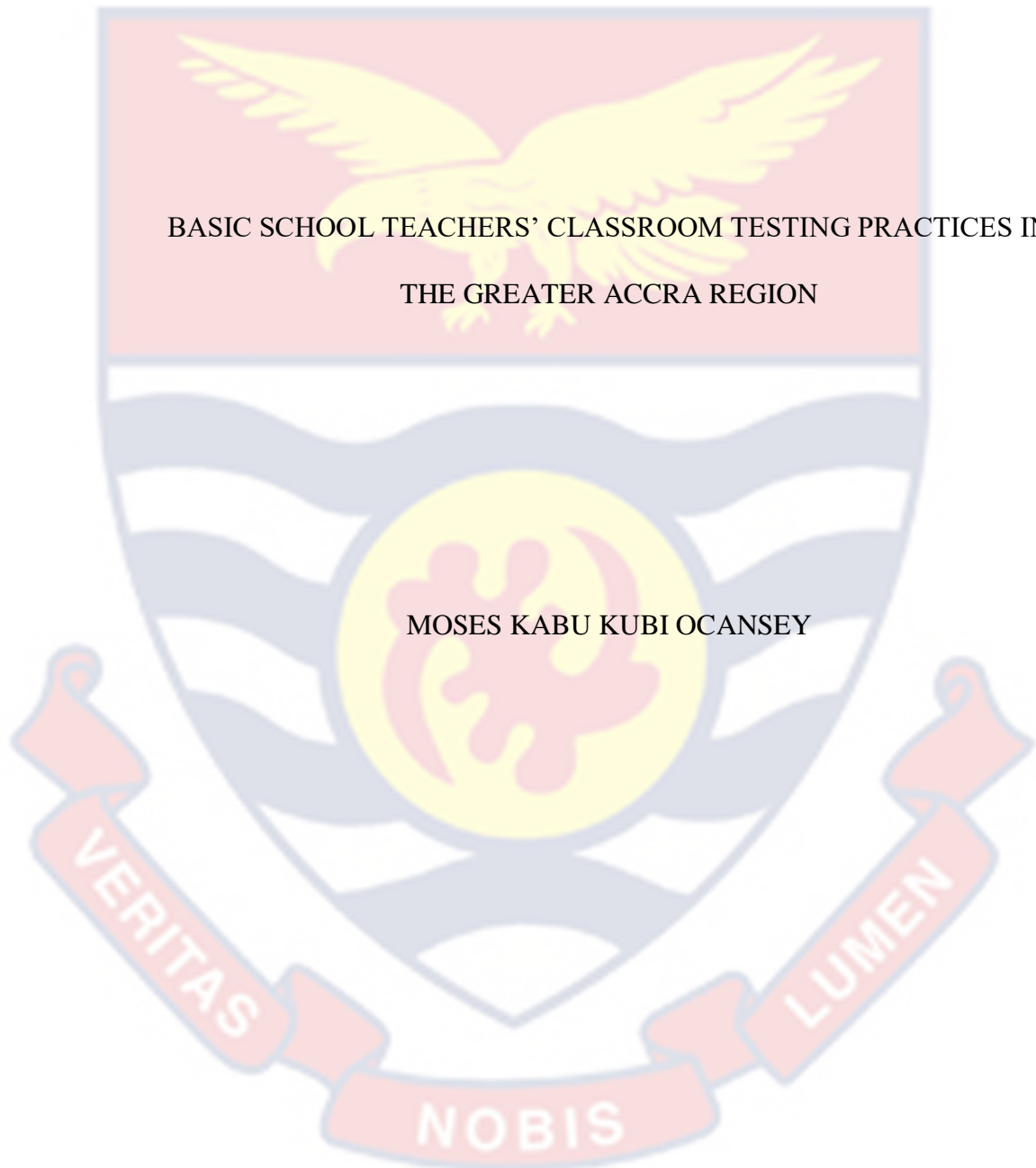


UNIVERSITY OF CAPE COAST

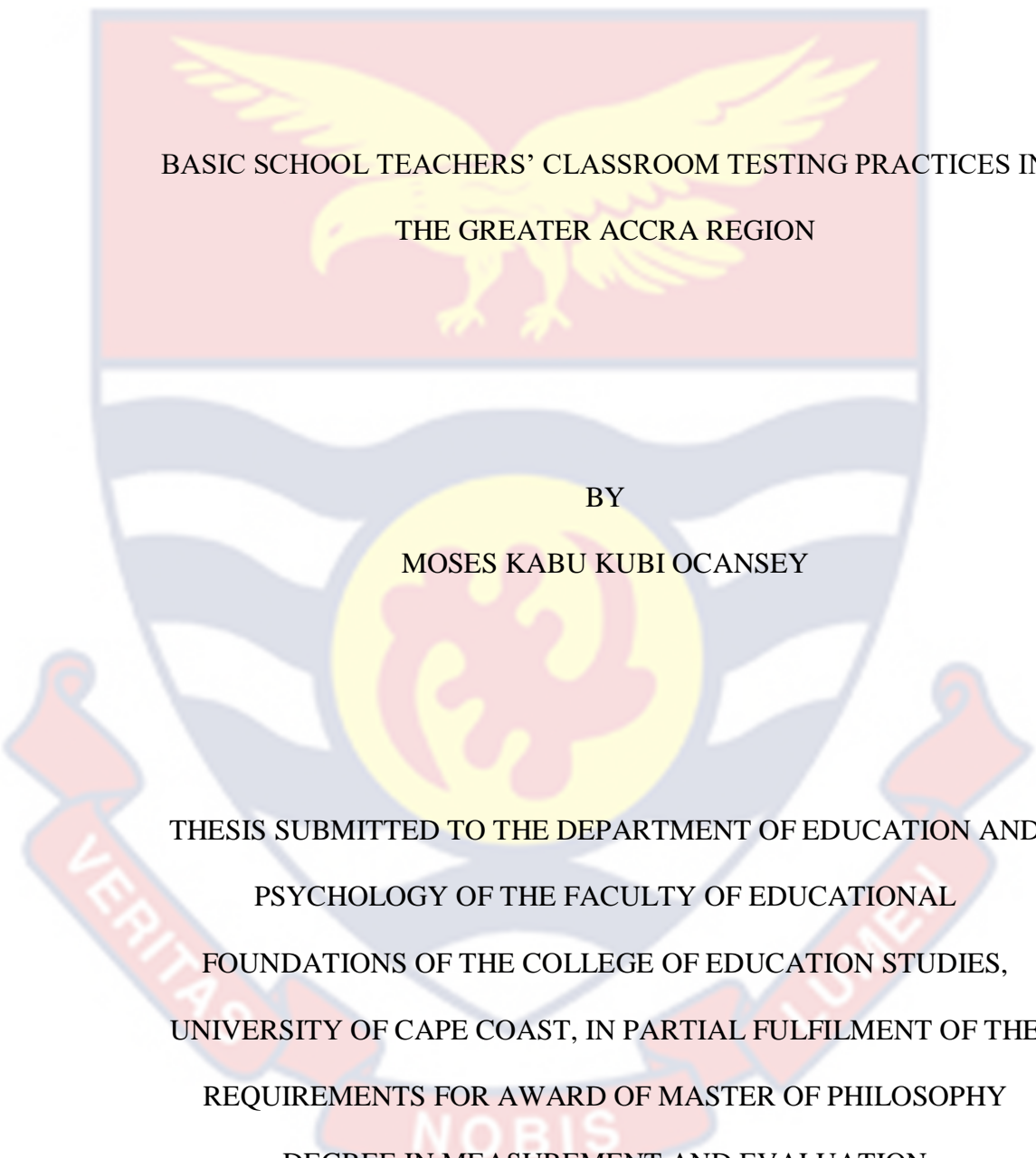


BASIC SCHOOL TEACHERS' CLASSROOM TESTING PRACTICES IN  
THE GREATER ACCRA REGION

MOSES KABU KUBI OCANSEY

2021

UNIVERSITY OF CAPE COAST



BASIC SCHOOL TEACHERS' CLASSROOM TESTING PRACTICES IN  
THE GREATER ACCRA REGION

BY  
MOSES KABU KUBI OCANSEY

THESIS SUBMITTED TO THE DEPARTMENT OF EDUCATION AND  
PSYCHOLOGY OF THE FACULTY OF EDUCATIONAL  
FOUNDATIONS OF THE COLLEGE OF EDUCATION STUDIES,  
UNIVERSITY OF CAPE COAST, IN PARTIAL FULFILMENT OF THE  
REQUIREMENTS FOR AWARD OF MASTER OF PHILOSOPHY  
DEGREE IN MEASUREMENT AND EVALUATION

DECEMBER 2021



DECLARATION

**Candidate's Declaration**

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

Candidate's Signature..... Date.....

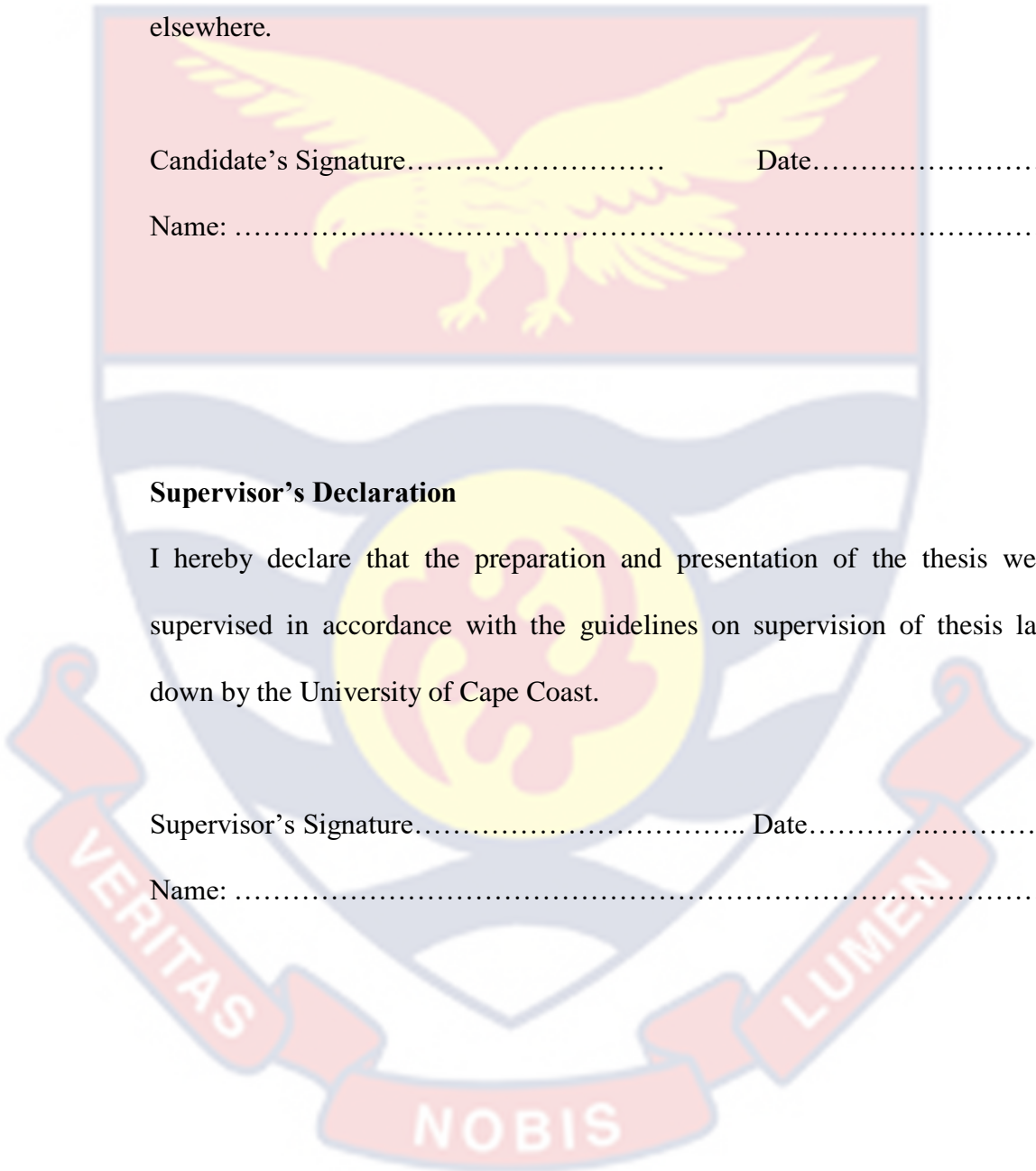
Name: .....

**Supervisor's Declaration**

I 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.

Supervisor's Signature..... Date.....

Name: .....



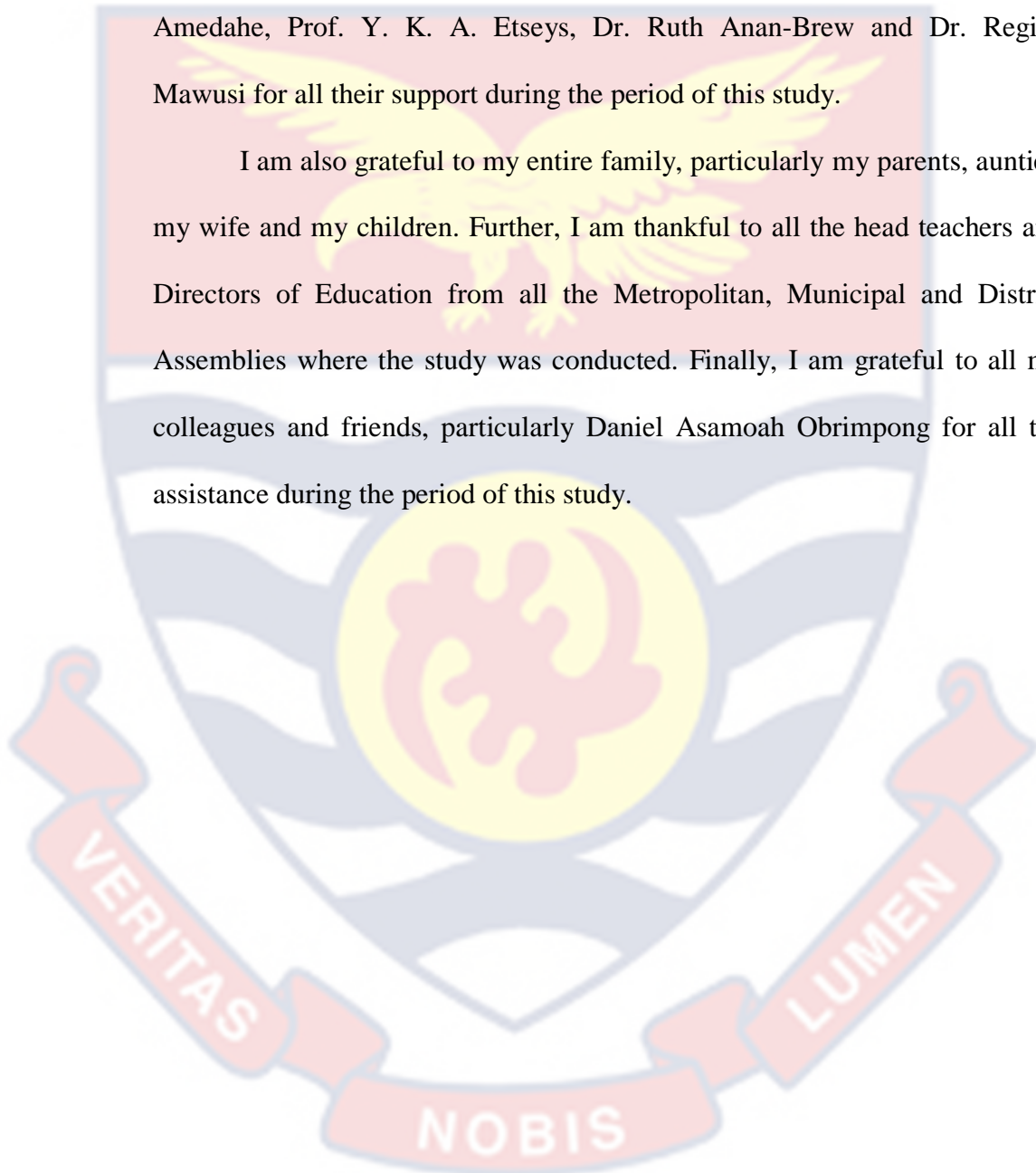
## ABSTRACT

The purpose of this study was to investigate basic school teachers' testing practice within the Greater Accra Region of Ghana. Mixed methods approach using the descriptive survey design was adopted for the study. A sample of 382 teachers was used for the study selected via stratified random sampling from a population of junior high school teachers in the Greater Accra Region of Ghana. Data were collected using questionnaire and interview guide. Quantitative data were analysed using both descriptive and inferential statistics. Qualitative data on the other hand were analysed using thematic analysis. The study revealed that in the construction of achievement test, the teachers determined the purpose of a test before constructing test items, related the instructional objectives of the subject matter to the test and selected the test format suitable for testing stated objectives. The study also found that in test administration, the teachers often made them aware of when the test will be given, the rules and regulations governing the conduct of the test and the format of the test and ensured that sitting arrangement allowed enough space to prevent malpractices. Regarding the scoring of essay tests, the study revealed that majority of the teachers used the analytic method in scoring essay tests, constantly followed the marking scheme and scored essay tests only when they were physically sound and mentally alert. Based on the findings, it was recommended that teacher training institutions make testing and measurement courses compulsory while at the same time regional education offices organise intermittent testing workshops for teachers to improve their testing practices.

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Prof. Eric Anane for his professional guidance, advice, encouragement and support in carrying out this study. I am also grateful to my lecturers, Prof. Francis Amedahe, Prof. Y. K. A. Etseys, Dr. Ruth Anan-Brew and Dr. Regina Mawusi for all their support during the period of this study.

I am also grateful to my entire family, particularly my parents, aunts, my wife and my children. Further, I am thankful to all the head teachers and Directors of Education from all the Metropolitan, Municipal and District Assemblies where the study was conducted. Finally, I am grateful to all my colleagues and friends, particularly Daniel Asamoah Obrimpong for all the assistance during the period of this study.



DEDICATION

To my wife, Faustina Mamle Ameko and children, Honi Ocansey, Haniel

Ocansey, Hanet Ocansey and Hansel Ocansey.



## TABLE OF CONTENTS

	Page
DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENTS	v
DEDICATION	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xii
<b>CHAPTER ONE: INTRODUCTION</b>	
Background to the Study	1
Statement of the Problem	7
Purpose of the Study	9
Research Questions	10
Hypotheses	10
Significance of the Study	11
Delimitation of the Study	12
Limitations of the Study	13
Definition of Terms	13
Organisation of the Study	13
<b>CHAPTER TWO: LITERATURE REVIEW</b>	
Theoretical Review	15
Classical Test Theory (CTT)	15
Item Response Theory (IRT)	18
Conceptual Review	20
Concept of Assessment	20



Relevance of Assessment	24
Concept of Testing	25
Stages in Testing Practices	27
Classroom-based Achievement Tests	29
Principles of Constructing Classroom-based Tests	32
Validation of Tests	33
Validity	33
Kinds of validity evidence	34
Reliability	35
Methods of estimating reliability	36
Administration of Classroom Achievement Tests	37
Scoring of Essay Tests	40
Interpretation of Test Results	42
Criterion-Referenced Interpretation (CRI)	42
Norm-Referenced Interpretation (NRI)	44
Empirical Review	45
Testing Practices Teachers Follow when Constructing Achievement Test Items	45
Testing Practices Teachers Follow when Administering Achievement Test Items	51
Testing Practices Teachers Follow when Scoring Essay-Type Test Items	53
Differences in Testing Practices Based on Demographic Characteristics (gender, teaching experience, professional qualification, and knowledge of test construction)	56
Chapter Summary	60

## CHAPTER THREE: RESEARCH METHODS

Research Approach and Design	61
------------------------------	----

Study Area	62
------------	----

Population	62
------------	----

Sample and Sampling Procedures	63
--------------------------------	----

Data Collection Instrument	65
----------------------------	----

Validity of Questionnaire	67
---------------------------	----

Reliability of Questionnaire	67
------------------------------	----

Pilot testing	67
---------------	----

Data Trustworthiness of Qualitative Data	68
--	----

Data Collection Procedures	68
----------------------------	----

Ethical Considerations	69
------------------------	----

Data Processing and Analysis	70
------------------------------	----

Chapter Summary	70
-----------------	----

## CHAPTER FOUR: RESULTS AND DISCUSSION

Introduction	71
--------------	----

Demographic Characteristics of Respondents	71
--	----

Demographic Characteristics of Interviewed Respondents	74
--	----

Answers to Research Questions	74
-------------------------------	----

Research Question 1	74
---------------------	----

Research Question 2	79
---------------------	----

Research Question 3	82
---------------------	----

Hypothesis One	88
----------------	----

Hypothesis Two	90
----------------	----

Hypothesis Three	92
------------------	----

Hypothesis Four	95
Discussion	97
Testing Practices Teachers Follow when Constructing Achievement Test Items	97
Testing Practices Teachers Follow when Administering Achievement Test Items	100
Testing Practices Teachers Follow when Scoring Essay-Type Test Items	101
Differences in Testing Practices Based on Gender	103
Differences in Testing Practices Based on Teaching Experience	104
Differences in Testing Practices Based on Professional Qualification	105
Differences in Testing Practices Based on Knowledge of Test Construction	106
Chapter Summary	107
<b>CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS</b>	
Introduction	108
Summary of Study	108
Major Findings	110
Conclusions	113
Recommendations	114
Suggestions for Further Research	115
<b>REFERENCES</b>	117
<b>APPENDICES</b>	134
<b>APPENDIX A: TEACHERS' QUESTIONNAIRE</b>	135
<b>APPENDIX B: INTERVIEW GUIDE FOR TEACHERS</b>	143

