make provision for extra sheets and writing materials				
27	allow students to start and stop test on time				
28	give more instructions during the time the students are taking the test				
29	inform students in advance areas for the test				

30	prepare classroom a day before test is taken				
31	test students after long vacations or important holidays				
32	inform student about the test format				
33	make provision for emergencies during the time the test is taken				
34	proof read all test items				
35	use "DO NOT DISTURB SIGN" at the entrance of classroom				

SECTION D

TEACHERS KNOWLEDGE ABOUT SCORING OF ACHIEVEMENT TEST

Please respond to the following statements on your knowledge about **Scoring of Achievement Test**. Indicate the extent to which you Strongly Agree-SA, Agree-A, Disagree-D and Strongly Disagree-SD to the statements below

Directions: Indicate with a tick [√] your level of knowledge of Scoring of Achievement Test. Where: ***SA = Strongly Agree, (4), A = Agree, (3) D = Disagree, (2) and SD = Strongly Disagree (1)***

	In scoring test items, I…….	SA	A	D	SD
36	mark papers just after the test is taken				
37	prepare scoring guide				
38	make sure test takers are kept anonymous				
39	grade the responses item by item				
40	keep scores of previous items out of sight				
41	periodically rescore previously scored items				
42	shuffle scripts before scoring				
43	score essay test when I am physically sound and mentally alert in a sound environment				
44	constantly follow scoring guide				

45	am influenced by the first few papers read when scoring test items				
46	score a particular item on all papers at a sitting				
47	provide comments and errors correct on scripts				
48	give extra marks to students based on Handwriting, Gender etc.				

SECTION E

KINDS OF ACHIEVEMENT TESTING FORMAT

Please respond to the following statements on your knowledge about **Kinds of Achievement Testing Format**. Indicate the extent to which you Very Often-VO, Often-O, Sometimes-S and Never-N to the statements below**. Directions**: Indicate with a tick [√] your level of knowledge in **Achievement Testing Format**.

	Kinds of Achievement Testing Format	**VO**	**O**	**S**	**N**
49	Assessing work samples				
50	role-play				
51	Constructed-Response Items				
52	Experiments/Demonstrations				
53	Projects				
54	Exhibitions				
55	Writing Samples				
56	Story Telling				
57	Presentations				

58	Drama				
59	Report writing				
60	Computer simulation task				

SECTION F

CHALLENGES OF ACHIEVEMENT TEST

Please respond to the following statements on your knowledge about **The Challenges That You Encounter in Using Achievement Testing.** Indicate the extent to which you Strongly Agree-SA, Agree-A, Disagree-D and Strongly Disagree-SD to the statements below

Directions: Indicate with a tick [√] your level of knowledge on the challenges of achievement test. Where: ***SA = Strongly Agree, (4), A = Agree, (3) D = Disagree, (2) and SD = Strongly Disagree (1)***

	Challenges	**SA**	**A**	**D**	**SD**
61	The school assessment system makes it difficult to use achievement testing				
62	Lack of funds to embark on some activities and projects				
63	Lack of support from the school authorities in terms of logistics and facilities				
64	Lack of motivation from school authorities				
65	Developing achievement testing task is difficult				
66	Inadequate time allotted on the timetable for various subjects does not permit the use of achievement testing.				
67	Inadequate time to prepare in terms of gathering information and materials to be used for achievement testing.				
68	Large class size makes it difficult to assess students using achievement testing.				