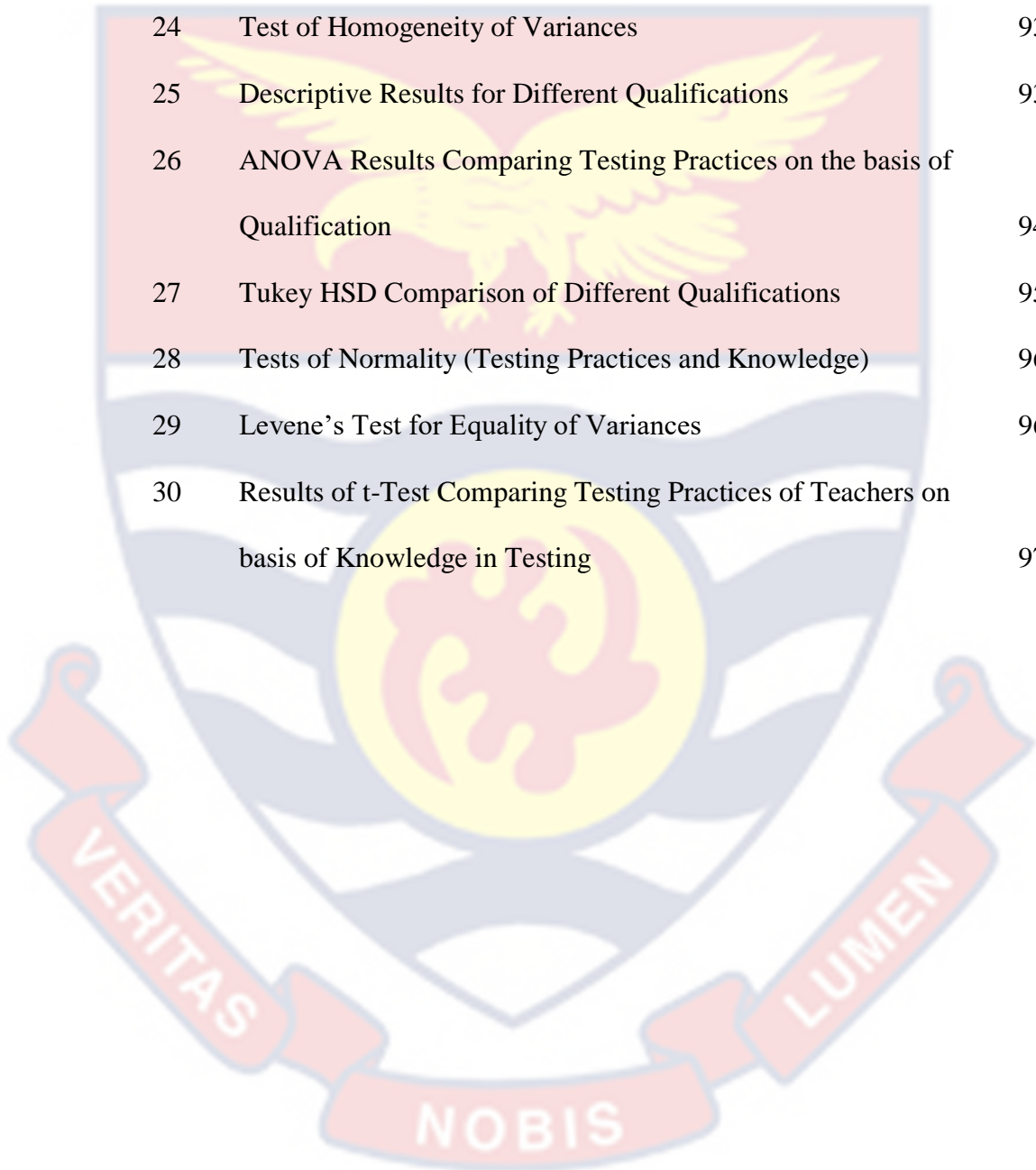
APPENDIX C: RELIABILITY OUTPUT	145
APPENDIX D: INTRODUCTORY LETTER	146
APPENDIX E: ETHICAL CLEARANCE	147



## LIST OF TABLES

Table	Page
1 Demographic Characteristics of Respondents	71
2 Classes Taught	72
3 Subjects Taught by Teachers	73
4 Ever Taken a Course in Testing (Educational Measurement)	73
5 Testing Practices in Constructing Achievement Test	75
6 Construction of Objective Type Tests	76
7 Considerations in Arrangement of Objective Test Items	77
8 Determining Time to Allot for Essay Test	78
9 Preparing Students in Advance for Test	80
10 Test Administration Practices	81
11 Method Used in Scoring Essay Tests	82
12 Scoring of Essay Tests	83
13 Test Scoring Practices	84
14 Method used in Interpreting Results of Test N=382	85
15 Method used in norm-referenced interpretation	86
16 Method used in criterion-referenced interpretation	87
17 Tests of Normality (Testing Practices and Gender)	89
18 Levene's Test for Equality of Variances	89
19 Results of t-Test Comparing Testing Practices of Male and Female Teachers	89
20 Tests of Normality (Testing Practices and Years of Teaching)	91
21 Levene's Test for Equality of Variances	91

22	Results of t-Test Comparing Testing Practices of Teachers on basis of Years of Experience	91
23	Tests of Normality (Testing Practices and Educational Qualification)	92
24	Test of Homogeneity of Variances	93
25	Descriptive Results for Different Qualifications	93
26	ANOVA Results Comparing Testing Practices on the basis of Qualification	94
27	Tukey HSD Comparison of Different Qualifications	95
28	Tests of Normality (Testing Practices and Knowledge)	96
29	Levene's Test for Equality of Variances	96
30	Results of t-Test Comparing Testing Practices of Teachers on basis of Knowledge in Testing	97



## CHAPTER ONE

### INTRODUCTION

The process of teaching and learning is incomplete without an assessment of students to identify the extent to which set targets are being reached (Inko-Tariah & Okon, 2019). In the classroom, tests are mostly used in assessment of students. The assessment process involves test construction, administering and scoring. To be able to effectively carry out the entire process, teachers need some specific skills and competences. It has thus been stressed that teachers' competence and skills can influence to a great extent the quality of tests that are made (Chan, 2009; Darling-Hammond, 2012). As a result, the focus of this research is on basic school teachers' classroom testing practices in Ghana's Greater Accra Region.

#### **Background to the Study**

Testing practices is defined by NEA (2014) as, the use of tools to examine someone's knowledge of something to determine what that person knows or has learned. At the school level, educators create tests to measure their students' understanding of specific content or the effective application of critical thinking skills. Such tests are used to evaluate student learning, skill level growth and academic achievements at the end of an instructional period, such as the end of a project, unit, course, semester, program or school year (Din, 2020). It is vital to evaluate students' development in schooling. Assessment is the collection of data to make educational judgments about students, offer feedback concerning their growth, strengths and shortcomings, determine effectiveness of teaching methods, and curriculum appropriateness (American Federation of Teachers [AFT], National Council on Measurement

in Education, [NCME], & National Education Association [NEA], 2014). As a result, educational testing comprises collecting data in order to make educational decisions. While assessing students' cognitive capacity, academic skills, and intellectual progress, some methodologies are utilised to find out how students perform on a specific learning and instructional objectives (Bichi, 2016).

Assessment is a significant domain of the educational system because it helps teachers make important judgments regarding pupils. Some of the decisions made about students such as in terms of management of instructions, placement of students in various educational programmes, assignment of students to some specific and suitable groups, guiding and counselling of students, selecting of students for some opportunities, and for credential and certification of students (Nitko, 2001). Scriven (1994) outlined two types of assessment which can be applied to obtain desired outcomes. Formative and summative assessments are the two types of assessment.

Formative assessment as indicated by Etsey (1992), is a systematic manner for gathering records on students during teaching and learning within the classroom, while summative assessment is utilised mostly at the completion of instructions to find out the success of students and provide proof of competence or programme effectiveness. Testing is one of the most used methods of assessment. A test, according to Tamakloe, Amedahe, and Atta (1996), is a device or technique that measures a specified part of a person's behaviour in a learnt activity or discipline. A test, according to Crooker and Algina (2008), is a conventional technique for gathering a sample



of behaviour from a certain area. As a result, testing is an important part of educational assessment.

Tests are used in schools to get information about pupils' learning (Quansah, Amoako & Ankomah, 2019). Tests are extremely important in the educational system since they give a platform for achieving any significant educational goals (Hamafyelto, Hamman-Tukur & Hamafyelto, 2015). As a result, developing high-quality tests is critical for evaluating students' performance.

From the forgoing, there are some principles and skills required to ensure that test items constructed are valid and reliable (Bichi, 2016). This does not imply that instructors must be specialists in educational measurement and assessment in order to create valid and reliable test items, but there are certain basic test development abilities and skills that teachers should have. According to Silker (2003), skills in test development enable a teacher to develop test items with exactness, fitting language, fairness, and appropriate grading measure. Such skills can assist teachers to construct items that can bring to the fore clear and succinct responses from students, ensure that tests are suitable for students of all ages, capacities and gender, and carry out testing in such a way that students complete on time and are not terrified of the test (Ali, 1999).

In the absence of the right test construction skills, teachers may end up obtaining false results from their assessment. This is because incorrect test items might impair students' understanding and ability to offer accurate responses to test questions, thus jeopardizing the conclusion formed about the level of knowledge and competence of students (Koksal, 2004; Leighton &

Gokiert, 2005). Aside this, there have been claims that poor test construction may be connected to examination malpractices in schools (Esomonu, 2002; Paulson, 2003).

When it comes to classroom or teacher-made exams, mastering test creation abilities and competencies is critical. In schools, teacher-created tests are widely utilized as a significant tool for evaluating pupils' development (Koloi-Keaikitse, 2017). These tests, which are created by the teacher, are also known as classroom achievement tests. Students may be forced to answer questions that lack fundamental psychometric qualities (i.e. validity, reliability, and usability) because they are developed and administered by teachers who may not have enough expertise of test construction techniques (Inko-Tariah & Okon, 2019). This emphasizes the need for teachers to be familiar with assessment techniques such as test design, administration, and scoring. However, only a small portion of teachers receive instruction in tests or measurement and evaluation instruction (Asamoah-Gyimah, 2002).

Scholars have suggested that test creation by instructors is not encouraging on a global scale (Hamafyelto, Hamman-Tukur & Hamafyelto, 2015; Kazuko, 2010). For example, Campbell and Evans (2000) looked at pre-service teachers with some level of training in educational measurement and identified that many of the assessment processes recommended in the curriculum were not implemented by student teachers. As a result, if tests are not properly constructed, teachers may wind up with erroneous information regarding student learning. In Africa, similar arguments have been made. For example, Ololube (2008) assessed the test development skills of Nigerian instructors and discovered that teachers who are not professionals had weak

test creation skills. Onyechere (2000) also discovered that most teachers created bad test items which resulted in poor outcomes. These teachers accepted that they had weak test construction skills and so resorted to the use of past questions in assessing students.

In Ghana, there has been some attention on the test construction practices of teachers. This is because several authors have shown that testing practices and skills of teachers are limited. For example, Quansah and Amoako (2018) discovered that teachers in senior high schools in the Cape Coast Metropolis had a negative attitude toward test creation, particularly when it comes to test planning, item development, item review, and item assembly. This had an impact on the quality of the tests used to evaluate students. This was supported in the view of Asamoah, Sundame, Adusei, Ocansey, and Ntim (2019) that high school teachers have poor abilities in the creation of end-of-term examinations. According to Anhwere (2009), both trained and untrained teachers in Ghana's classrooms, from elementary to tertiary level are involved in the development, administration and scoring of classroom tests, in spite of their level of know-how and training in measurement and evaluation. Teachers make test-related decisions that have far-reaching consequences for students. As a result, teachers must have sufficient knowledge and skills in the science and art of test creation, administration, and scoring (D'Agostino, 2007).

The demographic characteristics of teachers have been studied in connection to their competence in test construction. For example, according to Dubem cited in Inko-Tariah and Okon (2019), teachers' test construction competence is determined by their personality and training rather than their

gender, because good test construction skills are learned through training. Petters and Okon (2014) also revealed that knowledge of test development did not vary on the basis of gender.

In general, test construction competence is a vital tool for every teacher in ensuring that the objectives and targets of teaching and learning are achieved (Quansah & Amoako, 2018). Regardless, policymakers have largely ignored teacher training and preparation in the areas of how tests are constructed, administered and scored (Anhwere, 2009). In the basic schools in Ghana, it appears proficiency and competency in testing of students is lacking. This greatly affects the nature of tests constructed as well as the procedures in administering and scoring (Chan, 2009; Darling-Hammond, 2012).

Basic schools in the Greater Accra Region like all other basic schools have tests for their students throughout the academic year. Such tests help identify the areas where the students need to be assisted to perform better. This is due to the fact that tests are frequently used in formative assessment to evaluate and promote students' learning (Roediger & Karpicke, 2018; Roediger, Putnam, & Smith, 2011). The scores obtained from formative assessment should in principle, somewhat correlate with scores during summative assessment, even if this has not been evident in most Ghanaian classrooms (Anane, 2011).

According to Anane and Adu-Mensah (2019), it is believed by experts in measurement that educational reforms in curriculum creation (objective and standards-based), utilization of standardised tests, and high-stakes testing programmes and plans are required to guarantee that tests are developed and utilized appropriately. There is also a great need for reforms to involve

enhancing teachers' professional abilities in the scoring and grading of tests in the classroom, so as to guarantee that teachers adhere to standard testing practice that would not render students' test results useless (Anane & Adu-Mensah, 2019; Marzano, 2000). There are over one thousand six hundred and seventeen (1617) teachers within the Greater Accra Region who in one way test their students. Currently there isn't enough study on basic school teachers' classroom testing practices within the Greater Accra Region, therefore the current study sought to investigate the issue. On this basis, it is important to determine the testing practices of basic school teachers. It is in the light of this that the study is conducted.

#### **Statement of the Problem**

The relevance of testing in schools has gained recognition across different societies and countries as a significant part of education at all levels (Borich, 2011; Sarita, 2005). However, it is generally argued that teachers have low knowledge in testing or construction (Armah, 2018) which affects their ability to administer the appropriate test for their students. In several places around the world, there have been studies on the testing competencies of teachers. For example, Inko-Tariah and Okon (2016) discovered that most test items created by lecturers lacked fundamental psychometric features, based on a sample of 440 lectures in Nigeria. Similarly, Muzenda (2017) found that teachers in Zimbabwe were unaware of the normal procedures for creating, marking, scoring, and grading assessments. Pidu, Darsikin, Eliaumra, and Ika (2018) found that senior high school teachers in Indonesia had poor test construction skills. In Botswana, however, Koloi-Keaikitse (2017) found that instructors felt more knowledgeable in test construction than in other

techniques like employing classroom assessment. This assertion indicates that, ones knowledge plays a key role in their ability to test their students and also understand the appropriate test to use during the process.

In Ghana, there have been some studies on the testing competencies of teachers in schools. For instance, Quansah, Amoako and Ankomah (2019) and Asamoah et al. (2019) reported that teachers in high school have limited skills when it comes to construction of end-of-term examination. When issues with the test's content representativeness and relevance, as well as its dependability and fairness, were uncovered, this became evident. Yeboah (2018) however reported that after training, senior high school teachers have acquired competencies and skills in constructing tests. Anane and Adu-Mensah (2019) also examined how basic school teachers scored and graded tests and found that teachers considered academic and non-academic issues in the scoring and grading of students.

Even though there have been studies in Ghana, it appears most of the studies cited in this work have focused on senior high school teachers except that of Anane and Adu-Mensah (2019). Also, the studies have been conducted mainly in the Cape Coast Metropolis (Quansah et al., 2019; Asamoah, et al., 2019) and Agona West Municipality (Yeboah, 2018) to neglect of other metropolis such as the current study area.

Besides, several variables have been noted to affect teachers' test construction skills such as gender, years of teaching experience, professional training, and knowledge in tests construction (Anane & Adu-Mensah, 2019; Armah, 2018; Inko-Tariah & Okon, 2019; Quansah & Amoako, 2018). While studies (Inko-Tariah & Okon, 2019; Adodo, 2014) posit that teachers' tests

construction skills are not influenced by these demographic variables of teachers such as gender, years of teaching experience, professional training, and knowledge in tests construction, others (Zhang & Burry-Stock, 2003; Muzenda, 2017) argued that teachers' test construction skills depend on these variables. From the foregoing, it is evident that there are inconsistent research findings on the demographic variables that determine tests construction skills.

Currently, there is little to no data on the knowledge and skills of basic school teachers on their competence to assess their students and it is unclear whether or not basic school teachers in the Greater Accra Region have the skills and competencies to assess their students, and whether or not their demographic variables affect their test construction practices. Since junior high school teachers are involved in the construction, administration and scoring of tests, this study aimed at exploring basic school teachers' testing practices in the Greater Accra Region.

### **Purpose of the Study**

The study sought to examine basic school teachers' testing practices within the Greater Accra Region of Ghana. Specifically, the study sought to:

1. investigate the testing practices basic school teachers follow when constructing achievement test items;
2. investigate the testing practices basic school teachers follow when administering achievement test items;
3. investigate the testing practices basic school teachers follow when scoring essay-type test items; and

- ascertain whether gender, teaching experience, professional qualification and knowledge of test construction affect the testing practices of basic school teachers.

### Research Questions

The research questions which guided the study are outlined below:

- What testing practices do basic school teachers follow when constructing achievement test items?
- What testing practices do basic school teachers follow when administering achievement test items?
- What testing practices do basic school teachers follow when scoring essay-type test items?

### Hypotheses

The hypotheses tested in the study are shown below:

H<sub>0</sub>1: There is no statistically significant difference in basic school teachers' assessment practices on the basis of their gender.

H<sub>1</sub>2: There is statistically significant difference in basic school teachers' assessment practices on the basis of their gender.

H<sub>0</sub>2: There is no statistically significant difference in basic school teachers' assessment practices on the basis of their years of teaching experience.

H<sub>1</sub>2: There is statistically significant difference in basic school teachers' assessment practices on the basis of their years of teaching experience.

H<sub>0</sub>3: There is no statistically significant difference in basic school teachers' assessment practices on the basis of their professional qualification.



H<sub>3</sub>: There is statistically significant difference in basic school teachers' assessment practices on the basis of their professional qualification.

H<sub>0</sub>: There is no statistically significant difference in basic school teachers' assessment practices on the basis of their knowledge on test construction.

H<sub>1</sub>: There is statistically significant difference in basic school teachers' assessment practices on their knowledge on test construction.

### **Significance of the Study**

The current study would highlight the practices of junior high school teachers in the development, administering and scoring of tests. The results of the study would be important for stakeholders in education like the Ghana Education Service (GES) and the Ministry of Education (MoE), heads of basic schools and teachers to be in know of their tests construction skills. In the first place, the Ghana Education Service (GES) and the Ministry of Education (MoE) would be enlightened as to the level of proficiency and practices of basic school teachers in terms of test construction, administering and scoring. This can help them identify the gaps in the knowledge of teachers so that measures such as training workshops can be organised for the teachers to bridge the gaps.

Secondly, the results would help basic school headteachers know the gaps in the knowledge of teachers about testing. With this knowledge, they can plan their own measures of training to help equip their teachers with the necessary skills required in test construction, administering and scoring (Quansah et al., 2019). For basic school teachers, the findings of the study

would enlighten them as to their level of competence and expertise regarding specific aspects of the testing procedure. Based on this, basic school teachers can take steps of enhancing their own competencies in testing.

Again, the results of the study would bring to bear whether or not teachers' test construction skills are dependent on their demographic variables like gender, professional qualification, years of teaching experiences and knowledge of test construction. This would expose them to developing the habit of ensuring such demographic features for effective test construction (Inko-Tariah & Okon, 2019).

Besides, the results of the study would augment existing literature regarding the test construction of basic school teachers, and add to the discussions of the current study variables. Lastly, the results of this study would become a reference material for future researchers and other stakeholders who may find it useful.

### **Delimitations**

The study concentrated on teachers' test construction at the basic school level in the Greater Accra Region of Ghana. Specifically, the scope of the study covered test construction, administering, and scoring practices of teachers. The study also covered teachers' demographics such as gender, years of teaching experience, professional qualification and knowledge of test construction. Regarding the participants of the study, the study covered only basic school teachers teaching at the junior high level in the Greater Accra Region of Ghana. The study was limited to both multiple choice and essay tests administered by teachers in the schools.

### **Limitations**

The major challenge that was encountered in the study was the respondents answering the questionnaires in a timely manner. This is because some of the teachers saw the study as a form of evaluation of their teaching and as such were hesitant in responding to the items on the questionnaire. Also, the use of the questionnaire in collecting data means the data was self-reported. This did not adequately provide a vivid picture of the actual testing competencies and practices of teachers.

### **Definition of Terms**

The main terms used in the study are operationally defined in the context of how they are used in the study:

**Test construction:** This is used in the study to refer to the set of activities involved in developing a classroom test.

**Test administering:** This term refers to the techniques and processes used to conduct a test in order to minimize measurement error and enhance the possibility of a fair, valid, and accurate evaluation.

**Test scoring:** This is used to describe the process involved in arriving at evidence of the performance of students on a test.

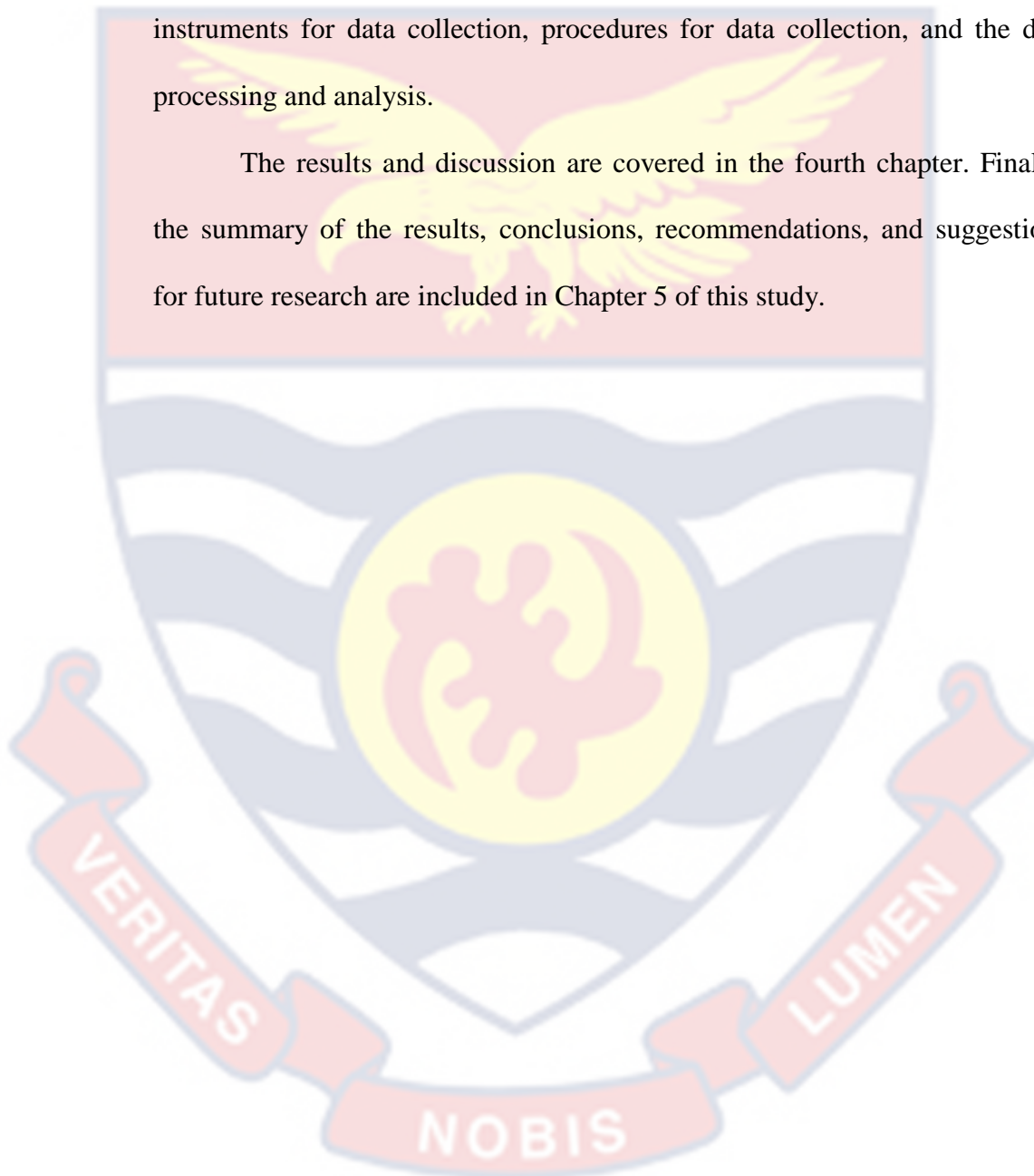
**Testing practices:** It involves studying a topic and then trying to recall the newly studied material.

### **Organisation of the Study**

The study was broken down into five chapters. The background to the study, statement of the problem, purpose of the study, hypotheses, significance, the study's delimitation and limitations, definition of terminologies, and organisation of the study are all covered in the first chapter.

The second chapter is a review of literature in the study. There was a theoretical framework, conceptual review and a review of related empirical research. The methodology of the study is discussed in the third chapter. This deals with the research design, study area, population, sampling procedures, instruments for data collection, procedures for data collection, and the data processing and analysis.

The results and discussion are covered in the fourth chapter. Finally, the summary of the results, conclusions, recommendations, and suggestions for future research are included in Chapter 5 of this study.



## CHAPTER TWO

### LITERATURE REVIEW

The main aim of this study is to investigate basic school teachers' testing practices in Ghana's Greater Accra Region. The study's literature review is covered in this chapter. The review is divided into sub-sections: theoretical review, conceptual framework, conceptual review and empirical review.

#### **Theoretical Review**

In this section, the theories that form the foundation for the study are reviewed. The theories include the Classical Test Theory and the Item Response Theory and they are discussed in detail in this section.

#### **Classical Test Theory (CTT)**

Classical test theory is a theory which explains how errors in measurement impact on the outcomes of the measurement (Marcoulides, 1999). According to Schumacker (2010), classical test theory (CTT) deviates from the early 20<sup>th</sup> century methodologies for assessing differences among people. CTT originated from three accomplishments: 1. acknowledgement of measurement errors; 2. recognition of the error as an unplanned variable; and 3. recognition of relationship and its indexing.

Charles Spearman established means of correcting a correlation coefficient for attenuation because of faulty measurement in 1904 (Allen & Yen, 1979). He also discovered how to compute the dependability index necessary for the correction, according to Allen and Yen (1979). Classical Test Theory is said to have begun with his finding. George Udny Yule, Truman Lee Kelley, Louis Guttman, and those who are part in developing

Kuder-Richardson Formulas are some of the other scholars who contributed to the Classical Test Theory (Traub & Fisher, 1997).

Traditional test theory has generally been utilised in finding out reliability of some instruments and other attributes for decades. Classical test theory is a test score theory which focuses on three main notions, according to Hambleton and Jones (1993). These are the test score (observed score), true score, and error score. In the theory, true scores and error scores are deemed to be not related and error score which is zero also not related to the rest of the scores.

In classic test theory, a student or testee has an unobservable true score that would be earned if no measurement errors were evident (Magno, 2009). Regardless, if the instrument utilised is not adequate and comprehensive, the score assigned to each individual may not accurately reflect their true talents. Measurement error is to blame for the variance between the observed and the true score (Bichi, 2016). Error is commonly thought of as a random element in a normal distribution.

The consequence for examinees of CTT is that tests are flawed, inaccurate instruments are used and that the score achieved by a person is referred to as the individual's true score (Bichi, 2016). This means that an individual's genuine score will not change even if they take the same exam many times. CTT's observed score is usually the real score modified by some amount of mistake, which can have a positive or negative influence on the observed score. The standard deviation of each examinee's distribution of random errors, according to Kaplan and Saccuzo (1997), shows the degree of measurement error.

The distribution of random errors is commonly assumed to be uniform across all individuals taking the test (Bichi, 2016). The standard deviation of mistakes is the fundamental measure of error in CCT. Practically, the standard error of measurement is calculated by combining the test's reliability with the standard deviation of the observed score. The lower the standard error of measurement, the more definite the property being assessed is, and the individual score is likewise closer to the genuine value. On the other side, the higher the standard error of measurement, the less correctly a property is measured.

For CTT models, item statistics are viewed to be sample-dependent. The indication is that the individuals who took part in the item standardization are influential in the test item statistics (Zanon, Hutz, Yoo & Hambleton, 2016). However, it would be easier to create tests if item data were not linked to the examinee sample. Similarly, examinee test results are influenced by the selection of test items. Thus, testees must either take the identical test items or items which are statistically equivalent in order to be compared. However, Zanon et al. pointed out that there are situations when better testing might occur if test items were chosen based on the examinees' abilities. Examinees who scored well on a construct, for example, may be given more difficult objects. Individualizing test item selection, on the other hand, would reduce the ability to compare examinees.

In summary, Classical Test Theory (Spearman, 1904; Novick, 1966) was (and still is) used to predict a person's latent feature based on a total score on an instrument. It has been questioned, however, because the interpretation of responder characteristics is dependent on the test utilized. If an easier exam

is given, examinees would seem to be smart; nevertheless, if a more difficult test is given, respondents would appear less clever.

Regardless of this, the theory was considered applicable in the current study since it gives understanding as to how the procedures and skills used in constructing and conducting tests can affect the results of students and the interpretation of the performance of students. On this basis, the CTT can help explore the practices of teachers in test construction.

### **Item Response Theory (IRT)**

Item Response Theory (IRT) was propounded in the 1950s and 1960s by Frederic Lord and other psychometric experts with the goal of providing a system for assessing replies without relying on the same test questions (Lord, 1952; Lord & Novick, 1968). (Hambleton & Jodoin, 2003). As a result, IRT arose from traditional measurement theory in order to address several of its flaws (Hambleton, 1994).

IRT is conceptualized as a statistical theory made up of a number of mathematical models with the features listed below. These are: a) predicting a person's score on the basis of his or her talents or latent qualities; and b) using a function termed the "item characteristic curve", establishing a link between a person's item performance and the set of attributes that underpin item performance (Hambleton, Swaminathan & Rogers, 1991). These attributes are attainable in the event that the IRT model really matches the data that is available. The implication of this is that, when the same test items are utilised with different participants, the test items would retain their properties such as difficulty and discrimination, and people's scores on a specific construct that



indicates ability or latent features will not be influenced by the test items they were given (Zanon, Hutz, Yoo & Hambleton, 2016).

Generally, IRT is viewed in psychometrics, as a paradigm for designing, analyzing, and scoring tests, and other instruments that assess skills, attitudes, and some other properties of individuals (Hambleton, Swaminathan & Rogers, 1991). It is a theory based on the relationship between test takers' overall performance on a specific characteristic or property that the item was developed to assess and their performance on the test item. As a result, item response theory sees each item's difficulty as data that must be included into item scaling.

Item response theory gets its name from the fact that it focuses on the item rather than the test level like traditional test theory. As a result, IRT simulates each examinee's answer to each test item for a particular skill. The theory is considered applicable in the current study since it enhances understanding of how focusing on the items of a test and developing it well can help obtain valid and reliable results. This implies that teachers constructing tests should have skills and proficiency into how to construct specific items on a test.

The item characteristic curve is the basic building block of item response theory; all the other constructs of the theory depend upon this curve. Therefore, considerable attention will be devoted to this curve and its role within the theory. There are two technical properties of an item characteristic curve that are used to describe it. The first is the difficulty of the item. Under item response theory, the difficulty of an item describes where the item functions along the ability scale (Himelfarb, 2019). For example, an easy item

functions among the low-ability examinees and a hard item function among the high-ability examinees; thus, difficulty is a location index. The second technical property is discrimination, which describes how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location. This property essentially reflects the steepness of the item characteristic curve in its middle section. The steeper the curve, the better the item can discriminate (Kondratek, 2023). The flatter the curve, the less the item is able to discriminate since the probability of correct response at low ability levels is nearly the same as it is at high ability levels. Using these two descriptors, one can describe the general form of the item characteristic curve. These descriptors are also used to discuss the technical properties of an item. It should be noted that these two properties say nothing about whether the item really measures some facet of the underlying ability or not; that is a question of validity. These two properties simply describe the form of the item characteristic curve (Akbari, 2020).

### **Conceptual Review**

The major concepts in the study are discussed in detail in this section.

### **Concept of Assessment**

Instructors utilize assessment to ascertain the level at which students have effectively learned through a systematic method of evaluation. According to Khan (2019), assessment is a means of collecting data for a certain purpose or aim. In other terms, assessment is the process of evaluating learners' performance and knowledge. According to Rhalmi (2020), assessment is the act of collecting or obtaining data about pupils from various sources so as to get an idea of their level of knowledge and how they can use the knowledge

they possess. The word “assess” is frequently used when discussing issues like skills, abilities, performance, aptitude, and competence.

Some of the earlier authors have given good description for assessment. For instance, Le Grange and Reddy (1998, p.3) were of the view that “assessment occurs when judgments are made about a learner’s performance, and entails gathering and organizing information about learners in order to make decisions and judgments about their learning.” From the view of Le Grange and Reddy, it can be seen that assessment is a crucial aspect of the teaching and learning process since it supports student learning and enhances education.

Teachers in every setting are responsible for promoting student learning and motivation, documenting student performance, and reporting outcomes for accountability reasons. This means that classroom assessment is major duty for teachers. The main aim of Classroom Assessment is “to empower both teachers and their students to improve the quality of learning in the classroom” through an approach that is “learner-centered, teacher-directed, mutually beneficial, formative, context-specific, and firmly rooted in good practice” (Angelo & Cross, 1993, p. 4).

Every form of assessment, according to Shulman (2004), begins with defining precisely what is to be assessed. Teachers do this so that pupils can achieve their goals. As a result, specific skills are required for the assessment procedure. In essence, the assessment in education course material is intended to impart competences and abilities to student teachers. Teachers must grasp fundamental concepts about how learners learn, as well as the topic areas they

teach so as to effectively assess learners and make appropriate judgments on the basis of the results (William, 2010).

The three types of assessment utilized in the classroom. These are assessment for learning, assessment of learning, and assessment as learning:

#### *Assessment for Learning (Formative Assessment)*

According to the “assessment for learning” principle, assessment and instruction are to be unified as a whole unit. The strength of this assessment is not due to the employment of a specific evaluation technique but from appreciating the amount of learning that happens during the normal and routine course of a school day – and the amount of information teachers may gather about student learning from this content (McNamee & Chen, 2005). As a consequence, “assessment for learning” is a continual process which gives teachers the opportunity to assess learners daily and modify their teaching methods to meet what they find from the assessment (Burns, 2005). Such assessment also gives learners the opportunity to receive with timeliness, appropriate and concrete feedback which can help improve their academic work. After a lesson has been delivered, it is necessary to evaluate and assess whether all the learners were able to grasp the material taught while at the same time providing a sense of challenge for the above average learners. It is also essential at the end of a lesson to assess the knowledge students have obtained and what they do not know yet; the means to enhance lessons effectively and alternative means of teaching what was taught. Our attempts to improve our teaching skills are centered on this ongoing assessment of instructional practice.

#### *Assessment of Learning (Summative Assessment)*

Another type of assessment which is “assessment of learning” is a descriptive picture which tells the instructor, learners and parents the extent to which learners have accomplished specific learning tasks and objectives (Burns, 2005). It offers data on the academic accomplishments of learners. In spite of this, the data obtained from this assessment has no bearing or impact on the immediate teaching and learning process. According to Clarke (2006), summative assessment emphasizes a summary of students’ level of development across time in order to communicate, inspire, and evaluate their standards and progress. Clarke argued further that this type of assessment prioritizes the instructor engaging the student and interacting with him or her concerning his or her performance. Summative assessment therefore uses accountability to establish the level of achievement of learners on precise and definite activities at the end of teaching and learning (Black & William, 2001; Clarke, 2006; Harlen, 2006).

#### *Assessment as Learning*

As students learn, their metacognitive talents are strengthened and reinforced through assessment. This kind of assessment is essential in assisting learners to develop the habit of lifelong learning. This type of assessment therefore involves peer and self-assessment where individual learners gain insight into data about themselves concerning their previous level of knowledge and apply it to novel tasks in the future (Burns, 2005). Learners feel a sense of possession and boost their own confidence when they utilize data from peer, teacher and self-assessment in making modifications and enhancements in their own level of knowledge and understanding.

## Relevance of Assessment

Assessment in the classroom is relevant in so many respects. First and foremost, assessment provides teachers with the information and abilities needed to provide effective and relevant feedback on the work of students, according to Hattie and Timperley (2007). The evaluation process aids in validating how students' grades are calculated, as well as identifying and praising certain characteristics in students' work and guiding students to improve their work (Khan, 2019). Students are encouraged to improve their performance when they get comments. According to studies, when students perceive development and achievement instead of the failure and loss that comes with being in comparison with highly successful peers, they are more likely to be motivated and confident learners (Davies, 2004; Stiggins, Arter, Chappuis & Chappuis, 2004). In the past, evaluation systems were designed to assign rewards and punishment. While this approach can succeed in certain cases, it often leads to pupils seeing themselves as failures. If this continues, people will lose confidence and quit trying. Students, on the other hand, might grow to regard themselves as competent learners if they are involved in the evaluation process and receive appropriate feedback (Sparks, 1999).

Secondly, assessment helps teachers make classroom decisions. They obtain information to decide what has to be done for improvement (Khan, 2019). Assessment aids in providing information and guidance in teaching and learning in this way. This means that a good classroom assessment proposal obtains data on learners so that instructors may utilize to help them make better instructional decisions (Burns, 2005). It also gives teachers the needed data to inform them about the capabilities of students. In order to plan

successful education, teachers must also grasp what pupils misunderstand and from where the misunderstanding emerge. A strong classroom assessment strategy also gives pupils a blueprint to follow. Students should have the chance to be assessed anytime so that they may use it to direct and inform their academic work. Assessment is also used to meet the various requirements of kids.

Students can also use assessment to help them define learning targets. Learners need the chance and prospects to reflect on their level of academic work and what they need to do to accomplish their academic targets on a regular basis (Burns, 2005). When learners participate in an active manner in the assessment of their own academic work and formulating goals to attain them, they attain considerable progress in guiding their learning and knowing themselves as learners (Alberta Assessment Consortium, 2005).

Finally, assessment helps grade and give the overall performance of students. Several stakeholders like parents, employers, educational and governmental institutions can use grades to get an overview of a student's learning (Alberta Assessment Consortium, 2005). This can aid in the development of a roadmap for learners' future improvement.

### **Concept of Testing**

The concept of assessment covers a wide variety of tasks such as testing (Khan, 2019). In essence, a test is a type of assessment. As a result, all tests are assessments, but not all assessments are tests. Test is a systematic method for evaluating students' performance and collecting data on their abilities, knowledge, and attitudes. A test, according to Rhalmi (2020), is a tool, technique, or approach used to assess pupils' knowledge or competence

to execute a certain assignment. This is why, as previously said, testing is seen as a sort of assessment. The primary aim of classroom testing is to acquire accurate, trustworthy, and meaningful data on student accomplishment. Thus, tests would have to meet certain criteria like validity and reliability.

According to Gronlund (cited in Oduro-Okyireh, 2008), “the key to effective achievement testing is careful planning” (p. 15). Testing is a process that requires a lot of careful and systematic planning and preparation. The testing preparation process assists the test creator in determining how to quantify the desired learning results (Armah, 2018). By implication, for tests to satisfy adequate criteria, those who are responsible for the development of tests must have a deep and exact insight into the desired learning objectives and the accompanying subject matter to assist the construction of test items.

According to Crocker (2019), the common types of test include the following:

- A placement exam is a tool used by educators to assist students to be placed in a specific level or area of a programme or institution.
- Diagnostic test: The teacher uses this to determine where he or she should begin teaching the learners. This test informs the teacher of his or her pupils’ present standing. The test also aids teachers and students in determining their strengths and shortcomings.
- An aptitude test is used to forecast a student's likely success on a course.
- Achievement test: This is used to assess a student’s existing knowledge and apply it to upcoming assignments (for an external exam etc.). They



are developed to assess the knowledge and abilities of students following some form of teaching.

### **Stages in Testing Practices**

Testing practices are in series of stages. These are discussed in this section.

#### **Test Construction**

Test construction involves the processes and steps that teachers use in constructing test items for assessment. Test items have to go through a series of steps and procedures because they have to be constructed in ways that the outcomes of the test would meet the criteria of reliability and validity. As a result, assessment tools should be carefully chosen or built in order to produce valid and accurate results (Silker, 2003). Test building skills are the abilities required to create high-quality test objects using established test construction principles (Ali, as cited in Agu, Onyekuba & Anyichie, 2013). Ujah (2001) defines these competences as objectivity, communication, item validation skills, and the capacity to use suitable procedures for determining the extent to which test tools are reliable.

According to Silker (2003), test creation skills enable a teacher to create test items with correctness, accuracy, good language utilisation, fairness, and effective grading measure. According to Agu, Onyekuba, and Anyichie (2013), teachers do not need to be specialists in educational measurement and assessment to create tests that meet criteria of validity and reliability, but there exists some fundamental test-creation skills that all teachers should have. According to Koksal (2004), it is critical for the instructor to ask the following questions while creating a test: Is the assignment fully clear? Is it feasible to have more than one right response

(objective items)? Can test-takers get the right answer even if they don't have the ability that's intended to be tested? Is there enough time for test takers to complete the task(s)? All of these questions send the signal that test construction should be done with some skills and competencies as well as following some principles.

Agu et al. (2013) have emphasized that the quality of a teacher-administered exam is directly tied to its capacity to deliver the type of data required about students' performance. The results of such assessments enable teachers to assess the effectiveness of their instruction to some extent. Poorly constructed test items, on the other hand, can lead to erroneous learning assessments and misleading information about student performance and instructional efficacy (Education Up Close, 2005). Any feature of a test item that diverts the attention of the test taker away from the main focus diminishes the item's efficacy (Frey, 2007). Also, any item that is answered properly or wrongly due to extraneous variables related to the item gives both the examinee and the examiner erroneous feedback.

According to Koksal (2004), there are some factors that are consistently related to poorly developed tests which need to be dealt with to ensure that quality tests are developed. These include:

- No mention of the target test-takers, the skill or capacity that the test was designed to assess, the amount of time allotted for the test items, or the number of points awarded to test takers for each right response.
- Separate sections are not described properly.
- Test items with several viable answers due to a lack of conceptualization.

- On the papers, there is no mention of the time allotted for each activity. Only the entire amount of time allowed to complete all of the tasks was specified.
- The level of pupils is not taken into account while creating the exam.
- Instructions that are unclear.
- Students are required to complete tasks that are not in accordance with the duties assigned to them during classroom teaching.
- Items that aren't reflective of what the teacher wants to test aren't included.
- Making a collage of certain test objects.

### **Classroom-based Achievement Tests**

Achievement tests are useful for assessing key components of a topic, correctly reflecting the focus on key domains of education, and determining the appropriateness of a learner's understanding in a particular topic (Frey, 2007). Achievement test can be classroom-based (teacher-made assessments) or standardized (designed by test professionals for broader objectives) (Agu et al., 2013). The most instructionally relevant achievement assessments, on the other hand, are classroom-based and, if well-designed, offer teachers with precise, correct and helpful data on their students' retention of knowledge in certain school topics (Childs, 1989).

Achievement tests, in general, are a global method of determining a student's abilities, emotions, attitudes, views, and achievements in relation to what they have studied over a period of time (Nitko, 2004). Also, achievement tests are mostly "paper and pencil" based in the classroom. Objective or essay tests are the most common kinds of achievement examinations (Amedahe,

2008; Etsey and Afedo 2008; Nitko, 2001; Osterlind, 2006). Specifically, objective test demands that a test-taker provides a short answer (typically less than a line), whereas essay tests need extended responses (Armah, 2018).

### *Objective tests*

There are two forms or kinds of objective tests utilised within educational institutions. They include the supply and selection types. The selection type mainly covers “multiple choice”, “true and false”, and “matching”, while the supply type includes “completion”, “fill-in-the-blanks”, and “short answers” which all require that test-takers provide their own solutions (Armah, 2018). The types of objective tests have their different set of benefits and drawbacks. Multiple choice and supply tests help to assess specific and individual learning targets.

Lower-order thinking skills like information retention and understanding are assessed using the supply/completion type. This sort of exam is used to measure relatively simple learning outcomes. Students' capacity to discern the validity of assertions of facts, concepts, and principles is assessed using the “true or false” type. The value of a “true or false” item is mostly in measuring knowledge of factual information. It's however tough to prepare true or false stuff (Salvia & Ysseldyke, 2001).

Also, “true or false” items are divided into four categories: basic true or false, which has only two options (true or false); complex “true or false”, which has three options (“true or false and opinion”); compound true or false, which has three options (“true or false and conditional completion response”); and multiple true or false (Salvia & Ysseldyke, 2001).

It has been pointed out that objective tests do not allow for deception or bragging and are better suited for the measurement of low-level tasks like knowledge and understanding. They are also cost effective in terms of scoring time and are subject to item and statistical analysis (Etsey, 2008; Amedahe, 2010). On the other hand, the objective test is challenging to develop since it takes a lot of time to organize and produce, and it also cannot be utilized to measure higher and complicated learning outcomes as well as the essay type test, particularly in the areas of synthesis, analysis, and assessment (Armah, 2018).

#### *Essay type tests*

An essay test is one in which the student is free to construct his or her own answer using his or her own arguments (Etsey, 2008). When opposed to objective tests, these tests usually include fewer items, but each one requires a lengthy response. Essay tests can be restricted response or extended response (Armah, 2018).

The restricted answer type put some form of limitation on the test-taker in terms of the scope and dimensions, whereas the extended response does not put any limitations on respondents in any way (Amedahe, 2010). However, according to Armah (2018), extended-response essay questions set no restrictions on the student in terms of the point(s) he or she must cover or the style of organization he or she will utilize. Extended-response essay questions, according to Amedahe (2010, p.25), allow students to exhibit their abilities to use factual information, assess factual knowledge, arrange thoughts, and convey ideas in a logical and cohesive written format.

## Principles of Constructing Classroom-based Tests

One thing to keep in mind while creating classroom-based exams is that effective classroom-based assessment entails following conventional test-building methods (Armah, 2018). Every classroom instructor is required to have and use the necessary abilities for creating high-quality items for class evaluations. Etsey (2004) highlighted some steps of constructing classroom-based tests. They are indicated below:

1. Determine the test's purpose.
2. Decide on the item format you'll employ.
3. Decide what will be tested.
4. Make a list of the individual items.
5. Go over the items again.
6. Make a scoring sheet.
7. Create instructions, and 8. Assess the exam.

Gay (2011) gave some basic guidelines which can be useful for developing test items. They are presented below:

1. Each time one teaches a course, there is the need for new test items to be created.
2. In the course of teaching a specific subject, one is likely to do it in a somewhat different way, emphasizing new ideas and concepts. Using outdated tests almost always indicates that you are not testing on what you have really taught.
3. Give yourself plenty of time to complete the test.

4. Avoid writing the test the night before it is supposed to be taken. Allow one to two weeks to fine-tune your exam questions and structure.
5. Throughout the semester, create a bank of questions.
6. You don't have to write the questions right before the exam to get a head start. As you receive test questions, jot them down.
7. Pay attention to how the exams are laid up. The design should be simple, straightforward, and with the ease of reading. In the event that there are a lot of questions, put them in a group.