69	Some topics are difficult to be assessed using achievement testing				

APPENDIX B

RELIABILITY TEST RESULTS OF THE INSTRUMENT

Case Processing Summary

		N	%
Cases	Valid	30	100.0
	Excluded[a]	0	.0
	Total	30	100.0

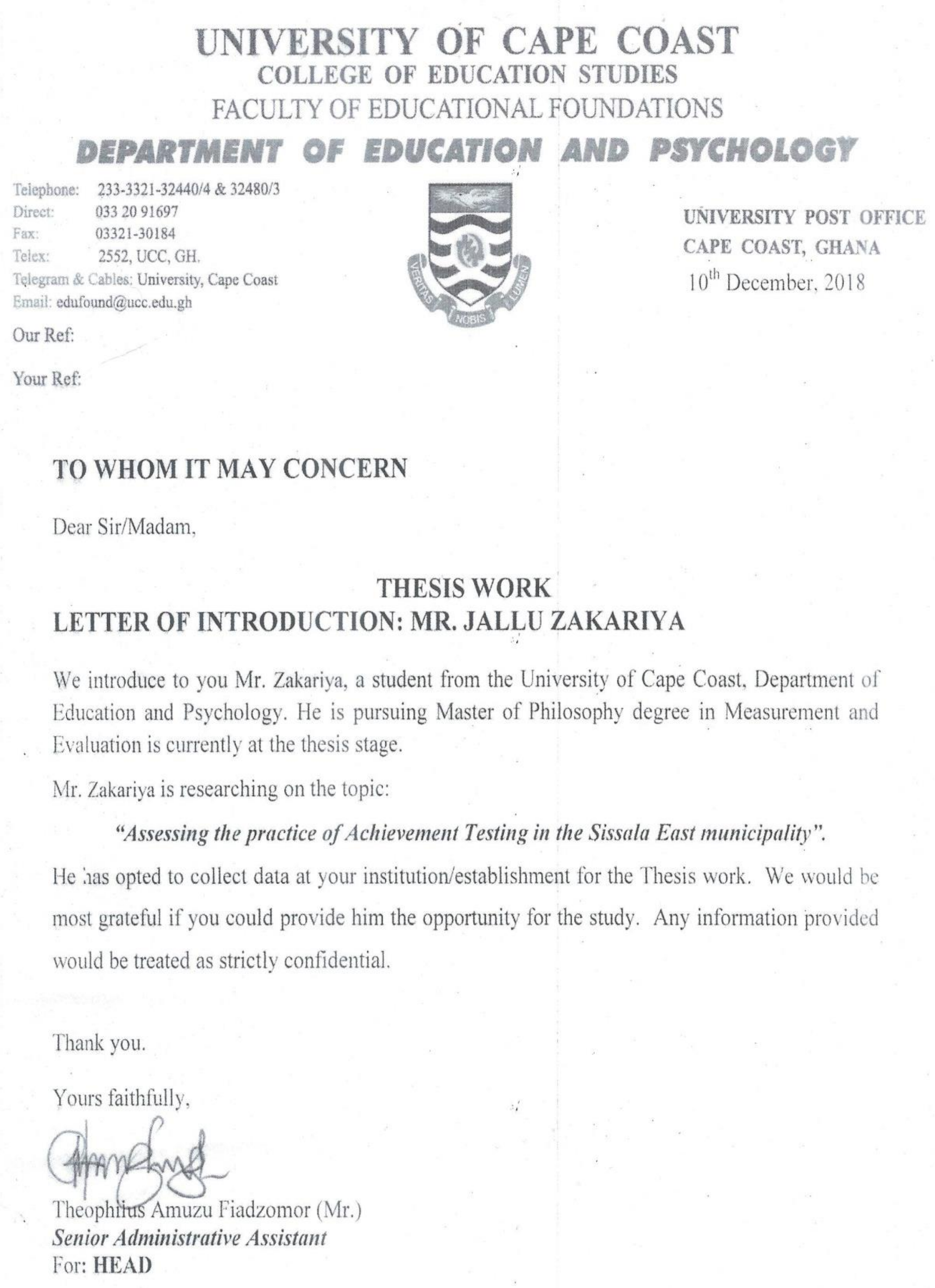
a. List wise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.806	69

APPENDIX C

INTRODUCTORY LETTER



UNIVERSITY OF CAPE COAST
COLLEGE OF EDUCATION STUDIES
FACULTY OF EDUCATIONAL FOUNDATIONS
DEPARTMENT OF EDUCATION AND PSYCHOLOGY

Telephone: 233-3321-32440/4 & 32480/3
Direct: 033 20 91697
Fax: 03321-30184
Telex: 2552, UCC, GH.
Telegram & Cables: University, Cape Coast
Email: edufound@ucc.edu.gh

VERITAS NOBIS LUMEN

UNIVERSITY POST OFFICE
CAPE COAST, GHANA
10th December, 2018

Our Ref:

Your Ref:

TO WHOM IT MAY CONCERN

Dear Sir/Madam,

THESIS WORK
LETTER OF INTRODUCTION: MR. JALLU ZAKARIYA

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

Mr. Zakariya is researching on the topic:

"Assessing the practice of Achievement Testing in the Sissala East municipality".