### **Validation of Tests**

Two main psychometric properties of tests that are considered very significant in test construction are validity and reliability of tests. These are discussed in this section.

### **Validity**

Validity has been viewed as “the extent to which evidence and theory support the interpretations of test scores necessitated by the suggested applications” (Din, 2020). Also, validity is recognized by Nitko (2001), as the accuracy with which one interprets and applies the outcomes of a student's evaluation. The import of this is that for validity to be ensured, the results of students in tests must be supported with series of evidence. It is also obvious that test score interpretations and uses are linked to validity. Only when the values suggested by the assessment findings are suitable are the interpretations and applications of the results deemed to have met the criteria of validity. Another implication of the validity definition is that validity is a question of degree rather than a "all-or-nothing" situation. Instead of just valid or invalid,

the validity of how a test is interpreted must be thought of in the context of being strong or being weak (Yeboah, 2017).

### **Kinds of validity evidence**

Evidence for validity can be in three main forms. These are content-related, criterion-related and construct-related forms.

#### *Content-related evidence*

In situations where the evidence for interpreting test scores is based on the appropriateness of test material, the techniques used to establish and generate test content should be stated (Moss, 1992; Yeboah, 2017). It should also rationalise the test's purpose in terms of the population to be tested and the construct to be measured. William (2001, p. 4) contended that "content-related validity should be concerned not just with test questions, but also with the responses generated, and the link between them". This means that content-related data must be expanded to include behavior elicited that truly corresponds to the assessment task's aims. Also, it has been indicated that "the circumstances under which student performances are obtained can have major implications for the validity of the interpretations from an assessment" (Crooks, 2001, p. 270). However, low motivation, test anxiety, and unsuitable assessment circumstances can all jeopardize how valid a student's test results are.

#### *Criterion-related validity*

Criterion-related validity, according to Tamakloe, Amedahe, and Atta (2005), refers to the test's ability to predict a student's performance in certain conditions. As a result, criteria-related data may be used to forecast a person's standing on a criterion measure of interest using test scores (Armah, 2018).



This is done by comparing the student's test results to a standard criterion, which is a direct and independent predictor of the exact behaviour that the test is supposed to assess.

Criterion-related evidence might be contemporaneous or predictive, according to Kinyua and Okunya (2014). Concurrent validity connotes the extent to which a person's current state on a criterion may be projected from their previous achievement on a test (Nitko & Brookhart, 2007). On the other side, predictive validity is when the results of a specific test are used in predicting the results of a different test.

#### *Construct validity*

In the view of Deville (1991, p. 46), "construct validity-evidence is directly concerned with the theoretical relationship of a variable 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". The process of operationalizing a concept involves developing a collection of observable behaviours or traits that are assumed to be related to the hidden construct. Assuming linkages between the construct of interest and other relevant constructions or actions is required when syntactically specifying the construct (Benson, 1998; Gregory, 1992).

#### **Reliability**

Armah (2018) defines reliability as the consistency of the results acquired. This means that the scores for each individual should be consistent from one instrument administration to the next and from one question to the next. As a result, the extent of consistency between two tests of the same phenomenon is referred to as reliability (Hussain, Jamil, Siraji & Maroof,

2012). In the classroom, reliability refers to the likelihood that students would produce comparable results when given the same examinations at different times.

### **Methods of estimating reliability**

The approaches for estimating dependability differ in that they take into account various causes of inaccuracy (Liaquat, Asif, Siraji & Maroof, 2012). The test-retest technique, equivalent-forms estimate, split-half approach, and internal consistency method, according to Liaquat et al., are all typical methods for measuring reliability.

#### *Test-retest method*

This method entails giving the same exam to a group of students repeatedly, spanning a time period from minutes to years. In this method, the results of the two tests are correlated, yielding an indication of the test's dependability (Etsey, 2012). There are a variety of time periods that may be used here. Because the estimate of reliability varies along with period between the tests, the period between the tests should be taken into account while assessing reliability coefficients. This depicts the student's genuine performance.

#### *Equivalent-forms estimate*

Equivalent-forms dependability is calculated by delivering two copies of a test to the same set of participants on a single day (with equal content, means, and variances) and comparing the results (Armah, 2018). This method tests how well an individual can extrapolate an individual's score to what he is likely to obtain on a test with similar but different questions (Hussain et al., 2012). Any change in performance is termed a blunder once more. In this

situation, rather than assessing variations from one period to the next, variations due to the precise knowledge are assessed. The difficulty and structure of the items in equivalent versions of a test should be the same. In addition, administrative instructions for both tests are intended to be the same (Liaquat et al., 2012).

#### *Split-half method*

In this method, a single test is administered to the learners in a split-half format. The exam is then scored in two parts, with the two results for each student being connected to provide a reliability estimate (Liaquat et al., 2012). The split-half approach can accurately reflect a student's genuine performance since the student is examined twice using the identical assessment task. The split-half technique has the benefit of requiring just one variant of the exam to be administered.

#### *Measures of internal consistency*

The Cronbach alpha is a means for measuring internal consistency. It represents the correlation of tests split into two halves. Internal consistency method is utilised in situations where test items are marked pass-fail or when a correct response is worth more than one point (Salvia & Yesseldyke, 2001). It is based on the assumption that the extent to which the results of students are consistent over time provides the foundation for establishing the overall reliability coefficient (Nitko, 2001).

#### **Administration of Classroom Achievement Tests**

The physical and psychological environment within which tests are taken is referred to as test administration (Amedahe & Gyimah, 2003). The guiding idea in test administering is to give all test-takers an equal chance to

exhibit their mastery of the subject matter (Oduro-Okyireh, 2008). The importance of maintaining consistent test administration situations cannot be overstated. Particularly, it is important if the test is to produce consistent, trustworthy, and legitimate results free of chance mistakes.

In test administration, it is of necessity that the teacher engages in preparation of his pupils for the exam ahead of time (Etsey, 2004). In order to optimize their performance, Etsey noted that students should be informed about when (day and time) during which the test would be administered, the situations and circumstances of the test (number of items, where test would take place and whether test-takers would be given chance to consult their books), and the topic areas (giving out specific questions or target areas) that the test would cover. Students should also be told about the significance of the test results, the amount of weight to specific topics, the types of items (either objective or essay), and the scoring and grading procedure.

The order in which tests are performed is crucial (Armah, 2018). This is because assessments should not be given either before or right after a long vacation, vacations, or other significant occasions in which pupils are physically or psychologically involved. Students' focus will be harmed if tests are offered when they would ordinarily be doing other enjoyable activities, such as eating or engaging in sports (Amedahe & Gyimah, 2003; Etsey, 2004).

Additionally, appropriate work space, silence in the area, excellent lights and ventilation system, and a suitable temperature are all physical factors that must be considered to enable optimal performance by the test takers (Etsey, 2004). The most important guiding concept in giving any classroom examination, according to Rukundo and Magambo (2010), is that

all test takers should be provided with the opportunity to show their accomplishment of the desired learning goals. Factors that might tamper with the measurement's validity must also be kept under control. Despite the absence of reliable evidence on the influence of physical and environmental factors on test accomplishment, test takers should feel as comfortable as possible by reducing and minimizing distractions. This is because any environment factors which distract test takers can have a minor impact on student performance, but they can have a significant effect on younger test takers (Linn & Miller, 2005).

Aside from these, interruptions during testing, such as giving directions, should be limited and always linked to the exam. At regular intervals, the time that has been used so far during the test and the time remaining must be announced so that students may devote their attention to the test items (Armah, 2018). When feasible, examiners should write the time on the whiteboard in 15-minute intervals through to the completion of the test. However, towards the final moments of the test, time can be indicated within every five minutes.

Furthermore, test takers must be required to begin and end the test on time (Amedahe & Gyimah, 2003; Etsey, 2004). During testing, teachers should constantly attempt to reduce test anxiety in children. As a result, teachers should avoid issuing threats to students that the exam is essential and that they must put in a lot of efforts in order to complete the test in a timely manner work harder to finish on time, and that there can be grave repercussions for failure in the test (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996).

### Scoring of Essay Tests

Essay tests are seen to be less objective, therefore there's a higher chance that the grades pupils earn on them, and the judgments they make based on them, will be invalid or unreliable (Anhwere, 2009). As a result, it's critical to use extreme caution while creating, marking, and administering essay assessments. Writing thorough requirements, indicating prompts which would explain the learner's goal, establishing rubrics for marking, and providing some form of training for the markers while also encouraging group-marking are some of means of enhancing essay tests to contribute to the education process in a valid and trustworthy manner (Capper, 2007).

Further, in essay tests, the learner's skill is measured in great part by the individual reading the document and the grading procedure utilized (Anhwere, 2009). How well an essay test is evaluated has a significant impact on its efficacy. One of the most common and reasonable arguments levied against the use of grading in essay test is that it is unreliable (Charmey, 1984).

According to Shibboleth (2011), grading essay questions and examinations is one of the very time-demanding and frustrating components of classroom assessment for most instructors. Teachers are sometimes hesitant to devote the time necessary to check essay exams. It almost goes without saying that correct grading requires a large amount of time and efforts from the perspective of the instructor.

When it comes to assessing essay assessments, it's critical that the instructor compose a comprehensive ideal answer ahead of time. This is critical since it will be used to evaluate each student's response. The repercussions of not doing so might be devastating. In addition, rather of

evaluating each student's whole test individually, student papers should be anonymously examined, and all replies to a single question should be marked one at a time.

Due to the lack of uniformity in the replies of the students that took the exam, essay questions are the most difficult to verify. In addition, there are a number of distractors in students' responses that might influence subjective essay item evaluation (Runeson, 2006). Handwriting, style, grammar, neatness, and student knowledge are examples of these detractors.

An essay test can be graded in two ways: holistic and analytic (Kubszyn & Borich, 2013). In holistic scoring, each essay question receives a total score based on the teacher's overall opinion or assessment (Shibboleth, 2011). Outstanding, very satisfactory, fair, and bad are the four categories in which essay responses are evaluated. Each of these categories is then given a score value. The maximum score is given to an outstanding response, while the lowest score is given to a weak response.

The essay is graded analytically in terms of its components. Separate points are awarded for arrangement of ideas, language and spelling, and supporting arguments or proofs in this type of essay. Teachers must ensure neutrality in assessing students' replies because an essay test is difficult to check (Hopkins et al, 1990).

Airisian (1994) provided the following criteria for the instructor to consider in order to reduce subjectivity in scoring an essay test:

- Before posing an essay question, determine what criteria form an excellent response.
- In the exam item, explain these factors.

- Read all of the responses to a single essay question before moving on to the next.
- After the initial grading, reread the essay responses a second time.

### **Interpretation of Test Results**

In the process of testing, interpretation of test results is the last stage. In interpreting test results, the instructor seeks to make meaning of the test results so that the aim of the test can be achieved. Thus, after a “measure of achievement has been obtained, the results need to be put in a form that is easily interpretable” (Gronlund, 1988, p. 155). In the literature, two main methods of interpreting test results are identified. These are the criterion-referenced and the norm-referenced interpretation. Even though these two are different, Gronlund (2003, p. 27) posited that “strictly speaking, the terms norm-referenced and criterion-referenced refer only to the method of interpreting results”. This means that a single test result can be interpreted using both types.

### **Criterion-Referenced Interpretation (CRI)**

Criterion-referenced interpretation (CRI) is a method of interpreting results which is mainly based on the extent of mastery of some specific skills. CRI thus describes the performance of students with regard to their ability to carry out some required task (Glaser & Nitko, cited in Nitko, 1996). In CRI, the test score gives an indication of the knowledge or skill level in a specific area possessed by the student taking the test. The focus of this interpretation is not to compare students to other students but to test the student’s level of competence in the subject area being assessed. For instance, a teacher can say that a student was able to solve eight out of 10 Mathematics questions in



Simultaneous Equation. Therefore, in CRI, the results of students give a description of their own levels of accomplishment.

In the classroom there are some tests referred to as criterion-referenced tests (CRT) which are utilised to test the mastery of specific learning outcomes or objectives (Gronlund, 1988). Gronlund argued that there was the need to set some standard in using criterion-referenced tests so that it would be easy to distinguish between students who have mastered a specific task and the students who do not have such mastery. In doing this, the “percentage-correct score is the most widely used method of judging whether learning targets have been mastered, in reporting the results of classroom CRTs” (Gronlund, 1988, p. 119).

From the forgoing, it is clear that CRI allows for comparison of students’ performance to some specific content area rather than for comparison of students to some norm group. Thus, to make an appropriate criterion-referenced interpretation, the content or area being assessed should be unambiguously defined in order to ease the process of describing a student’s level of accomplishment. Usually, teachers use percentages and raw scores as means of describing mastery. Gronlund (1988) has argued that specifying the criteria for standard performance may be challenging since there are usually not many clear directions to go by. In most cases, teachers rely on their personal judgments which are rooted in their years of experience in teaching. In CRI, teachers usually arbitrarily set a standard which occasionally receives upward or downward adjustment depending on prevailing circumstances (Gronlund, 1988).

### **Norm-Referenced Interpretation (NRI)**

Norm-referenced interpretation is the method of interpretation which is based on comparing individual students to some reference group which is known as the norm group. Therefore, in NRI, the scores obtained by a student are described in relation to a reference group who also took the same test (Nitko, 1996). In the classroom, NRI involves comparing and describing the performance of student in relation to other students in the class. The average test performance and distribution of scores of the norm group set the standard for the comparison. For instance, a teacher can state that the performance of a student is superior to 50% of the group or an individual placed 10<sup>th</sup> in a class of 20 students.

It is clear that NRIs compare the scores of students to a set of scores previously obtained by other students. Therefore, it can be said that, “how much a student knows is determined by his or her relative ranking within the norm group” (Oduro-Okyireh, 2008, p. 45). On this basis, appropriate and meaningful interpretation of scores depends on ensuring that the group being assessed is similar to the norm group. In the view of Gronlund (1988), in doing comparison, the teacher can use the raw scores of students in the form of rankings, percentile ranks or stanines. The percentile ranks and stanines are both generated from the raw scores. Oduro-Okyireh opined that in the ranking of raw scores, the scores of the individual students are arranged in descending order, thus, from the high score to the low score in a form of a frequency distribution. In interpreting results using this technique, the total number of students needs to be quoted for it to make sense.

Percentile rank gives an indication of the relative place of a student in a class by calculating the proportion of individuals who scored at or beneath the score of the individual student (Gronlund, 1988; Nitko, 1996). On the other hand, the stanine scale is “a system of standard scores that divides the distribution of raw scores into nine parts” (Oduro-Okyireh, 2008, p. 45). This means that the least stanine score is 1, the maximum is 9, and stanine 5 is deemed as the median score. This is because stanine 5 is at the centre of the data of scores. Stanines follow the normal distribution system with mean score of 5 and a standard deviation of 2 (Gronlund, 1988; Nitko, 1996).

### **Empirical Review of Previous Literature**

This section reviews previous research studies which are connected to the current investigation. The review is done under separate subheadings based on the main objectives of the study.

### **Testing Practices Teachers Follow when Constructing Achievement Test Items**

Test construction is an important part of the work of teachers. In carrying out this important task, there are some principles that need to be followed and some skills that teachers need to have. Some researchers have conducted studies to determine the extent to which teachers construct tests by following the right principles or with the right skills. Some of these studies are discussed in this section.

Pidu, Darsikin, Eliaumra, and Istadewi (2018) wanted to know how well physics instructors are at developing competency tests in junior high schools. The study took place in the Sindue Region of Kabupaten Donggala. The research was purely descriptive in nature and sampled 17 physics

teachers. A test, a documentation study, and an interview schedule were used in gathering the data. Teachers' knowledge, planning tests, table function specifications, and the nature of how tests are developed were among the teacher competencies tested in the study. Descriptive means and inferential means were used in analyzing the data gathered. The findings showed that teachers' ability to develop tests was limited. Quality of tests developed was poor and the language utilised was mainly average at best. The results show that the physics instructors did not strictly follow the rules when creating the tests.

The test-development abilities of Senior High School (SHS) teachers in the Cape Coast Metropolis were researched by Quansah, Amoako, and Ankomah (2019). Their study was qualitative and they focused on analyzing test documents in specific subjects. These subjects were Integrated Science, Core Mathematics, and Social Studies. The subjects were randomly selected (lottery approach) from three SHS in the Cape Coast Metropolis. The findings indicated that the teachers' ability to prepare end-of-term exams was restricted. When issues with the test's content representativeness and relevance, as well as its dependability and fairness, were uncovered, this became evident. Also, a suggestion was made that head teachers organise some academic seminars for teachers in the quest to improve the test development abilities of teachers.

Agu, Onyekuba, and Anyichie (2013) have created and validated a "Test Construction Skill Inventory (TCSI)" to measure the ability of teachers in senior high schools to develop classroom-based test items. The 30-item measure designed by the researchers was subjected to factor analysis. It was revealed that 25 items met the criteria to be valid. With a coefficient of 0.73,

the TCSI was also shown to be trustworthy, and secondary school teachers judged virtually all of the 25 items to be critical abilities for effective classroom-based test construction. From the results, it was realized that there were some skills required for test construction which were mostly lacking for teachers in senior high schools in Anambra State in Nigeria.

Anhwere (2009) investigated whether tutors follow the fundamental guidelines prescribed for testing in teacher training institutions, including the development, administration and scoring of classroom or instructor-made tests. The study was descriptive and involved 230 male and 80 female tutors in colleges of education in Ghana. Data was gathered using questionnaire. Mean and standard deviation, frequency counts and percentages, and independent samples-t test were utilised in analysing data that were gathered. From the findings, tutors in colleges of education did not follow the fundamental testing principles when developing tests for their students and were of the view that assessment procedures and activities were an additional burden on their teaching activities.

Teachers' attitudes regarding test construction were measured using a standardised instrument designed and verified by Quansah and Amoako (2018). In the Cape Coast Metropolis, the designed tool was given to 432 teachers in senior high schools. Four areas were discovered using an exploratory factor analysis: planning, item development, item review, and assembly. The factor loadings of the items were then examined using a confirmatory factor analysis. The instrument's items stayed at 32 on a four-point Likert scale after careful examination. Further investigation found that SHS teachers had a generally unfavorable attitude toward test development.

This suggests that the teachers did not always adhere to the rules of test development or construction. The researchers proposed that the Ghana Education Service (GES) together with the heads of various SHS, provide adequate teacher monitoring while developing tests.

Koloi-Keaikitse (2017) looked at teacher perceptions of ability in classroom assessment methods, which is contrary to the bulk of the studies in this section. A total of 691 teachers from Botswana's government elementary, junior and senior secondary schools were polled. The findings revealed that instructors usually felt more knowledgeable in test development than in other practices like utilizing the assessment data in making appropriate decisions during teaching and learning. This conclusion was counter to the majority of the research, which claimed that teachers were less experienced and competent in test development.

Inko-Tariah and Okon (2019) examined lecturers' knowledge of test development similar to that of Koloi-Keaikitse (2017) discussed earlier. Descriptive survey was adopted and the study was made up of 200 participants from 440 teaching staff. Data were collected using a self-structured instrument. Mean scores were utilized to answer the research question, and independent t-test and ANOVA were utilised in testing the hypotheses at the 0.05 level of significance. It was shown that the lecturers had a good level of understanding and competence in test development.

Mellati and Khademi (2018) also looked at teachers' assessment literacy and how it affected their present assessment techniques as well as the outputs of their students. A semi-structured interview schedule, observation checklist and the "Writing Competence Rating Scale (WCRS)" were used in

obtaining data. A total of 85 respondents took part in the study sampled from four universities in Iran. Specifically, the results brought to light that the level of knowledge about assessment of the respondents significantly affected the outcomes of students and teachers' assessment awareness led to successful assessment development in the classroom.

In Ghana, Ollennu and Etsey (2015) examined the extent to which positioning of items in multiple-choice test affected the level of performance of students during the Basic Education Certificate Examination (BECE). The sample comprised of 810 Form 3 pupils from 12 different Junior Secondary Schools (JSS). The study employed a quasi-experimental design. Test was used in gathering data for the study. By utilising ANOVA, the results showed that when item order was changed for English Language, Mathematics, and Science subjects, the difference in performance was statistically significant. Thus, the researchers concluded from the study that the ordering or arrangement of items of an objective test is an important practice to follow.

According to Etsey (2004), item arrangement must go from easier to more challenging forms in test development, and each part must be ordered in order of increasing complexity. Testing specialists have advocated that during tests, things should go from simple to challenging mainly to ensure that test takers are confident particularly from the start of the test (Sax & Cromack, 2005).

Caldwell and Pate (2013) also looked at the impact of three different test item formats on item statistics and student performance. Three particular item-writing rules were used to write fifteen pairs of directly comparable test items that followed (standard scale) or deviated from (nonstandard scale). The

differences in item difficulty and discrimination between the two scales as a whole and for each guideline separately were assessed. The two measures were also used to compare student performance. According to the study, test items which were not standard proved challenging for students than test items which were standard. Also, test items which were not standard did not clearly distinguish between individuals who performed at high level and those who performed at low level. Caldwell and Pate noted that item discrimination practices were not followed much and so led to poor student performance. From the results, it is evident that test development should be done in line with the appropriate guidelines and principles.

Furthermore, DiBattista and Kurzawa (2011) looked at students' responses to 1198 multiple-choice items on sixteen classroom assessments in a variety of subjects. Item discrimination coefficients were poor in more than 30% of the items. In addition, 45% of the distractors were incorrect. Overall, it was recognized that the quality of many multiple-choice examinations might need to be significantly improved, owing to poor item analysis and discriminating index analysis.

In test construction, Quaigrain and Arhin (2017) investigated the link between difficulty index and discrimination index (DI) along with distractor efficiency in the development of tests. The research involved 247 first-year students at Cape Coast Polytechnic pursuing a Diploma in Education. In the Educational Measurement course, a fifty-question multiple-choice test was given at the close of the semester. To increase the quality of the evaluation, Quaigrain and Arhin discovered that items with moderate difficulty and good discriminating power with functional distractors need to be included in future



examinations. This is because most test timings were not regarded to be the best without DI.

Zakariya (2020) investigated the practices of teachers in the development of tests by focusing on junior high schools within the Sissala East Municipality. This research was descriptive and 248 teachers were sampled using multistage sampling procedure. Specifically, purposive, stratified and simple random sampling procedures were used. After obtaining data through questionnaire, the findings revealed that the teachers moderately followed most test development guidelines. Also, majority of the teachers followed test administration standards averagely. The findings further revealed that the majority of the teachers lacked strong scoring abilities, which had a negative impact on the test scores of their students. Also, it was clear that the majority of tests were done without adherence to appropriate guidelines. In the usage of achievement tests, it was shown that large class size was a major challenge. According to the findings, it can be seen that generally, teachers in the Sissala East Municipality lacked the abilities necessary for test development, administration and scoring.

From the literature reviewed in this section, it is apparent that teachers in different places rarely adhered to the right practices or had the skills and proficiency in test construction. In a few of the studies, however, teachers were found to be well-equipped in test construction skills and practices.

### **Testing Practices Teachers Follow when Administering Achievement Test Items**

After tests are constructed, the next activity is the administering of the tests. In administering tests, teachers are expected to follow some principles

and practices. The degree to which teachers follow the right practices has been the subject of several studies. Some of these studies are discussed in this section. Armah (2018) investigated how well lecturers followed standard practices in developing and administering tests at the University of Cape Coast. The study employed a descriptive survey approach. For the study, 131 lecturers and 40 staff members from the Examinations Unit were selected using a census approach. By using questionnaire, the findings revealed that lecturers had a general knowledge of test development practices. Lecturers stated that they specified the goal of each test before it was developed, selected the format of the test items, and developed a grading system just at the time when the test was developed. Also, it was discovered that lecturers were familiar with the practices in administering tests. During the exam administration, students were informed of the test's norms and restrictions. The survey also revealed that lectures struggled to keep big groups under control during quizzes and exams. The study concluded that lecturers should get regular training on the development and administration of tests. Furthermore, it was suggested university administration should set up funding for workshops, seminars, and other educational opportunities for lecturers on the development and administration of test items.

Hudaya (2017) looked at the assessment of teachers' literacy in terms of their level of preparation in evaluating the performance of students, their practices in implementing language assessment principles in the classroom, and the utility of the questionnaire they created. In all, 43 teachers were given a 31-item questionnaire to give their responses. The data gathered were analysed descriptively. According to the findings, 79 % of teachers felt

prepared to assess students' performance. Also, it was found that in the administration of tests, teachers followed and implemented the appropriate principles. Generally, the conclusion that was drawn was that teachers could improve their instructions and understand how to utilize tests to assist their students increase their performance in the target language if they employ the principles of language assessment to develop tests.

Rukundo and Magambo (2010) also showed that it is essential to encourage fairness during tests through seating arrangements and supervision which would make students do independent work during test administration. Tyler (1991) also revealed that in test administration, teachers are expected to plan with detail the entire process for administering tests, use proper seating arrangements to avoid malpractices, and make intermittent announcements about the time for the tests.

Guskey's (2006) research on the reasons behind the grading of teachers revealed that teachers primarily utilised their personal experiences when they were students to determine how to score and grade tests. As a result, Guskey urged that teachers make every effort to ensure that grading processes are as transparent, unambiguous, and unbiased as possible. Guskey stated that teachers should avoid utilizing personal ideas and biases as a source of influence when scoring and awarding grades, and that their grading procedures and processes should be fair and impartial above all.

From the forgoing, it is clear that in most studies, teachers had the right skills in test administration. It remains to be seen as to the test administering skills of the teachers in the current study.

### **Testing Practices Teachers Follow when Scoring Essay-Type Test Items**

Essay-type tests require adherence to some principles in its scoring. Teachers' ability to follow these principles has been of interest to different researchers. Some of the studies in this regard are discussed in this section. In Ghana, Oduro-Okyireh (2008) wanted to know if teachers in senior high schools in the Ashanti Region followed the main principles when it came to testing. It also sought to determine if pre-service testing training helped people improve their testing skills. Using cluster and simple random selection procedures, 265 teachers teaching varied subjects in 26 schools were recruited for the study. The results showed that teachers, on the whole, followed basic test formulation, administration, and scoring requirements. Seven of 10 test development ideas, 12 of 18 test administration principles, and six of nine test scoring principles were applied by the teachers. Furthermore, teachers stated that they utilised both norm-referenced and criterion-referenced strategies in interpreting test results. Teachers, according to the study, scored examinations according to principles as well.

Hussain and Sajid (2015) conducted a review of test construction, including its cyclic formulation process, stages (deciding content, stating objectives, generating a table of specification, and fixing items), and evaluation methods (item difficulty and item discrimination). Teaching and testing are inextricably linked, and knowing the norms, standards, and ethics of test development and evaluation was essential for being professional when grading a student's performance.

Benzehaf (2017) surveyed 40 teachers in senior high schools in the El Jadida region to learn about their assessment techniques and identify the roadblocks to effective formative assessment in the classroom. Data were

gathered using questionnaire and an interview. From the results, a wide range of assessment processes are utilized. These covered the use of home-based assignments and in-class tests, although they were generally used for summative purposes. Benzehaf further mentioned that scoring essay-type assessments must adhere to certain guidelines.

Green, Johnson Kim, and Pope (2007) also looked at defining ethical conduct and analyzing educators' ethical judgements in connection to evaluation in their research. The results showed that educators were required to follow some ethical standards in the scoring of essay-type tests.

Teachers' scoring and grading techniques in basic schools were investigated by Anane and Adu-Mensah (2019). The study used a stratified sampling strategy to acquire data from 278 primary and junior high school teachers. From the results of this descriptive study, teachers considered students' regularity and promptness in class attendance, interest in the topic, clarity of work done, and the amount of questions answered while grading. Also, it was reported that there was no significant difference in the scoring practices of teachers from primary schools and those from junior high schools.

Furthermore, Quagrain (1992) performed a research in Ghana on instructors' abilities to utilize essay exams and discovered that while some teachers carried out a review of their essay tests, other teachers did not. Most of the teachers in the study of Quagrain created their scoring system after the examination, whereas only a minority did so before the test. Etsey (2006) also explored the test development and administering practices of primary school teachers in Ghana. His findings revealed that primary school teachers

moderately followed the standards required for test development and administering.

From the literature reviewed, it is realised that scoring essay-type tests presents significant challenges to teachers. This is because the scoring requires some skills and adherence to some principles.

### **Differences in Testing Practices Based on Demographic Characteristics (gender, teaching experience, professional qualification, and knowledge of test construction)**

There have been some studies which have investigated the role of demographic characteristics in the test construction skills of teachers. These demographic characteristics include gender, teaching experience, professional qualification, and knowledge of test construction. Oduro-Okyireh (2008) investigated whether pre-service testing training improves senior secondary school teachers' competency in real testing practice in Ghana's Ashanti Region. It was found that training provided before teaching favorably impacted on the testing practices of the teachers, however the effect was minor. In essence, teachers' test construction abilities were influenced by their test construction expertise. Because assessment competency is crucial to teacher effectiveness, the findings led to the proposal that every teacher get complete training in measurement and assessment before fully beginning their profession of teaching.

Yeboah (2018) investigated the assessment course's impact on the competences and skills among teachers in Ghana's Agona West and East Municipalities. The investigation was conducted using a descriptive survey approach. In all, 204 teachers handling core subjects from six public SHSs

were involved in the study by selecting them through purposive sampling procedure. The data were gathered via a questionnaire. The data were analysed using statistical procedures such as frequencies, percentages, and document content analysis. The study's findings revealed that, to a large extent, instructors stated that they gained competences and abilities as a result of taking the course, which aided them in planning a test while keeping the exam's aim in mind. The training, according to teachers, has assisted them in avoiding ambiguity in exam items. However, the course's material should be more practical, according to the respondents. From the results, it is evident that the kind of training and knowledge that teachers have influenced their test construction skills.

According to the findings of Yeboah (2018), teachers should use the abilities and skills learned in the course while engaging in assessment of their students. This would allow instructors to obtain reliable information about their learners in order to make the best judgments possible. Furthermore, instructors who were responsible for teaching the assessment course should use practical scenarios and instances to make the material more fascinating and engaging. This would help understand what is being taught, allowing them to apply their expertise to addressing challenges in a new setting. It would, once again, remove from the minds of some students the impression that the course is tough.

Inko-Tariah and Okon (2019) also evaluated lecturers' knowledge of test construction processes. The research used an analytical descriptive survey methodology, and it was discovered that lecturers' knowledge of test creation

techniques was unaffected by gender, years of experience, professional training, or educational degree.

Zhang and Burry-Stock (2003) examined the assessment strategies of teachers and the extent to which years of experience and training impacted on assessment practices. A MANOVA technique was used to examine data from 297 instructors who completed the Assessment Practices Inventory. It was realized that as students progressed through the grades, teachers were increasingly reliant on objective tests for classroom evaluation, and they became more worried about assessment quality. It was also shown that irrespective of the years of experience, teachers who had some level of training in measurement, reported stronger assessment abilities in testing and giving feedback on test results.

The influence of classroom evaluation procedures on students' success objectives were investigated by Alkharusi (2008). A total of 1,636 students and 83 science teachers were sampled from Muscat public schools. The students were mainly in the 9<sup>th</sup> grade. The findings of hierarchical linear modeling methods revealed that class setting characteristics, years of experience of teachers, and assessment procedures interacted considerably with students' characteristics in determining students' accomplishment objectives. Alkharusi also demonstrated that teachers' teaching experience influenced their ability to develop tests.

Acar-Erdol and Yldzl (2018) wanted to know how teachers from elementary, secondary and high schools assess their students in the classroom. The data for the study was gathered using surveys and observation forms. There were 288 teachers in the research group. It was revealed that teachers



chose “assessment for learning” method in their classrooms. In their classroom assessment techniques, teachers mostly employed traditional means of assessment. Teachers recognized student characteristics as the most important component impacting classroom assessment procedures, however it was observed that they did not represent this reality in their practices. It was recommended that teachers should adopt alternate assessment techniques that focus on student self-assessments, and evaluation method variety should be encouraged. The researchers also noted that the level of knowledge and experience of teachers affected their level of proficiency in test construction.

In a similar vein, Adodo (2014) discovered a link between instructors’ teaching experience and their ability to evaluate students’ cognitive achievement. As a result, the amount of teaching expertise had an impact on testing techniques. Din (2020) discovered that lecturers’ ability to develop tests is influenced by their degree of education and the skills they have acquired via training.

Moreover, Irvine (2019) found that the superior the teacher’s academic qualification, the better the teaching skills and assessment practices. Suah Irvine (2019) discovered that years of teaching experience influenced the assessment practices of teachers, as beginner teachers have a higher inclination of utilizing questions developed by other teachers. This signifies a lower perception of assessment competency. However, Moawad (2020) found that academic qualifications do not influence academic staff’s assessment practices. Moawad (2020) further revealed that assessment practices of teachers depended principally on the purpose they had set for the class, rather than their educational qualifications.

Salihu (2019) evaluated teacher competency in test development and Economics content validity with the focus on Nasarawa State senior secondary schools in Nigeria. The study used a content analysis and a co-relational research approach. A total of 95 Economics teachers responded to the “Teachers Ability Questionnaire on Test Construction (TAQTC)”. The data showed that a significant difference existed in test construction skills of teachers who were professionals and those who were not professionals. Also, significant difference was shown in the Economics content validity of teachers in public and private schools. According to the findings, item construction conferences/workshops should be held to improve teachers' ability to construct valid tests.

Other research, such as that of Dubem reported by Inko-Tariah and Okon (2019), found that lecturers' test construction competency was not reliant on the gender of the instructor since excellent test construction abilities are learned via training. Furthermore, Petters and Okon (2014) looked at gender and test construction knowledge and found no significant differences.

### **Chapter Summary**

The theoretical framework, conceptual review, and empirical review of the study were all looked over. The Classical Test Theory and the Item Response Theory were among the ideas examined. The study's concepts were also looked through. There was also a review of previous empirical literature. According to the findings, some teachers possessed the necessary abilities and followed conventional procedures in test development, administration, and scoring, while others did not.

## CHAPTER THREE

### RESEARCH METHODS

The aim of the current study was to examine the testing practices of basic school teachers in Ghana's Greater Accra Region. The methods used to carry out this investigation are described in this chapter. The research design, study area, population, sampling procedure, data collection instrument, data collecting procedures, and data processing and analysis are all covered in this chapter.

#### **Research Approach and Design**

In this work, a mixed methods approach was used. A mixed methods approach includes the researcher utilizing both qualitative and quantitative research methodologies in a study. A mixed methods research approach, incorporates the philosophical foundations of qualitative and quantitative procedures and the integration of both within a study. Mixed methods approach is used in the pragmatist paradigm. Creswell (2009) defines pragmatism as “a lack of commitment to a single philosophical system”. The implication is that, both quantitative and qualitative assumptions were considered in this study. In other words, the researcher chooses the quantitative and qualitative research procedures, strategies, and processes that suitably fulfill the study's goals. To clarify, mixed methods require more than just the collection and analyses of qualitative and quantitative data and extend to blending the two approaches such that the overall strength is higher than quantitative or qualitative on its own.

This study used a sequential explanatory mixed research technique in particular. A mixed methods approach is one that uses both quantitative and

qualitative approaches at the same time rather than two or more entirely quantitative or qualitative methods. They also claim that for a study to be categorized as mixed method research, it must incorporate design components that represent some or all of the qualitative and quantitative research aspects.

The mixed methods approach was chosen because it helped to obtain in-depth information from a large number of teachers about their testing practices. The quantitative and qualitative methods offset each other's weaknesses.

### **Study Area**

The Greater Accra Region happens to be the region which is small among the regions in Ghana in terms of geography. This region makes about 1.4 percent of Ghana's total land mass. However, in terms of population, the Ashanti Region region lies second behind the Greater Accra Region. The Greater Accra area is the most urbanized in Ghana, with most of the population living in cities. The Greater Accra Region's capital city is Accra, which is also Ghana's capital.

The area has a population of 5,455,692 people according to the 2021 census, making it Ghana's most populated (total number of inhabitants) region followed by the Ashanti Region with a population of 5,440,463 (Ghana Statistical Service (GSS), 2021). A major feature of the region is that there are several schools within the region and education is considered very fundamental in the region.

### **Population**

According to Creswell (2005), a population is a collection of individuals who share traits, such as all basic school teachers in a certain area. Teachers in public basic schools in the Greater Accra Region were the target

population. The Greater Accra Region is divided into 31 districts and municipalities which would constitute the target population. The accessible population, on the other hand, was made up of instructors from six municipalities selected schools in the Greater Accra Region. These are Ga North, Ga West, Ga South, Ablekuma North, Ningo Prampram and Ada West Municipalities. In all, there are 313 junior high schools in the six selected Municipalities. The total number of teachers in all these schools is 8,044. This data was obtained from the Regional Education Office which also indicated that there are over one thousand six hundred (1600) schools and over forty one thousand (41000) teachers within the district.

### **Sample and Sampling Procedures**

A sample in research, according to Kumar (1999), is a subset of the population that has been carefully chosen for a study. It is representative of the study's population. The study used a sample of 382 people. This is based on Singh and Musuka (2014) table for calculating sample size with a 95% confidence level and 5% accuracy. For this investigation, a simple random sample approach was applied. Stratified random sampling is an approach that divides the population into substantially homogenous sub-groups usually referred to as strata and selecting individuals from the groups randomly (Albright, Winston, & Zappe, 2010). The benefit of stratification is that it helps the researcher to guarantee that the study includes all of the population's subgroups and that the sample correctly reflects the population on the basis of certain features.

To select the six districts from the thirty-one, simple random sampling was used where the names of the districts were placed in a box after being

written on a piece of paper and then six selected to represent the population for the study. Ogah (2013) shared a similar viewpoint, stating that stratified sampling guarantees that relevant population features are accurately reflected in the sample. The people were divided into groups based on their districts.

The researcher guaranteed that the size of each sub-group in the sample is proportionate to the size of the group in the actual population by utilizing stratified random sampling. This means that the sample size for each stratum was proportional to the population's size. The formula for the computation was  $n = N / N \times S = s$ .

where

$n$  = Population of sub-group

$N$  = Main population

$S$  = Sample Size

$s$  = sample for sub-group.

For example,

$$1402/8044 \times 382 = 67.$$

The sample distribution is shown in table below.

District	Population	Sample
Ga North	1402	67
Ga West	1521	72
Ga South	1440	68
Ablekuma North	1354	64
Ningo Prampram	1172	56
Ada West Municipality	1155	55
Total	8044	382

Aside the 382 teachers who were sampled for the quantitative data, 30 teachers were sampled for the qualitative data. Several scholars have posited that a sample of six to 30 is adequate for a qualitative study (Saumure & Given, 2008). This is because qualitative research focuses on the quality of data collected rather than the quantity of persons involved. As a consequence, the qualitative part of the current study's total sample of 30 people was considered adequate.

The 382 and 30 respondents were chosen using a purposeful sample approach. In the view of Babbie (1990), purposeful sampling entails choosing a sample based on the investigator's own understanding of the population and the features, as well as the nature of the study objectives. Purposive sampling, in other terms, means that the population is picked non-randomly based on a certain feature. The 382 respondents for the quantitative study and 30 respondents for the qualitative study were sampled from the six districts, thus, five teachers from each district to be interviewed.

### **Data Collection Instrument**

A questionnaire and an interview guide were used in collecting data. A questionnaire is a series of questions about a single topic or a related range of topics that is presented to a randomly chosen group of people in order to collect information about an issue (Best & Kahn, 2006). Questionnaires provide anonymity and boost the likelihood of getting truthful replies for the researcher (Leedy & Ormrod, 2005). This was one of the reasons for using a questionnaire as one of the study's instruments.

However, there are certain drawbacks to using surveys. Questionnaires, according to Cohen, Manion, and Morrison (2003), do not

enable respondents to freely give responses as interviews do, and they do not give the researcher the chance to rectify misconceptions or wrong replies caused by confusion or misinterpretation of questions. This constraint will not affect the present study since the questionnaire will have items that are clearer and do not allow for misconceptions.

The questionnaire was adapted from the instrument of Anhwere (2009) on the test construction practices of teachers. The questionnaire had four sections. The first section (A) covered the demographic characteristics of the respondents. The second section (B) dealt with the practices teachers followed in constructing tests. The third section (C) covered the practices teachers followed in administering tests. The final section (D) covered the practices teachers followed in scoring essay tests.

An interview guide which was adapted from Osman (2021), was also used to collect the qualitative data. Specifically, semi-structured interview guide was used. As such, even though there are fixed questions, the length, depth and type of questions asked will be idiosyncratic. This is so because after asking a main question from the guide, follow-up questions were asked depending on the responses of the interviewee. The use of a semi-structured interviewing method was judged acceptable in order to provide participants some leeway in expressing their opinions and highlighting portions of specific attention (Ekanayake, 2011). In this regard, the decision to use the semi-structured interview guide helped to gather more intense and deep data. Each of the selected schools had one teacher interviewed. This means 30 teachers were interviewed in the study.



### **Validity of Questionnaire**

The validity of the questionnaire was established despite the difficulty in establishing construct validity. This was performed by emailing a working document of the questionnaire and the interview guide to my supervisor to check for content validity. Content validity has been viewed as the degree to which the elements on a data collection tool cover the whole range of what it is supposed to assess (Rubin & Babbie, 2016). As a consequence, the instrument was deemed to be content valid since the items on the questionnaire sufficiently assessed the study's objectives.

### **Reliability of Questionnaire**

After the pilot test, the Cronbach co-efficient alpha of the questionnaire was calculated to determine its reliability. After executing the pilot test, the total dependability co-efficient was found to be 0.789. The total dependability co-efficient achieved after the main investigation was 0.835. Both the pilot and the main research proved the reliability of the various portions of the questionnaire. For Section B, the pilot reliability co-efficient was 0.789 while the main data reliability coefficient was 0.814. For Section C, the reliability co-efficients were 0.802 (pilot) and 0.845 (main). Finally, for Section D, the reliability co-efficients were 0.792 (pilot) and 0.825 (main). The interview guide was pretested among five (5) participants to ensure that it met the criteria for its intended purpose before administering.

### **Pilot testing**

A pilot test was carried out to validate the instrument. The pilot testing was conducted with 50 teachers in the Greater Accra Region's Accra Metropolis. The instrument's pilot testing is crucial since it helped establish

the instrument's dependability as well as alter or examine some of the instrument's components.

### **Data Trustworthiness of Qualitative Data**

The trustworthiness of the qualitative data was established through the following components: (a) Credibility; (b) transferability; (c); dependability; and (d) confirmability. Triangulation of sources is thought optimal for proving the credibility of qualitative data. Triangulation of sources entails getting information from a variety of persons at various times and locations (Din, 2020). Aside from that, an inquiry audit, which entails having an outsider assess the data collecting technique, data analysis, and the study's primary findings, is an excellent way to determine the reliability of qualitative data.

### **Data Collection Procedures**

The researcher acquired ethical approval from the University of Cape Coast's College of Education Ethical Review Board prior to data collection. After this, an introduction letter was taken from the Department of Education and Psychology in the University of Cape Coast. This letter was delivered to the schools that had been chosen in order to acquire permission to gather data.

Teachers' informed permission was requested at each school. They were asked for their agreement and told that participation was completely optional and that they might opt out at any moment. The respondents would be given the questionnaire to answer after which the interview would follow simultaneously. Both the questionnaire and the interview would be done within twenty minutes. The study's goal was explained to them so that they could decide whether or not to participate. To guarantee a high return rate on

the questionnaires, the researcher gathered data in person. The data was collected during a four-week period. A 100% return rate was realized.

### **Ethical Considerations**

In this investigation, the researcher followed various ethical guidelines. The College of Education Ethical Review Board of the University of Cape Coast provided ethical approval. The study was found to be ethically sound after receiving ethics clearance. The researcher examined issues such as permission, autonomy, secrecy, anonymity, plagiarism, and discrimination. First and foremost, the participants' agreement was secured before to the study's execution.

The researcher verified that the individual had autonomy after gaining consent. The investigator did not impose his choices on the participants in this regard. They were given complete freedom in how they responded to the questionnaire and interview guide. In addition, the researcher guaranteed that the participants' identity was protected throughout the study. As a result, any personal information, such as names, that may lead to simple identification of individual respondents was not required in the study. However, data such as participant gender was gathered because the gender of participants was a crucial aspect for the study's goal.

In addition, the researcher verified that the information was kept private. Participants were guaranteed secrecy, and the researcher went to great lengths to secure it. Except in the case of a life or death danger, the data obtained was not shared with a third party without the respondents' agreement. Also, the data was kept in a secure location to guarantee that no one could use it without the researcher's permission.

### **Data Processing and Analysis**

On completion of data collection, data was manually checked for errors, coded, and put into the Statistical Product and Services Solution (SPSS) version 21. The quantitative research was both descriptive and inferential. The demographic data was descriptively analyzed using frequencies and percentages. The data for research questions 1, 2, and 3 were analyzed using means and standard deviation. Hypothesis one was tested using the Independent Samples t-Test, whereas hypotheses two, three, and four were tested using One-Way ANOVA. The qualitative data was analyzed using thematic analysis.

### **Chapter Summary**

The techniques for carrying out the study were described in this chapter. The study used a mixed techniques approach with a descriptive survey methodology. The study employed a sample of 382 teachers drawn from a population of junior high school teachers in chosen areas in Ghana's Greater Accra Region by stratified random selection. A questionnaire and an interview guide were used to collect data. Both descriptive and inferential statistics were used to analyze quantitative data. On the other side, theme analysis was used to analyze qualitative data. Ethical guidelines were followed.

## CHAPTER FOUR

## RESULTS AND DISCUSSION

**Introduction**

The aim of this study is to look into how basic school instructors test their students in Ghana's Greater Accra Region. The study's findings and comments are presented in this chapter. The demographic data of the respondents is shown first, followed by the primary data results.

**Demographic Characteristics of Respondents**

The demographic characteristics of the respondents cover their gender, educational qualification and years of teaching. These are shown in Table 1.