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

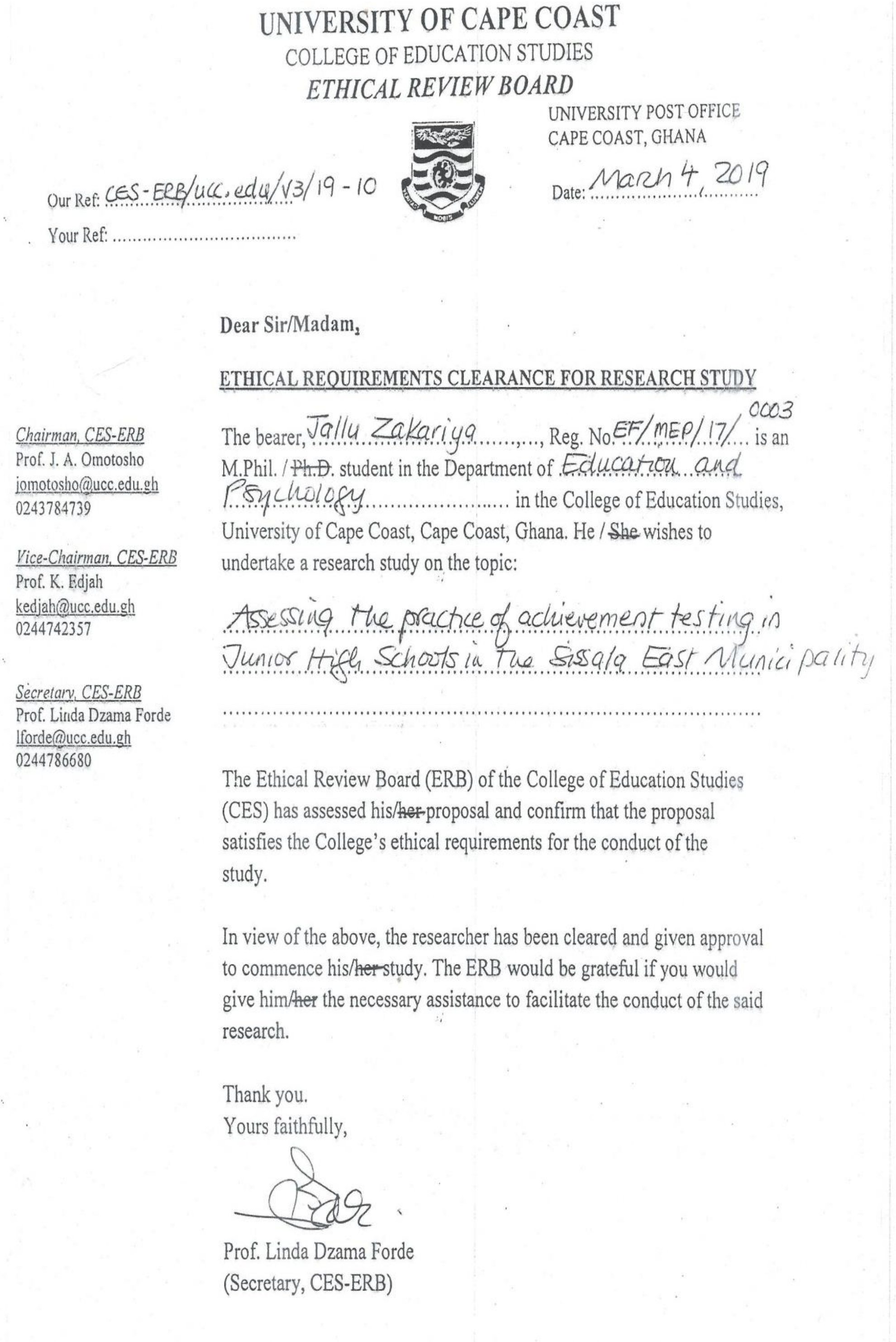
Thank you.

Yours faithfully,

Theophilus Amuzu Fiadzomor (Mr.)
Senior Administrative Assistant
For: **HEAD**

APPENDIX D

ETHICAL CLEARANCE



UNIVERSITY OF CAPE COAST
COLLEGE OF EDUCATION STUDIES
ETHICAL REVIEW BOARD

UNIVERSITY POST OFFICE
CAPE COAST, GHANA

Our Ref: CES-ERB/ucc.edu/v3/19-10

Date: March 4, 2019

Your Ref: ……………………………

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

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

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

Dear Sir/Madam,

ETHICAL REQUIREMENTS CLEARANCE FOR RESEARCH STUDY

The bearer, Jallu Zakariya……, Reg. No EF/MEP/17/0003 is an M.Phil. / ~~Ph.D.~~ student in the Department of Education and Psychology…………………… in the College of Education Studies, University of Cape Coast, Cape Coast, Ghana. He / ~~She~~ wishes to undertake a research study on the topic:

Assessing the practice of achievement testing in Junior High Schools in The Sissala East Municipality
……………………………………………………………………………

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

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

Thank you.
Yours faithfully,

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