**Table 1.1: Demographic Characteristics of Respondents** N=382

Item	Frequency (f)	Percentage (%)
<b>Gender</b>		
Male	208	54.5
Female	178	45.5
<b>Highest Educational Qualification</b>		
Teacher's Diploma	91	23.8
HND	6	1.6
Degree	72	18.8
Degree (B. Ed)	194	50.8
Master's Degree	19	5.0
<b>Years of Teaching</b>		
Less than five years	53	13.9
Five years or more	329	86.1

Source: Field Survey (2021)

It can be seen in Table 1.1 that more than half of the respondents were males (54.5%) while the remaining 45.5% were females. This means more males were involved in the study than females. In terms of the educational qualification, it is shown that 194 respondents corresponding to 50.8% held

Bachelor of Education Degrees. Also, 23.8% of the respondents held Teachers' Diploma. The rest of the respondents either held University Degrees other than education (18.8%), Master's Degree (5%) or HND (1.6%). This data means that most of the teachers held Education Degrees. Finally, Table 1 showed that a majority of the respondents (86%) had taught in their schools for five years or more. This means that the respondents in the study had more years of teaching experience and as such could provide the relevant information needed in the study.

The respondents were asked to indicate the classes they teach and their respective subjects. The results obtained are shown in Tables 2 and 3.

**Table 1.2: Classes Taught**

Class Taught	Frequency (f)	Percentage (%)
Form 1	185	48.4
Form 2	240	62.8
Form 3	195	51.1

Source: Field Survey (2021)

Table 1.2 shows that 185 respondents corresponding to 48.4% were teaching Form 1 classes. Also, 62.8% of the respondents were teaching Form 2 classes. Finally, 51.1% of the respondents were teaching Form 3 classes. Since the question was a multiple response question, the respondents could indicate more than one class. The data means that some of the teachers were teaching multiple classes.

**Table 1.3: Subjects Taught by Teachers**

Subject Taught	Frequency (f)	Percentage (%)
Science	206	53.9
Mathematics	150	39.3
English	201	52.6
BDT	98	25.7
RME	134	35.1
Ghanaian Language	95	24.9
ICT	135	35.3

Source: Field Survey (2021)

Table 1.3 reveals that 53.9% of the respondents were teaching Science in their schools while 52.6% were also teaching English. The other subjects taught by the respondents were Mathematics (39.3%), ICT (35.3%), RME (35.1%), BDT (25.7%) and Ghanaian Language (24.9%). The data shows that some teachers were teaching more than one subject.

The respondents were further asked if they had ever taken a course in testing. The results are shown in Table 4.

**Table 1.4: Ever Taken a Course in Testing (Educational Measurement)**

Answer	Frequency (f)	Percentage (%)
No	131	34.3
Yes	251	65.7
Total	382	100.0

Source: Field Survey (2021)

It can be seen in Table 1.4 that 65.7% of the respondents indicated that they have ever taken a course in testing or educational measurement. About 34.3% had however not taken any course in testing or educational measurement. These results give a good indication that most teachers have a background in testing or educational measurement.

### **Demographic Characteristics of Interviewed Respondents**

The respondents who were interviewed were 30 in all. Out of these, 17 of them were males while 13 of them were females. In terms of age, 15 of them were aged between 26 and 40 years. Out of the remaining 15 respondents, 10 were aged above 40 years while the remaining five respondents were less than 26 years of age. In terms of educational level, 20 respondents indicated that they held Bachelor of Education Degree, 7 respondents held Teacher's Certificate while the remaining three respondents had University Degree but not in the field of Education. The respondents taught at JHS 1, 2 and 3 levels.

### **Answers to Research Questions**

**Research Question 1:** What testing practices do basic school teachers follow when constructing achievement test items?

This research question investigated the testing practices basic school teachers follow when constructing achievement test items. The data were analysed using mean and standard deviation. Using a scale of "Never=1", "Sometimes=2", "Very Often=3" and "Always=4", a cut-off of 2.5 was set. Thus, mean scores of 2.5 and above were deemed to be high while mean scores below 2.5 were deemed to be low. A higher mean implies that the respondents followed the said practice often while a lower mean indicates that respondents rarely followed the said practice. The results are presented in Table 5.



**Table 1.5: Testing Practices in Constructing Achievement Test**

Statement	Mean	SD
In the construction of achievement test:		
I establish the purpose of a test before constructing test items.	3.39	0.71
I connect the objectives of the lesson to the test.	3.53	0.64
I choose the format which suits the objectives for which the test is carried out.	3.36	0.80
I utilise a table of specifications in establishing which items should be part of the test.	2.66	0.91
I prepare more items than the total number of items which would be needed in the test.	2.51	1.01
I write the items on the test some considerable period before the test day so I can review and edit where needed.	2.96	0.88
I design the marking scheme at the moment the test is constructed.	3.19	0.92
I review items on the test after developing the test and putting aside for some time and re-reading them.	3.08	0.86
I write clear and specific directions for the whole test and parts of the test.	3.31	0.81
I copy test items directly from textbooks.	2.31	0.83
I evaluate the test to ascertain it's clarity, practicality, efficiency and validity	3.29	0.82

Source: Field Survey (2021)

The results in Table 1.5 reveal that while creating accomplishment tests, teachers link the subject matter's instructional objectives to the exam (M=3.53, SD=0.64). The respondents also indicated that in the construction of achievement test, they determine the purpose of a test before constructing test items (M=3.39, SD=0.71). The respondents also selected the test format suitable for testing stated objectives (M=3.36, SD=0.80). Aside this, the respondents indicated that in the construction of achievement test, they write

clear and specific directions for the whole test and parts of the test ( $M=31$ ,  $SD=0.81$ ).

Moreover, the respondents noted that they evaluate test to ascertain its clarity, practicality, efficiency and validity before constructing ( $M=3.29$ ,  $SD=0.82$ ). They also construct a marking system at the moment the test items are written ( $M=3.19$ ,  $SD=0.92$ ) and go over the test items after setting them away for a few days ( $M=3.08$ ,  $SD=0.86$ ). The data show that these were the most common techniques used by respondents when creating accomplishment assessments.

The respondents were also asked a series of questions on things they do during the construction of objective type tests. The results are shown in Table 6.

**Table 1.6: Construction of Objective Type Tests**

Question	Yes		No		Not Familiar	
	F	%	F	%	F	%
Do you usually do distracter analysis of your multiple-choice test items?	223	58.4	83	21.7	76	19.9
Do you usually estimate the difficulty level of your objective test items?	264	69.1	77	20.2	41	10.7
Have you ever computed the discrimination index of your objective type tests?	79	20.7	151	39.5	152	39.8

Source: Field Survey (2021)

As indicated in Table 1.6, more than half of the respondents (58.4%) usually did distracter analysis of multiple-choice test items. In a similar vein, majority of the respondents (69.1%) usually estimated the difficulty level of objective test items. In terms of computing discrimination index of objective

type tests, 39.8% of the respondents indicated that they were not familiar with discrimination index while 39.5% indicated that they did not compute discrimination index. Only 20.7% of the respondents indicated that they did discrimination index of objective type tests.

Further, the respondents were asked to indicate what they considered in the arrangement of objective test items. The results from the respondents are shown in Table 7.

**Table 1.7: Considerations in Arrangement of Objective Test Items**

Considerations	Frequency (f)	Percentage (%)
The nature of item used.	37	9.7
The level of challenge of the different items (from the easiest to the most challenging).	58	15.2
The specific objectives being assessed.	186	48.7
The subject matter being assessed.	77	20.2
None of the above	24	6.3

Source: Field Survey (2021)

Table 1.7 shows that 48.7% of the respondents considered the learning outcomes being measured in the arrangement of objective test items. Also, the subject matter being measured was also considered by 20.2% of the respondents in the arrangement of objective test items. About 15% also considered the level of challenge (easiest to most challenging) in the arrangement of objective test items. These were the main considerations of the respondents in the arrangement of objective test items.

Finally, the respondents were asked to how they get to know the time to allot for essay tests. The results are shown in Table 8.

**Table 1.8: Determining Time to Allot for Essay Test**

How to know time allotted is adequate	Freq. (f)	Percent (%)
By using.....		
the number of items on the test to determine the amount of time needed	168	44.0
the time that the test developer would take to complete the test	33	8.6
the time that about 90% of the test takers would need to complete the test.	157	41.1
Other	24	6.3

Source: Field Survey (2021)

Table 1.8 reveals that 44% of the respondents indicated that they considered the number of items on the test in determining the amount of time to allot for the test. Also, 41.1% of the respondents indicated that they consider the time that about 90% of the test takers would need to complete the test in allotting time for essay tests. These were the two main issues considered by the respondents in allotting time for essay tests.

The respondents who were interviewed expressed views which were not entirely different from the results obtained from the questionnaire. Among the views expressed, the common themes identified were “considering the purpose of the test”, “considering the topic and objectives of the topic”, “ensuring clarity and unambiguity” and “considering the characteristics of the learners”. From the themes, the respondents highlighted the fact that in the construction of test items, the practices they followed were giving considerations to reasons for the test, the topic of interest and objectives, clarity and learner characteristics. Some of the actual statements of the respondents are shown below:

*“I usually ensure that test follow clearly defined learning outcomes. I also make sure the test items are appropriate for measuring the desired learning outcomes and it should fit the particular uses that will be made of the results.” – IR 3*

*“For me, I mostly consider the purpose of the test and ensure that the items meet the purpose and objective of the test. I also prepare marking scheme at the same time test items are being written.” – IR 6*

*“I set test items to cover all topics, avoid ambiguity and make sure the items are at the level of the students.” – IR 2*

**Research Question 2:** What testing practices do basic school teachers follow when administering achievement test items?

This research question investigated the testing practices basic school teachers follow when administering achievement test items. The data were analysed using mean and standard deviation. Using a scale of “Never=1”, “Sometimes=2”, “Very Often=3” and “Always=4”, a cut-off of 2.5 was set. Thus, mean scores of 2.5 and above were deemed to be high while mean scores below 2.5 were deemed to be low. A higher mean implies that the respondents followed the said practice often while a lower mean indicates that respondents rarely followed the said practice.

First and foremost, respondents were asked to rate how frequently they included their pupils in each of the following processes when preparing them for the test. Table 9 summarizes the findings.

**Table 1.9: Preparing Students in Advance for Test**

Statement	Mean	SD
In preparing students in advance for the test, I make them aware of:		
When (date & time) the test will be administered.	3.52	0.75
The number of items and the place where the test would be administered.	3.12	0.85
The specific topic areas that the test would cover.	3.12	0.83
The test format (objective type or essay-type tests).	3.21	0.81
The weighting of the different areas of the test.	2.90	0.85
How the test would be scored and graded.	2.93	1.01
The rules and regulations that would govern the administering of the test	3.31	0.82
The importance of the results of the test.	3.15	0.89

Source: Field Survey (2021)

The results in Table 1.9 show that in preparing students in advance for test, the respondents often made them aware of when (date and time) the test will be administered ( $M=3.52$ ,  $SD=0.75$ ). Also, the respondents indicated that they often made students aware of the rules and regulations governing the administering of the test ( $M=3.31$ ,  $SD=0.82$ ). The respondents made it clear again that the format of the test (objective type or essay-type tests) ( $M=3.21$ ,  $SD=0.81$ ) and the importance of the results of the test ( $M=3.15$ ,  $SD=0.89$ ) are often made known to students in advance.

Further, the respondents were required to indicate how frequently they follow some test administration practices. The results are shown in Table 10.

**Table 2.0: Test Administration Practices**

Test Administration Practices	Mean	SD
I do not give tests at periods before or just after a long vacation, or some significant events.	2.55	1.00
I make sure sitting arrangement allows for fairness and no cheating.	3.27	0.77
I make sure there is good ventilation and lighting in the place for the test.	3.20	0.93
I use “Do Not Disturb. Examinations In Progress” sign when test is going on.	2.18	1.13
During examinations, I anticipate and take care of all likely emergencies.	2.81	1.03
I make announcements on the time to take the test at regular intervals during testing.	3.64	0.62
During examinations, I take a position where I have a view of the entire room so I can monitor any malpractices.	3.69	0.60
I keep remarks in the testing room to a minimum and only make a few which are directly connected to the test.	3.16	0.86
I inform students on the awful consequences of failure in the test going on.	2.93	0.95
I ask students to work at a fast rate during the testing so as to complete on time.	3.01	1.02

Source: Field Survey (2021)

Table 2.0 shows that during tests, respondents frequently take position where they can view the entire room and monitor malpractices ( $M=3.69$ ,  $SD=0.60$ ). During the test administration, they often make announcements of the time left for taking the tests at regular intervals ( $M=3.64$ ,  $SD=0.62$ ). The respondents often checked and made sure that the seating arrangement ensured fairness and no cheating ( $M=3.27$ ,  $SD=0.77$ ), as well as that the testing room has appropriate air and illumination ( $M=3.20$ ,  $SD=0.93$ ). Respondents also restricted their comments in the testing room to a bare minimum and made sure they were relevant to the exam ( $M=3.16$ ,  $SD=0.86$ ). These were the most common test administration practices among the responders.

The interviewed respondents shared their views on the test practices they follow in the administration of test. The common themes identified were “giving prior information on the test”, “proper seating arrangement”, and “determining the right time and venue for test”. These were the main practices of the respondents in terms of test administration. Some of the actual statements are quoted below:

*“In administering achievement test, I usually ensure that the candidates/ pupils are well spaced to avoid cheating/ malpractices.”* – IR 14

*“I give the learners all the information about the test before the test day. I let them know the importance of the test and the time for the test and the topic areas.”* – IR 21

**Research Question 3:** What testing practices do basic school teachers follow when scoring essay-type test items?

This research question aimed at finding out the testing practices basic school teachers follow when scoring essay-type test items. In answering this research question, several questions were posed to the respondents. In the first place, the respondents were asked to indicate which method they use in scoring essay tests. The results are shown in Table 11.

**Table 2.1: Method Used in Scoring Essay Tests**

Method	Frequency (f)	Percentage (%)
Holistic	90	23.6
Analytic	279	73.0
Other	13	3.4
Total	382	100.0

Source: Field Survey (2021)



The results in Table 2.1 show that the majority of respondents (73 percent) scored essay tests using the analytic technique. The analytic technique employs a grading system in which points are assigned to concepts and points that are clearly presented. This implies that most of the respondents do not use their personal impressions in marking essay tests.

The respondents were also asked to indicate how they go about the scoring of essay tests. The responses are shown in Table 12.

**Table 2.2: Scoring of Essay Tests**

Method	Frequency (f)	Percentage (%)
Student by student	303	79.3
Item by item	79	20.7
Total	382	100.0

Source: Field Survey (2021)

From the results in Table 2.2, it is clear that majority of the respondents (79.3%) scored all the items answered by each student before proceeding to the next student. Thus, they used the “student by student” approach in scoring.

Further, the respondents were asked a series of questions on their general practices in the scoring of essay tests. The results of these are shown in Table 13.

**Table 2.3: Test Scoring Practices**

Statement	Yes		No	
	f	%	F	%
In scoring essay tests, do you constantly follow the marking scheme as you score?	250	65.4	132	34.6
In scoring essay tests, do you randomly rearrange the answer scripts after scoring each item?	132	34.6	250	65.4
In scoring essay tests, do you score all answers to a specific question at a sitting without disruption?	138	36.1	244	63.9
Do you score essay tests only when you are physically and mentally well and attentive?	270	70.7	112	29.3
Do you keep previously scored items out of sight during the time of scoring the rest of the items?	226	59.2	156	40.8
In scoring essay tests, do you write comments on the answer scripts to assist students in their learning?	288	75.4	94	24.6
Do you score the answer scripts of essay tests with the names of test-takers that you know?	108	28.3	274	71.7
In scoring essay tests, do you score the expressions like handwriting, general neatness, spelling etc., separately from the accuracy of the content?	263	68.8	119	31.2

Source: Field Survey (2021)

It can be seen from the results that majority of the respondents (65.4%) answered “Yes” to the question of whether they constantly follow the marking scheme as they score essay tests. Majority of the respondents (65.4%) however answered “No” to the question on whether they randomly rearranged

the answer scripts after scoring each item. Similarly, 63.9% answered “No” to the question on whether they scored all answers to a particular question at a sitting without disruption. Thus, the respondents did not score all responses at a sitting without interruption. In addition, 70.7% of the respondents indicated that they scored essay tests only when they were physically and mentally well and attentive. About 75.4% of the respondents also indicated that they wrote comments on the answer scripts to assist students in their learning.

Regarding the interpretation of results of tests, the respondents were asked a series of questions. In the first place, the respondents were asked to indicate the method they use in interpreting results of test. Their responses are shown in Table 2.4.

**Table 2.4: Method used in Interpreting Results of Test** N=382

Method	Frequency (f)	Percentage (%)
Norm-referenced interpretation	276	72.5
Criterion-referenced interpretation	306	80.1

Source: Field Survey (2021)

From Table 14, it can be seen that 80.1% of the respondents used criterion-referenced interpretation. This means that they described students' level of performance in terms of the learning tasks they can do. Also, 72.5% of the respondents indicated that they used norm-referenced interpretation which means they described students' level of performance in relation to other members of the class. The question was multiple-response question and so the respondents could choose more than one method. In essence, there are times that respondents used norm-referenced interpretation and there are times that

the respondents used criterion-referenced interpretation. However, criterion-referenced interpretation was used more than norm-referenced interpretation.

As a follow up, the respondents were asked to indicate the specific method used in each form of interpretation. The results for norm-referenced interpretation are shown in Table 15.

**Table 2.5: Method used in norm-referenced interpretation**

Method	Frequency (f)	Percentage (%)
Simple ranking of raw scores	129	46.8
Percentile ranks	108	39.1
Stanine system of standard scores	39	14.1
Total	276	100.0

Source: Field Survey (2021)

For those who used norm-referenced interpretation, Table 2.5 shows that 46.8% of them used simple ranking of raw scores while 39% used percentile ranks. The remaining 14% used stanine system of standard scores.

The results for criterion-referenced interpretation are shown in Table 16. In Table 16, it is shown that more than half of the respondents (57.2%) indicated that they use standard of 50% correct score in interpretation of scores while about 36.9% of the respondents set arbitrary standard and adjust up and down in interpretation of test scores.

**Table 2.6: Method used in criterion-referenced interpretation**

Method	Frequency (f)	Percentage (%)
Standard of 50% correct score	175	57.2
Set arbitrary standard and adjust up and down	113	36.9
None of the above	18	5.9
Total	306	100.0

Source: Field Survey (2021)

Finally, depending on their test results, the respondents were asked how they treat low-achieving students (i.e. pupils who fail to achieve specified educational objectives) in their class. Because this was an open-ended topic, the answers had a wide range of opinions. Among the views, some common themes were derived. They include the following:

- Different instruction in remedial teaching
- Remedial lesson with same subject matter
- Individual focused attention
- Further studies for such students
- Counselling and encouragement

Some of the specific statements of the respondents are quoted below:

*For me, I usually provide some remedial lessons for students who are not able to meet set targets in their results.* – Teacher 10

*I usually arrange individual counselling sessions for students who struggle to find out how to help them.* – Teacher 13

*I pay attention to students who struggle during the teaching period.* –  
Teacher 7.

For the respondents who were interviewed, they highlighted some practices they followed in scoring essay type tests. These practices included using analytic or holistic scoring, scoring without looking at the names, preparing scoring rubrics ahead of time and avoiding interruptions during marking. Some specific quotes are as follows:

*“Usually, when scoring essay tests, I make sure I have a scoring guide and I try as much as possible to avoid interruptions when scoring.”* –  
IR 2

*“Either I use holistic or analytic scoring style; I use any of them depending on the questions.”* – IR 23

*“I try to avoid anything or activity which may compromise the scoring of the tests. So I avoid interruptions to the best of my ability.”* – IR 3

**Hypothesis One:**

H<sub>0</sub>: There is no statistically significant difference in the testing techniques of elementary school instructors based on their gender.

H<sub>1</sub>: There is a statistically significant variation in the testing techniques of elementary school instructors based on their gender.

The goal of this hypothesis was to see if there was a substantial variation in basic school instructors' testing procedures based on their gender. The hypothesis was tested using an independent samples t-test with a significance level of 0.05.

Before doing the independent t-test, the normality and homogeneity tests were completed.

The results are shown in Tables 17 and 18.

**Table 2.7: Tests of Normality (Testing Practices and Gender)**

	Gender	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Practices	Male	.096	208	.210	.972	208	.708
	Female	.143	174	.180	.975	174	.884

Source: Field Survey (2021)

It is expected that the data for independent samples t-test should be normally distributed. Normality was tested using the Kolmogorov-Smirnov and Shapiro-Wilk statistics. It can be seen in Table 2.7 that the significant values are all above .05. From the table therefore, normality can be assumed for the data.

The results of the Levene's test for homogeneity of variance are presented in Table 18.

**Table 2.8: Levene's Test for Equality of Variances**

	F	Sig
Equal variances assumed	.047	.828
Equal variances not assumed		

Source: Field Survey (2021)

From Table 2.8, it could be seen that the significant value of .828 is greater than .05 the significant level. This implies that equal variances can be assumed. Based on this, it was appropriate to run independent samples t-test. The results of the independent samples t-test are shown in Table 19.

**Table 2.9: Results of t-Test Comparing Testing Practices of Male and Female Teachers**

Gender	N	Mean	SD	Df	t-value	Sig (2-tailed)
Male	208	119.86	10.19	380	.620	.536
Female	174	119.21	10.18			

Source: Field Survey (2021)

p>.05

Table 2.9 shows the results of the independent samples t-test. Table 19 shows that there is no statistically significant difference in testing techniques between male and female teachers [ $t(380) = .620, p > .05$ ]. Males had a mean score of 119.86, while females had a score of 119.21. Although there is a minor change in the mean scores, it is not statistically significant. The null hypothesis that there is no significant variation in basic school teachers' testing techniques based on their gender was accepted based on the data in Table 19. This suggests that male and female teachers used the same assessment methods.

**Hypothesis Two:**

$H_0$ : There is no statistically significant difference in the assessment techniques of elementary school instructors based on their years of experience.

$H_1$ : There is a statistically significant variation in the testing techniques of elementary school instructors based on their years of teaching experience.

The goal of this hypothesis was to see if there was a substantial difference in basic school teachers' testing procedures based on their years of experience.

The hypothesis was tested using an independent samples t-test with a significance level of 0.05. This is due to the fact that teaching experience was divided into two categories (less than 5 years and five years or more). Before doing the independent t-test, the normality and homogeneity tests were completed.

The results are shown in Tables 20 and 21.



**Table 3.0: Tests of Normality (Testing Practices and Years of Teaching)**

	Years	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Practices	5 or more	.079	278	.146	.982	278	.193
	Less than 5	.112	104	.200	.963	104	.808

Source: Field Survey (2021)

Normality was tested using the Kolmogorov-Smirnov and Shapiro-Wilk statistics. It can be seen in Table 3.0 that the significant values are above .05. From the table therefore, normality can be assumed for the data.

The results of the Levene's test for homogeneity of variance are presented in Table 3.1.

**Table 3.1: Levene's Test for Equality of Variances**

	F	Sig
Equal variances assumed	1.579	.210
Equal variances not assumed		

Source: Field Survey (2021)

From Table 21, it could be seen that the significant value of .210 is greater than .05 the significant level. This implies that equal variances can be assumed. Based on this, it was appropriate to run independent samples t-test.

The results of the independent samples t-test are shown in Table 22.

**Table 3.2: Results of t-Test Comparing Testing Practices of Teachers on basis of Years of Experience**

Years	N	Mean	SD	Df	t-value	Sig (2-tailed)
5 years or more	278	122.24	9.20	380	3.184*	.001
Less than 5 years	104	118.56	10.36			

Source: Field Survey (2021)

\*Significant,  $p < .05$

It is shown in Table 2.8 that there is a significant difference in the testing practices of teachers on the basis of the years of teaching experience [t

(380) = 3.184,  $p < .05$ ]. The mean score of those who had taught for five years or more was 122.24 while those who had taught for less than five years had a mean score of 118.56. Thus, teachers who had taught for more years had better testing practices. Based on the results in Table 8, the null hypothesis that there is no significant difference in the testing practices of teachers on the basis of teaching experience was rejected. This means that teachers differed in their testing practices on the basis of their teaching experiences.

### Hypothesis Three:

$H_0$ : There are no statistically significant differences in the testing techniques of elementary school instructors based on their professional qualifications.

$H_1$ : There is a statistically significant variation in the testing techniques of elementary school instructors based on their professional qualifications.

This hypothesis sought to determine whether there was a substantial variation in basic school teachers' testing techniques based on their professional credentials. At the 0.05 level of significance, one-way analysis of variance was utilized to test the hypothesis. This is due to the fact that professional qualifications were divided into five categories. Before doing the independent t-test, the normality and homogeneity tests were completed.

The results are shown in Tables 23 and 24.

**Table 3.3: Tests of Normality (Testing Practices and Educational Qualification)**

Practices	Qualification	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	Df	Sig.
	Teacher's Diploma	.097	91	.076	.956	91	.503
	HND	.407	6	.164	.640	6	.702
	University Degree	.120	72	.158	.949	72	.698
	University. Deg. B.Ed.	.082	194	.072	.970	194	.611
	Master's Degree	.332	19	.161	.818	19	.585

Source: Field Survey (2021)

The significant values across all the qualifications are higher than .05. This gives the indication that the data was normal. As a result, normality can be assumed for the data.

In using One-Way ANOVA, there was also the need to assess the homogeneity of variance among the groups. The results of the Levene's test for homogeneity are shown in Table 3.4.

**Table 3.4: Test of Homogeneity of Variances**

Levene Statistic	df1	df2	Sig.
1.106	4	377	.353

Source: Field Survey (2021)

It is shown in Table 3.4 that the significant level of .151 is greater than .05. This implies that homogeneity of variances can be assumed. Therefore, it is appropriate to carry out One-Way ANOVA.

The results of the ANOVA test are shown in Tables 25, 26 and 27.

**Table 3.5: Descriptive Results for Different Qualifications**

Qualifications	N	Mean	Std. Dev.
Teacher's Diploma	91	118.92	8.39
HND	6	115.00	7.75
University Degree	72	116.33	10.45
University Degree (B.Ed.)	194	120.94	10.14
Master's Degree	19	120.26	15.07
Total	382	118.29	10.18

Source: Field Survey (2021)

Table 3.5 shows the means and standard deviations of the various qualifications. It can be seen that teachers with University Degree (B.Ed.) had the highest mean score of 120.94 and a standard deviation of 10.14. Next to this was the teachers with Master's Degree (M=120.26, SD=15.07). The teachers with HND recorded the lowest mean score of 115.00. From the mean

scores, it is clear that there are differences among the different qualifications. Specifically, it is clear that, teachers with University Degree (B.Ed.) and Master's Degree had better testing practices than those with the other qualifications.

The results of the ANOVA test to reveal the significance of the difference observed among the groups are presented in Table 26.

**Table 3.6: ANOVA Results Comparing Testing Practices on the basis of Qualification**

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	1179.593	4	294.898	2.901*	.022
Within Groups	38328.522	377	101.667		
Total	39508.115	381			

Source: Field Survey (2021)

\*Significant,  $p < .05$

Table 3.6 shows that there is a significant variation in teacher testing techniques depending on their qualification [ $F(4, 377) = 2.901, p.05$ ]. The probability value (p-value) of 0.022 is lower than the .05 level of significance. This means that there was a statistically significant difference in the mean scores of the five distinct certifications. The null hypothesis, that there was no statistically significant difference in basic school teachers' testing techniques based on their professional qualifications, was shown to be false. Because a significant difference was discovered, a post-hoc analysis was required to establish which of the three mean values was responsible for the significant difference in the ANOVA findings.

Tukey's Post-Hoc test was used in doing the post-hoc analysis. The main aim of Tukey's test is to find out which groups in a specific sample differ from each other. It uses the "Honest Significant Difference," a number

that is a representation of the gap between groups and is used to equate one mean to another. The results are presented in Table 27.

**Table 3.7: Tukey HSD Comparison of Different Qualifications**

Highest Educational Qualification	N	Subset for alpha = 0.05
HND	6	115.00
University Degree (BA, B.Sc.)	72	116.33
Teacher's Diploma	91	118.92
Master's Degree	19	120.26
University Degree (B.Ed.)	194	120.94
Sig.		.586

Source: Field Survey (2021)

\*Significant,  $p < .05$

It is shown in Table 3.7 that the respondents with University Degree (B.Ed.) with a mean score of 120.94 had better testing practices compared to the other qualifications. Next to this, was, the teachers who had Master's Degree. The respondents with HND and University Degree (BA, B.Sc.) had lower mean scores indicating poorer testing practices.

#### **Hypothesis Four:**

$H_0$ : There is no statistically significant difference in basic school teachers' testing practices on the basis of their knowledge on test construction.

$H_1$ : There is statistically significant difference in basic school teachers' testing practices on their knowledge on test construction.

This hypothesis was aimed at finding out the significant difference in basic school teachers' testing practices on their knowledge on test construction. Knowledge on Test Construction was measured with a question on whether they had taken any course in testing. Those who responded 'Yes' was considered one group while those who responded 'No' was considered

another group. The independent samples t-test was used in testing the hypothesis at 0.05 level of significance. Normality and homogeneity tests were done first before carrying out the independent t-test.

The results are shown in Tables 28 and 29.

**Table 3.8: Tests of Normality (Testing Practices and Knowledge)**

Gender	Kolmogorov-Smirnov			Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	Df	Sig.	
Practices	No	.089	131	.135	.979	131	.242
	Yes	.097	251	.200	.968	251	.765

Source: Field Survey (2021)

The Kolmogorov-Smirnov and Shapiro-Wilk tests were observed. It can be seen in Table 28 that the significant values are above .05. From the table therefore, normality can be assumed for the data.

The results of the Levene's test for homogeneity of variance are presented in Table 3.9.

**Table 3.9: Levene's Test for Equality of Variances**

	F	Sig
Equal variances assumed	1.202	.274
Equal variances not assumed		

Source: Field Survey (2021)

In Table 3.9, it could be seen that the significant value of .274 is greater than .05 the significant level. This implies that equal variances can be assumed. The results of the independent samples t-test are shown in Table 30.

**Table 4.0: Results of t-Test Comparing Testing Practices of Teachers on basis of Knowledge in Testing**

Course in Testing	N	Mean	SD	Df	t-value	Sig (2-tailed)
No	131	118.12	9.42	380	-1.607*	.035
Yes	251	119.79	10.56			

Source: Field Survey (2021)

\*Significant,  $p < .05$ 

The results of the independent samples t-test are shown in Table 4.0. It can be seen that there is a significant difference in the testing practices of teachers on the basis of their knowledge in testing [ $t(380) = 1.607, p < .05$ ]. The mean score of the respondents who had taken a course in testing was high (119.79) than that of the respondents who had not taken a course in testing (118.12). This means that those who had taken a course in testing had better testing practices than those who had not taken a course in counselling. On the basis of the results, the null hypothesis that there is no significant difference in the testing practices of teachers on the basis of knowledge in testing was rejected.

## Discussion

### Constructing Achievement Test Items

The study discovered that teachers establish the aim of a test before creating test items while creating achievement tests. Teachers often link the test to the subject matter's teaching objectives. Because the objectives are linked to the exam, the teachers also chose a test style that would be appropriate for testing the stated objectives. Aside from that, respondents said they create clear and simple guidelines for the full exam and components of the test when creating accomplishment tests. Furthermore, the respondents stated that before developing a test, they analyze it for clarity, practicality,

efficiency, and validity. The teachers also devised a scheme for scoring as soon as the test items were written, and went over the test items after a few days had passed.

These findings are consistent with those of Armah (2018), who investigated how well lecturers followed standard practices in developing and administering tests at the University of Cape Coast. Lecturers stated that they specified the goal of each test before it was developed, selected the format of the test items, and developed a grading system just at the time when the test was developed.

Similarly, the data support Koloi-Keaikitse (2017)'s conclusion that teachers usually felt more proficient in test building procedures. Similarly, Inko-Tariah and Okon (2019) examined lecturers' knowledge of test construction processes and found that they had a good level of understanding of test building procedures.

Majority of the respondents usually did distracter analysis of multiple-choice test items and also usually estimated the difficulty level of objective test items. In terms of computing discrimination index of objective type tests, most of the respondents were not familiar with discrimination index or did not compute discrimination index. The present study's findings are in consonance with DiBattista and Kurzawa's (2011) findings, which looked at undergraduates' responses to 1198 multiple-choice questions on sixteen classroom assessments in diverse fields. They discovered that the quality of many multiple-choice exams might be significantly improved since item analysis and discriminating index analysis were not performed properly.



Caldwell and Pate (2013) also looked at how three different test item formats affected item statistics and student performance. They found that item discrimination practices were not followed much and so led to poor student performance. This was confirmed in the current study. In test construction, Quaigrain and Arhin (2017) investigated the link between difficulty index and discrimination index (DI) and distractor efficiency. To increase the quality of the evaluation, Quaigrain and Arhin discovered that items with moderate difficulty and good discriminating power with functional distractors should be included in future examinations. This is because without using DI most test times are not deemed to be the best.

When creating objective exams, teachers looked at the learning goals to be assessed, the subject matter to be measured, and the level of challenge posed by the items (arranged from easiest to most difficult). Finally, while allocating time for essay examinations, respondents reported that they took into account the quantity of items on the test as well as the time it takes roughly 90% of students to finish the test.

This study's findings corroborate those of Ollennu and Etsey (2015), who discovered that the sequence in which elements on an objective exam are ordered or arranged is an essential practice to follow. According to Etsey (2004), item arrangement must go from easier to more challenging forms in test creation, and each part must be ordered in order of increasing complexity. In a similar vein, Sax and Cromack (2005) revealed that testing specialists have advocated that during scheduled examinations, things should go from simple to challenging with the aim of building up the confidence in

the students. The findings discussed give the indication that test practices for test construction are mostly followed even though not all aspects.

### **Administering Achievement Test Items**

The study discovered that when teachers prepared students for tests in advance, they frequently informed them of the test's date and time. In addition, the respondents stated that they often made pupils aware of the laws and regulations controlling the test's conduct and format (objective type or essay-type tests). The importance of the results of the test was also often made known to students in advance.

These findings corroborate Armah's (2018) findings, which showed that instructors were knowledgeable of test administration concepts. During the exam administration, the lecturers informed pupils about the test's norms and restrictions. Hudaya (2017) also looked at teachers' assessment literacy in terms of their level of preparedness in evaluating students' performance, and discovered that the majority of instructors used test administration concepts.

Respondents said they typically stood where they can observe all students during examinations and move among them once in a while to look for malpractices. They often mentioned the amount of time left for the test intermittently during the exam administration. The responders often checked that the seating arrangement offers enough distance for students not to copy from one another, as well as that the exam area has appropriate air and illumination. In addition, responders frequently kept their comments in the testing room to a minimal and ensured that they were connected to the exam. These were the main practices followed by the respondents in the administration of tests.

The findings support the findings of Tyler (1991) that in test administration, teachers are expected to plan in detail for the administering of tests, use proper seating arrangements to avoid malpractices, and make intermittent announcements about the time for the tests. Rukundo and Magambo (2010) also showed that it is necessary to discourage cheating by special seating arrangements and careful supervision during test administration. The evidence is clear from the studies discussed that, testing practices for test administration are followed to some extent.

### **Scoring Essay-Type Test Items**

The majority of respondents employed the analytic technique to score essay assessments, according to the survey. The analytic technique employs a grading system in which points are assigned to concepts and points that are clearly presented. This indicates that the majority of respondents do not utilize personal impressions when grading essay examinations. In addition, the majority of respondents gave each student a score for all of the things they answered before moving on to the next student. Thus, they used the “student by student” approach in scoring.

These findings support Guskey's (2006) findings that instructors put in massive efforts to make sure that grading processes are as clear, unambiguous, and unbiased as feasible. This ensures that, above all, instructors' grading systems and processes are fair and unbiased, since they avoid utilizing personal ideas and insentient biases as influences when scoring and giving grades.

In terms of general scoring methods for essay examinations, the majority of respondents stated that they always grade essay tests according to

the marking scheme. The majority of respondents also stated that they do not rearrange the scripts they are marking randomly after marking each item, and that they do not score all replies to a single question in one session. The majority of the professors in the survey also stated that they only graded essay examinations when they were physically and intellectually fit, and that they offered students with comments on their response scripts to help them learn.

These findings were similar to those of Hussain and Sajid (2015), who conducted a survey of test construction and discovered that knowing and practicing the norms, standards, and ethics of test construction and evaluation is important in test scoring. Green, Johnson, Kim, and Pope (2007) examined educators' ethical judgements in regard to assessment and found that they were obliged to observe some ethical principles while scoring essay-type assessments.

In addition, Quagrain (1992) conducted study in Ghana on instructors' capacity to use essay tests, finding that even though some teachers verified their essay test items, some other teachers did not do so. Only a minority of teachers devised their grading scheme before the exam, whereas the majority did it afterward. Etsey (2006) performed study in Ghana to determine primary school teachers' test development and administration standards, as well as whether service length influenced assessment practice. The data suggested that the test construction and administration quality of elementary school instructors were low.

Regarding the interpretation of results of tests, criterion-referenced interpretation was used more than norm-referenced interpretation. This means that they described students' academic achievement in relation to the academic

tasks they could do level more than describing academic achievement with regard to other members of the class. The specific norm-referenced interpretations used were simple ranking of the actual scores obtained by students and percentile ranks. The specific criterion-referenced interpretations used were standard of 50% correct score and arbitrarily giving a score to describe maximum performance and adjusting such score up and down in interpretation of test scores.

The findings of Oduro-Okyireh (2008) that SHS teachers in the Ashanti Region followed the standard practices in their testing techniques, backed up the finding of this study. Oduro-Okyireh added that teachers reported using both norm-referenced and criterion-referenced techniques in how they interpreted test scores. Finally, respondents in this survey stated that they gave remedial classes, paid attention to individual students throughout sessions, and counseled and encouraged students who failed.

### **Differences in Testing Practices Based on Gender**

It was shown in this research again that there was no significant difference in testing techniques between male and female teachers. Although there were minor variations in the mean scores, they were not significant when compared statistically. The null hypothesis, that there was no substantial variation in basic school instructors' testing techniques based on their gender, was accepted based on the findings. This suggests that male and female teachers used the same assessment methods.

The findings support those of Inko-Tariah and Okon (2019), who examined lecturers' knowledge of test construction methods and discovered that test construction techniques did not differ substantially based on gender.

Similarly, Dubem, quoted by Inko-Tariah and Okon, found that lecturers' test construction competency was not reliant on the gender of the teacher since good test construction abilities are learned via training. Furthermore, Din (2020) looked at gender and test construction knowledge and found no significant differences.

From the findings it has become apparently evident that gender is not of essence in terms of test construction practices. This is not doubted since test construction practices are learned and not by birth or gender.

### **Differences in Testing Practices Based on Teaching Experience**

The findings also revealed that instructors' assessment procedures differed significantly depending on their years of teaching experience. Teachers with more years of experience have superior testing techniques. The null hypothesis that there is no significant variation in teacher testing techniques based on teaching experience was rejected based on the findings. This suggests that teachers' assessment procedures varied depending on their prior teaching experiences. Because teachers with many years of expertise have developed, delivered, and assessed examinations over a lengthy period of time, this conclusion is understandable.

The findings back up Alkharusi's (2008) findings in Oman, where he looked at the influence of classroom evaluation techniques on students' attainment objectives. The findings of hierarchical linear modeling approaches revealed that class setting elements, instructors' years of work experience, and assessment procedures interacted considerably with students' characteristics in determining students' accomplishment objectives. Alkharusi also demonstrated that teachers' teaching experience influenced their ability to develop tests.

Acar-Erdol and Yldzl (2018) wanted to know how elementary, secondary, and high school teachers judge their students in the classroom. The researchers discovered that a teacher's level of expertise influenced their ability to develop tests. In a similar vein, Adodo (2014) discovered a link between instructors' teaching experience and their ability to evaluate students' cognitive achievement. As a result, the amount of teaching expertise had an impact on testing techniques.

### **Differences in Testing Practices Based on Professional Qualification**

Furthermore, the study discovered that instructors' assessment techniques varied significantly depending on their qualifications. The null hypothesis, that there was no statistically significant difference in basic school teachers' testing techniques based on their professional qualifications, was shown to be false. Because a significant difference was discovered, a post-hoc analysis was required to establish which of the three mean values was responsible for the significant difference in the ANOVA findings. The post hoc analysis showed that teachers with University Degree (B.Ed.) and Master's Degree had better testing practices than those with the other qualifications. The teachers with HND and University Degree (BA, B.Sc.) had lower mean scores indicating poorer testing practices.

Having a Bachelor of Education degree requires that you take several courses in testing and educational measurement and so it does not come as a surprise that they had better testing practices compared to the other qualifications. The findings support the findings of Dubem cited in Inko-Tariah and Okon (2019) that lecturers competence in test construction is dependent on their level of education and the skills they through training.

Similarly, Acar-Erdol and Yıldızlı (2018) revealed that level of education and training influence level of proficiency in test construction. As people progress in terms of their educational level, there is a high likelihood that they would be take more courses in testing and measurement which will prepare them more for test construction, administration and scoring.

### **Differences in Testing Practices Based on Knowledge of Test Construction**

Finally, the study discovered that instructors' testing techniques differed significantly depending on their expertise of testing. Instructors who had completed a testing course had better testing methods than teachers who had not taken a counselling course. The null hypothesis that there is no significant variation in teacher testing techniques based on expertise in testing was rejected based on the findings.

Oduro-Okyireh (2008) investigated whether pre-service testing training improves senior secondary school teachers' competency in real testing practice in Ghana's Ashanti Region. Pre-service learning in educational measurement was found to have a favorable impact on real testing practice, however the effect was minor. In essence, instructors' test construction abilities were influenced by their test construction expertise.

Similarly, Yeboah (2018) investigated the assessment course's abilities and skills among practicing teachers in Ghana's Agona West and East Municipalities. The study's findings revealed that, to a large extent, instructors stated that they gained competences and abilities as a result of taking the course, which aided them in planning a test while keeping the exam's aim in



mind. The training, according to teachers, has assisted them in avoiding ambiguity in exam items.

In addition, Zhang and Burry-Stock (2003) examined the assessment strategies of teachers and the extent to which years of experience and training impacted on assessment practices. It was shown that irrespective of the years of experience, teachers who had some level of training in measurement, reported stronger assessment abilities in testing and giving feedback on test results.

From the forgoing discussion, it can be stated that knowledge in testing or measurement can significantly improve the testing practices of teachers. This makes testing a course that should be mandatory for all prospective teachers.

### **Chapter Summary**

The study's findings and debate were discussed in this chapter. This study addressed three research topics and four hypotheses. Teachers, on the whole, adhered to normal norms when it came to the creation, administration, and scoring of examinations. Teachers' testing procedures did not change based on gender, but they did differ based on teaching experience, educational qualifications, and testing knowledge.

## CHAPTER FIVE

### SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

#### Introduction

The preceding chapter covered the study's results and discussion. Summary, major findings, conclusions and recommendations are dealt with in this chapter. The chapter also covered suggestions for further research.

#### Summary of Study

The aim of this research was to investigate basic school teachers' testing practices within the Greater Accra Region of Ghana. Specifically, the study answered three research questions and tested four hypotheses:

#### Research questions:

1. What testing practices do basic school teachers follow when constructing achievement test items?
2. What testing practices do basic school teachers follow when administering achievement test items?
3. What testing practices do basic school teachers follow when scoring essay-type test items?

#### Hypotheses:

1.  $H_0$ : There is no statistically significant difference in basic school teachers' testing practices on the basis of their gender.  
 $H_1$ : There is statistically significant difference in basic school teachers' testing practices on the basis of their gender.
2.  $H_0$ : There is no statistically significant difference in basic school teachers' testing practices on the basis of their years of teaching experience.

$H_1$ : There is statistically significant difference in basic school teachers' testing practices on the basis of their years of teaching experience.

3.  $H_0$ : There is no statistically significant difference in basic school teachers' testing practices on the basis of their professional qualification.

$H_1$ : There is statistically significant difference in basic school teachers' testing practices on the basis of their professional qualification.

4.  $H_0$ : There is no statistically significant difference in basic school teachers' testing practices on the basis of their knowledge on test construction.

$H_1$ : There is statistically significant difference in basic school teachers' testing practices on their knowledge on test construction.

Literature were reviewed in this study. The review focused on the theoretical framework, conceptual review and the empirical review. The theories reviewed included the Classical Test Theory and the Item Response Theory. Concepts relating to the study were also reviewed. Previous empirical literature were also reviewed. The review showed that generally, some teachers had the skills and adhered to standard practices in test construction, administration and scoring while some other teachers did not.

Mixed methods approach using the descriptive survey design was utilised in the study. In all, 382 teachers were involved in the study selected via stratified random sampling from a population of junior high school teachers in selected districts in the Greater Accra Region of Ghana. Collection of data was done using questionnaire and interview guide. Quantitative data were analysed using both descriptive and inferential statistics. Qualitative data

on the other hand were analysed using thematic analysis. Ethical issues were adhered to.

### Major Findings

The study discovered that teachers establish the aim of a test before creating test items while creating achievement tests. Teachers often link the test to the subject matter's teaching objectives. Because the objectives are linked to the exam, the teachers also chose a test style that would be appropriate for testing the stated objectives. Aside from that, respondents said they create clear and simple guidelines for the full exam and components of the test when creating accomplishment tests.

Moreover, the respondents noted that they evaluate test to ascertain its clarity, practicality, efficiency and validity before constructing. The teachers also created a scoring guide at the moment the test items were written and went through the test items after they had been placed aside for some time.

Majority of the respondents usually did distracter analysis of multiple-choice test items and also usually estimated the difficulty level of objective test items. In terms of computing discrimination index of objective type tests, most of the respondents were not familiar with discrimination index or did not compute discrimination index.

Teachers examined the learning goals to be tested, the subject matter to be measured, and the difficulty level of the items while arranging objective tests (arranged based on their level of challenge). Finally, while allocating time for essay examinations, respondents reported that they took into account the quantity of items on the test as well as the time it takes roughly 90% of students to finish the test.

The study discovered that when teachers prepared students for tests in advance, they frequently informed them of the test's date and time. In addition, the respondents stated that they often made pupils aware of the laws and regulations controlling the test's conduct and format (objective type or essay-type tests). The significance of the exam results was frequently communicated to pupils ahead of time.

Respondents said they typically stand where they can observe all students during examinations and move among them once in a while to look for malpractices. They often mentioned the remaining time (time left to complete the test) at regular intervals during the exam administration. The responders often checked that the seating arrangement offers enough distance for students not to copy from one another, as well as that the exam area has appropriate air and illumination. In addition, responders frequently kept their comments in the testing room to a minimal and ensured that they were connected to the exam. These were the most common test administration practices among the responders.

The majority of respondents employed the analytic technique to score essay assessments, according to the survey. The analytic technique employs a grading system in which points are assigned to concepts and points that are clearly presented. This shows that most of the respondents do not utilize personal impressions when grading essay examinations. In addition, the majority of respondents gave each student a score for all of the items they answered before moving on to the next student. As a result, they utilized a "student by student" scoring method.

In terms of general scoring techniques for essay examinations, the majority of respondents stated that they always followed the marking system when scoring essay tests. The majority of respondents also stated that they do not rearrange the answer scripts at random after the items have been scored, and that they do not score all replies to a single question in one session. The majority of the professors in the survey also stated that they only graded essay examinations when they were physically and intellectually fit, and that they offered students with comments on their response scripts to help them learn. Criterion-referenced interpretation was preferred above norm-referenced interpretation for interpreting test outcomes.

Significant differences were not observed in the testing practices of male and female teachers, according to the research. Although there is a minor change in the mean scores, it is not statistically significant. There was, however, a significant variation in teacher testing practices based on number of years of experience. Specifically, teachers with more years of experience have superior testing practices.

Moreover, the study showed that there is a significant difference in the testing practices of teachers on the basis of their qualification. The post hoc analysis showed that teachers with University Degree (B.Ed.) and Master's Degree had better testing practices than those with the other qualifications. The teachers with HND and University Degree (BA, B.Sc.) had lower mean scores indicating poorer testing practices.

Finally, the study found that there was a significant difference in the testing practices of teachers on the basis of their knowledge in testing. The

teachers who had taken a course in testing had better testing practices than those who had not taken a course in testing,

### **Conclusions**

Several conclusions are drawn on the basis of the findings of the study.

Firstly, it was concluded that basic school teachers in the Greater Accra region followed standard practices in the construction of tests. Specifically, the teachers ensured that test items are constructed according to a purpose, focused on the objectives of the content, have clarity in terms of directions and are valid. Teachers also prepared marking scheme ahead of time and engaged in distractor analysis along with considering item difficulties in test construction.

Secondly, it was concluded that basic school teachers in the Greater Accra region followed standard practices in test administration both before and during the test. Before the day of the test, the teachers outlined clearly everything about the test to learners. During the administration of the test, teachers ensured that the arrangement and order in the examination room was conducive and did not favour malpractices.

Thirdly, it was concluded that basic school teachers in the Greater Accra region followed the standard practices in the scoring of essay tests. They consistently followed a marking scheme and marked only when they were in a good mental and physical state and in an environment without interruptions. Analytic method of scoring and criterion-referenced interpretation method were favoured more by teachers.

Further, it was concluded that gender was not an issue of consideration in terms of testing practices and as such both male and female basic school

teachers in the Greater Accra region had the same or similar testing practices. However, teaching experience and educational qualification were influential in testing practices since teachers who had more years of teaching and background in Education had better testing practices. Finally, basic school teachers in the Greater Accra region who had knowledge in testing or educational measurement had better testing practices compared to those without knowledge in testing or measurement. This means that testing practices can be dependent on knowledge in testing or measurement.

### **Recommendations**

Some recommendations are made on the basis of the findings:

1. It is recommended that basic school teachers in the Greater Accra region continue to construct test items on the basis of objectives and directions of the test and improve on areas that they were lacking like the item discrimination.
2. It is recommended that basic school teachers in the Greater Accra region continue ensuring that there is conducive environment before and during administration of tests since the study found that they were following the standard practices.
3. It is recommended also that basic school teachers in the Greater Accra region score and interpret essay tests by consistently ensuring that there is fairness as was indicated in the study that they were following.
4. Further, it is recommended that Ghana Education Service and other examination bodies give opportunity to both male and female teachers to be involved in any test construction, administration and scoring



tasks. This is because no gender difference was found in terms of testing practices of teachers.

5. It is recommended that basic school teachers in the Greater Accra region who may have difficulties with test construction, administration and scoring should continue engaging in testing practices since by means of experience they can improve.
6. Basic school teachers in the Greater Accra region should continue upgrade their educational level in order to improve their testing practices. This is because the study found that teachers with higher educational qualifications had better testing practices.
7. Finally, teacher training colleges and institutions make testing and measurement courses compulsory while at the same time regional education offices organise intermittent testing workshops for teachers. This can help enhance the knowledge of teachers in testing and as such improve their testing practices.

### **Suggestions for Further Research**

The suggestions below are given for further research:

1. Researchers can examine the actual test items constructed by teachers to find out if they are actually following the standard practices. This can help obtain data different from this current study since this study was mainly self-reported by the teachers.
2. Since most literature investigating testing practices have been focused on teachers, further study can be conducted into the views of students concerning testing practices of their teachers. This can help gain

another perspective on testing practices to help make recommendations for teachers to better their teaching practices.



## REFERENCES

- Acar-Erdol, T., & Yıldızlı, H. (2018). Classroom assessment practices of teachers in Turkey. *International Journal of Instruction*, 11(3), 587-602.
- Adodo, S. O. (2014). Evaluation of Secondary School Teachers' Competency in Evaluating Students' Cognitive and Psychomotor Achievement in Basic Science and Technology. *Journal of Emerging Trends in Educational Research and Policy Studies (JETERAPS)*, 5(3), 48-53.
- Agu, N. N., Onyekuba, C., & Anyichie, A. C. (2013). Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. *Educational Research and Reviews*, 8(8), 431.
- Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). McGraw-Hill.
- Akbari, A. (2020). Measuring the degree of difficulty and translation competence: The two intertwining variables in modern translation multiple-choice items testing. *Journal of Applied Research in Higher Education*, 12(3), 475-494.
- Alberta Assessment Consortium. (2005). *Refocus: Looking at assessment for learning* (2nd ed.). Alberta Assessment Consortium.
- Albright, S., Winston, W., & Zappe, C. (2010). *Data analysis and decision making* (4<sup>th</sup> ed.). Cengage Learning.
- Ali, A. A. (1999). *Basic research skills in education*. Orient Printing and Publishing.

- Alkharusi, H. (2008). Effects of classroom assessment practices on students' achievement goals. *Educational Assessment, 13*, 243–266.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole Publishing Company.
- Amedahe, F. K., & Gyimah, K. A. (2003). Measurement and evaluation.
- Amedahe, F. K. (2010). *Notes on educational research methods*. University of Cape Coast.
- Amedahe, F. K. (Ed.). (2008). *Teacher education conference proceedings*. Adwinsa Publications (Gh.) Ltd.
- American Federation of Teachers. (AFT). National Council on Measurement in Education. (NCME) & National Education Association. (NEA). (2014). *Standards for teacher competence in education assessment of students*. National Council on Measurement in Education.
- Anane, E. (2011). Predictive validity of continuous assessment scores of students' performance in end-of-semester examinations in colleges of education in Ghana. *Journal of Business and Educational Policies, 7*(1), 95-104.
- Anane, E., & Adu-Mensah, J. (2019). Teachers scoring and grading of students' responses to tasks: The Ghanaian basic school experience. *Academic Journal of Interdisciplinary Studies, 8*(1), 87-94.
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers*. Jossey-Bass.
- Anhwere, Y. M. (2009). *Assessment practices of teacher training college tutors in Ghana*. [Unpublished master's thesis, University of Cape Coast].

- Armah, C. (2018). *Test construction and administration practices among lecturers and staff of examinations unit of the University of Cape Coast*. [Unpublished master's thesis, University of Cape Coast].
- Asamoah, D., Sundame, B., Adusei, A., Ocansey, M. K., & Ntim, W. (2019). Ensuring content representativeness: Perceptions of senior high school teachers in Ghana. *International Journal of Education and Learning*, 6(1), 11-16.
- Asamoah-Gyimah, A. (2002). *An evaluation of the practice of continuous assessment in the senior secondary schools in the Ashanti Region of Ghana*. [Unpublished master's thesis, University of Cape Coast].
- Babbie, E. (1990). *Survey research methods* (2nd ed.). Wadsworth.
- Benson, J. (1998). Developing a strong programme of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, 17, 10- 17.
- Benzehaf, B. (2017). Exploring teachers' assessment practices and skills. *International Journal of Assessment Tools in Education*, 4(1), 1-18.
- Best, J. W., & Kahn, J. V. (2006). *Research in education* (10th ed.). Pearson Education Inc.
- Bichi, A. A. (2016). Classical test theory: An introduction to linear modeling approach to test and item analysis. *International Journal for Social Studies*, 2(9), 27-33.
- Black, P., & William, D. (2001). *Raising standards through classroom assessment*. King's College London School of Education.

- Borich, G. D. (2011). Interactions among group regressions: Testing homogeneity of group regressions and plotting regions of significance. *Educational and Psychological Measurement, 31*, 251-253.
- Burns, M. (2005). Looking at how students reason. *Educational Leadership, 63*(3), 26–31.
- Caldwell, D. J., & Pate, A. N. (2013). Effects of question formats on student and item performance. *American Journal of Pharmaceutical Education, 77*(4), 71-77.
- Campbell, C., & Evans, J. A. (2000). Investigation of pre-service teachers' classroom assessment practices during students teaching. *Journal of Educational Research, 93*(6), 350-355.
- Capper, J. (2007). *Testing to learn-Learning to test*. Delaware Publishing Company.
- Chan, J. (2009). Long-term effects of testing on the recall of nontested materials. *Memory, 18*(1), 49-57.
- Charmey, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English, 18*, 65-81.
- Childs, R. A. (1989). *Constructing classroom achievement tests*. Retrieved from <https://files.eric.ed.gov/fulltext/ED315426.pdf>
- Clarke, S. (2006). *Formative assessment in the secondary classroom*. Holder Murray.
- Cohen, L., Manion, L., & Morrison, K. (2003). *Research methods in education*. Routledge Falmer.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage Publications, Inc.

Creswell, J. W., & Plano-Clark, V. L. (2007). *Designing and conducting mixed methods research*. SAGE Publications.

Crocker, L., & Algina, J. (2008). *Introduction to classical & modern test theory*. Cengage Learning.

Crocker, M. (2019). *The importance of evaluation and testing in an educational system*. <https://www.teflcourse.net/blog/the-importance-of-evaluation-and-testing-in-an-educational-system-ittt-tefl-blog/>

Crooks, T. (2001). *The validity of formative assessment*. Paper presented at the British Educational Research Association annual Conference, University of Leeds, 13-15 September 2001.

D'Agostino, J. V. (2007). *Quantitative research, evaluation, and management section*. Ohio State University.

Darling-Hammond, L. (2012). *Powerful teacher education: Lessons from exemplary programmes*. John Wiley & Sons.

Davies, A. (2004). *Finding proof of learning in a one-to-one computing classroom*. Connections Publishing.

Deville, C. (1991). Testing in foreign language programs and testing programs in foreign language departments: Reflections and recommendations.

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *The Canadian Journal for the Scholarship of Teaching and Learning*, 2(2), 1-23.

Din, M. (2020). Evaluating university students' critical thinking ability as reflected in their critical reading skill: A study at bachelor level in Pakistan. *Thinking Skills and Creativity*, 35, 100627.

Education Up Close (2005). *Writing effective tests: A guide for teachers*.

<http://www.glencoe.com/sec/>

[teachingtoday/educationup.phtml/print/40](http://www.glencoe.com/sec/teachingtoday/educationup.phtml/print/40).

Ekanayake, A. (2011). Validity and reliability in case study research in

accounting: a review and experience.

Esomonu, N. P. (2002). Assessment of non-cognitive behaviours in secondary schools for national development. In R. C. Ebenebe, & L. R. Akudolu (Eds.). *Education for national development and integration* (pp.88-93). John Best Publishing.

Etsey, Y. K. A. (1992). What is continuous assessment? *The Oguaa Educator*, 10(1), 82-92.

Etsey, Y. K. A. (2006). *Notes on educational measurement and evaluation*. University of Cape Coast.

Etsey, Y. K., & Afedo, E. (2008). The Effects of Peer Observation and Intervention on Teacher Effectiveness in Ho Polytechnic. *Journal of Educational Development and Practice*, 2, 115-131.

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

Fraenkel, J. R., & Wallen, N. E. (2000). *How to design and evaluate research in education*. McGraw-Hill Higher Education.

Frey, B. B. (2007). *An introduction to quality test construction*.  
<http://www.specis1connection.ku.edu/cgi-bin/cgiwrap/speccom/mainphp?cat=assessment&section=main&section=qualitytest/main>.



- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Pearson
- Gay, R. L. (2011). *Educational research: Competencies for analysis and application* (4th ed.). Macmillan Publishing Company.
- Green, S. K., Johnson, R. L., Kim, D.-H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education, 23*, 999–1011.
- Gregory, R. J. (1992). *Psychological testing: history, principles and applications*. Allyn and Bacon.
- Gronlund, N. E. (1988). *How to construct achievement tests* (4th ed.). Prentice-Hall, Inc.
- Gronlund, N. E. (2003). *Assessment of student achievement* (7th ed.). Allyn and Bacon.
- Guskey, T. R. (2006). Making high school grades meaningful. *Phi Delta Kappan, 87*(9), 670-675.
- Hamafyelto, R. S., Hamman-Tukur, A., & Hamafyelto, S. S. (2015). Assessing teacher competence in test construction and content validity of teacher made examination questions in commerce in Borno State, Nigeria. *Journal of Education, 5*(5), 123-128.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10*(3), 229–244.
- Hambleton, R. K., & Jodoin, M. (2003). Item response theory: Models and features. In R. Fernández-Ballesteros, (Ed.), *Encyclopedia of psychological assessment* (pp. 509-514). Sage.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development.

*Educational Measurement: Issues and Practice*, 12(3), 38-47.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory: Measurement methods for the social sciences series (Vol. 2)*. Sage.

Harlen, M. (2006). *Closing the loop-learning from assessment*. Paper presented at the University of Maryland Eastern Shore Assessment Workshop, January, 2006.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.

Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, 33(2), 151-163.

Hudaya, D. W. (2017). *Teachers' assessment literacy in applying principles of language assessment*. Paper presented at the 1st Education and Language International Conference Proceedings Center for International Language Development of Unissula, May 2017.

Hussain, L., Jamil, A., Siraji, J., & Maroof, K. (2012). Development and standardization of intelligence test for children. *International Journal of Learning & Development*, 2(5).

Hussain, S., & Sajid, S. (2015). Test construction and evaluation: A brief review. *Indian Journal of Applied Research*, 5(6), 725-729.

Inko-Tariah, D. C., & Okon, E. J. (2019). Knowledge of test construction procedures among lecturers in Ignatius Ajuru University of Education, Port Harcourt, Nigeria. *Academic Research International*, 10(1), 130-138.

Irvine, J. (2019). Relationship between Teaching Experience and Teacher Effectiveness: Implications for Policy Decisions. *Journal of Instructional Pedagogies*, 22.

Kaplan, R. M., & Saccuzo, D. P. (1997). *Psychological testing: Principles, applications and issues*. Brooks Cole Pub. Company.

Kazuko, J. W. (2010). *Japanese high school mathematics teachers' competence in real world problem solving*. Keto Academy of New York and Teachers College Columbia University.

Khan, A. (2019). *Meaning, nature and functions of assessment*. <https://www.toppr.com/bytes/meaning-nature-and-functions-of-assessment/>

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

Koksal, D. (2004). Assessing teachers' testing skills in Elt and enhancing their professional development through distance learning on the net. *Turkish Online Journal of Distance Education*, 5(1), 122-131.

Koloi-Keaikitse, S. (2017). Assessment of teacher perceived skill in classroom assessment practices using IRT Models. *Cogent Education*, 4(1), 128-135.

Kondratek, B. (2023). UIRT: Stata module to fit unidimensional Item

Response Theory models.

Kubiszyn, T., & Borich, G. (2013). *Educational testing and measurement: Classroom application and practice* (10th ed.). John Wiley & Sons, Inc.

Kumar, R. (1999). *Research methodology: A step-by-step guide for beginners*. Sage Publications.

Kumar, R. (2005). *Research methodology: A step by step guide for beginners* (2nd ed.). Sage Publication.

Le Grange, L. L., & Reddy, C. (1998). *Continuous assessment: An introduction and guidelines to implementation*. Juta.

Leedy, D. L., & Ormrod, J. E. (2005). *Practical research: Planning and design* (8th ed.). Merrill/Prentice Hall.

Leighton, J. P., & Gokiert, R. J. (2005). *The cognitive effect of test item features*.

<http://www.education.ualberta.ca/educ/psych/crame/files/NCMEfinalLeighandGokiert.pdf>.

Liaquat, H, Asif, J. M., Siraji, J., & Maroof, K. (2012). Development and standardization of intelligence test for children. *International Journal of Learning & Development*, 2(5), 190-202.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Pearson.

Lord, F. (1952). *A theory of test scores* (Psychometric Monograph No. 7). Psychometric Corporation.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

- Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment, 1*(1), 1-11.
- Marcoulides, G. A. (1999). Generalizability theory: Picking up where the Rasch IRT model leaves off? In S. E. Embretson, & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 129-152). Erlbaum.
- Marzano, R. J. (2000). *Transforming classroom grading*. Association for Supervision and Curriculum Development.
- McNamee, G. D., & Chen, J.-Q. (2005). Dissolving the line between assessment and teaching. *Educational Leadership, 63*(3), 72–77.
- Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education, 43*(6), 1-18.
- Moawad, R. A. (2020). Online learning during the COVID-19 pandemic and academic stress in university students. *Revista Românească pentru Educație Multidimensională, 12*(1 Sup2), 100-107.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62*, 229–258.
- Muzenda, D. (2017). Effectiveness of teacher-made tests in goromonzi district primary schools. *IJRDO-Journal of Educational Research, 2*(2), 53-65.
- Nitko, A. J. (1996). *Educational assessment of students* (3rd ed.). Prentice-Hall, Inc.

Nitko, A. J. (2001). *Educational tests and measurements* (3th ed.). Prentice Hall, Inc.

Nitko, A. J. (2004). *Educational assessment of students* (4th ed.). Pearson Education, Inc.

Nitko, A. J., & Brookhart, S. M. (2007). *Education assessment of students*. New Jearsey: Merrill Prentice Hall.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1–18.

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

Ogah, J. K. (2013). *Decision making in research process: Companion to students and beginning researchers*. Adwinsa Publications (Gh.) Ltd.

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

Ololube, N. P. (2008). Evaluation competencies of professional and non-professional teachers in Nigeria. *Studies in Educational Evaluation*, 34(1), 44-51.

Onyechere, I. (2000). *New face of examination malpractice among Nigerian youths*. The Guardian Newspaper July 16.

Osman, S. (2021). BASIC SCHOOL TEACHERS' ASSESSMENT

PRACTICES IN THE SISSALA EAST MUNICIPALITY,  
GHANA. *European Journal of Education Studies*, 8(7).

Osterlind, N. L. (2006). Sampling designs in qualitative research: Making the sampling process more public. *The Qualitative Report*, 12(2), 238-254.

Osuala, E. C. (2005). *Introduction to research methodology* (3rd ed.). Africana First Publishers Ltd.

Paulson, T. U. (2003). *The study of teaching*. Holt, Rinehart and Winston Inc.

Petters, J. S., & Okon, M. O. (2014). Students' perception of causes and effects of examination malpractice in the Nigerian Educational System: The way forward for quality education. *Procedia - Social and Behavioural Sciences*, 114(21), 125-129.

Pidu, A. W., Darsikin, E., & Ika, I. (2018). The competency of junior high school physics teachers in constructing achievement test and its implication for the test quality in Sindue. *Advances in Social Science, Education and Humanities Research*, 174, 49-52.

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

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1), 130-133.

Quansah, F., & Amoako, I. (2018). Attitude of Senior High School (SHS) teachers towards test construction: Developing and validating a standardised instrument. *Research on Humanities and Social Sciences*, 8(1), 25-30.

Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in senior high schools in Ghana: Document analysis. *International Journal of Assessment Tools in Education*, 6(1), 1–8.

Rhalmi, M. (2020). *Definition of testing, assessment, and evaluation*.

<https://www.myenglishpages.com/blog/definition-of-testing-assessment-and-evaluation/#:~:text=Simply%20put%2C%20a%20test%20refers,such%20as%20validity%20and%20reliability.>

Roediger, H. L. III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 1–36). Elsevier Academic Press.

Roediger III, H. L., & Karpicke, J. D. (2018). Reflections on the resurgence of interest in the testing effect. *Perspectives on Psychological Science*, 13(2), 236-241.

Rubin, A., & Babbie, E. R. (2016). *Essential research methods for social work*. Cengage Learning.

Rukundo, L., & Magambo, W. (2010). *What teachers need to know about assessment*. National Education Association.

Runeson, P. (2006). A survey of unit testing practices. *IEEE software*, 23(4), 22-29.

Salihu, A. G. (2019). Assessing teachers ability on test construction and economics content validity in Nasarawa State Senior Secondary



Schools, Nigeria. *International Journal of Innovative Research in Education, Technology & Social Strategies*, 6(1), 1-16.

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment in special and remedial education*. Houghton Mifflin.

Sarita, K. (2005). *Improving assessment, evaluation and remedial*. Isha Books.

Saumure, K., & Given, L. M. (2008). *Convenience sample: The SAGE Encyclopedia of Qualitative Research Methods*. Sage Publications.

Sax, W., & Cromack, S. (2005). *Controversial issues in special education. Divergent perspectives* (2nd ed.). Ally and Bacon Press.

Schumacker, R. E. (2010). *Classical test analysis*.

[http://appliedmeasurementassociates.com/ama/assets/File/CLASSICAL\\_TEST\\_ANALYSIS.pdf](http://appliedmeasurementassociates.com/ama/assets/File/CLASSICAL_TEST_ANALYSIS.pdf).

Scriven, M. (1994). Evaluation as a discipline. *Studies in Educational Evaluation*, 20(1), 147-166.

Shibboleth, A. (2011). *Constructing and scoring essay tests*. Retrieved from <http://ed103shibboleth.blogspot.com/2011/04/constructing-and-scoring-essay-tests.html>

Shulman, L. S. (2004). *The wisdom of practice: Essays on teaching, learning, and learning to teach*. Jossey-Bass.

Silker, R. T. (2003). *Teachers and tests*. Basil Blackwell.

Silker, R. T. (2003). *Teachers and tests*. Basil Blackwell.

Singh, A. S., & Masuku, M. B. (2014). Sampling techniques & determination

of sample size in applied statistics research: An overview. *International Journal of economics, commerce and management*, 2(11), 1-22.

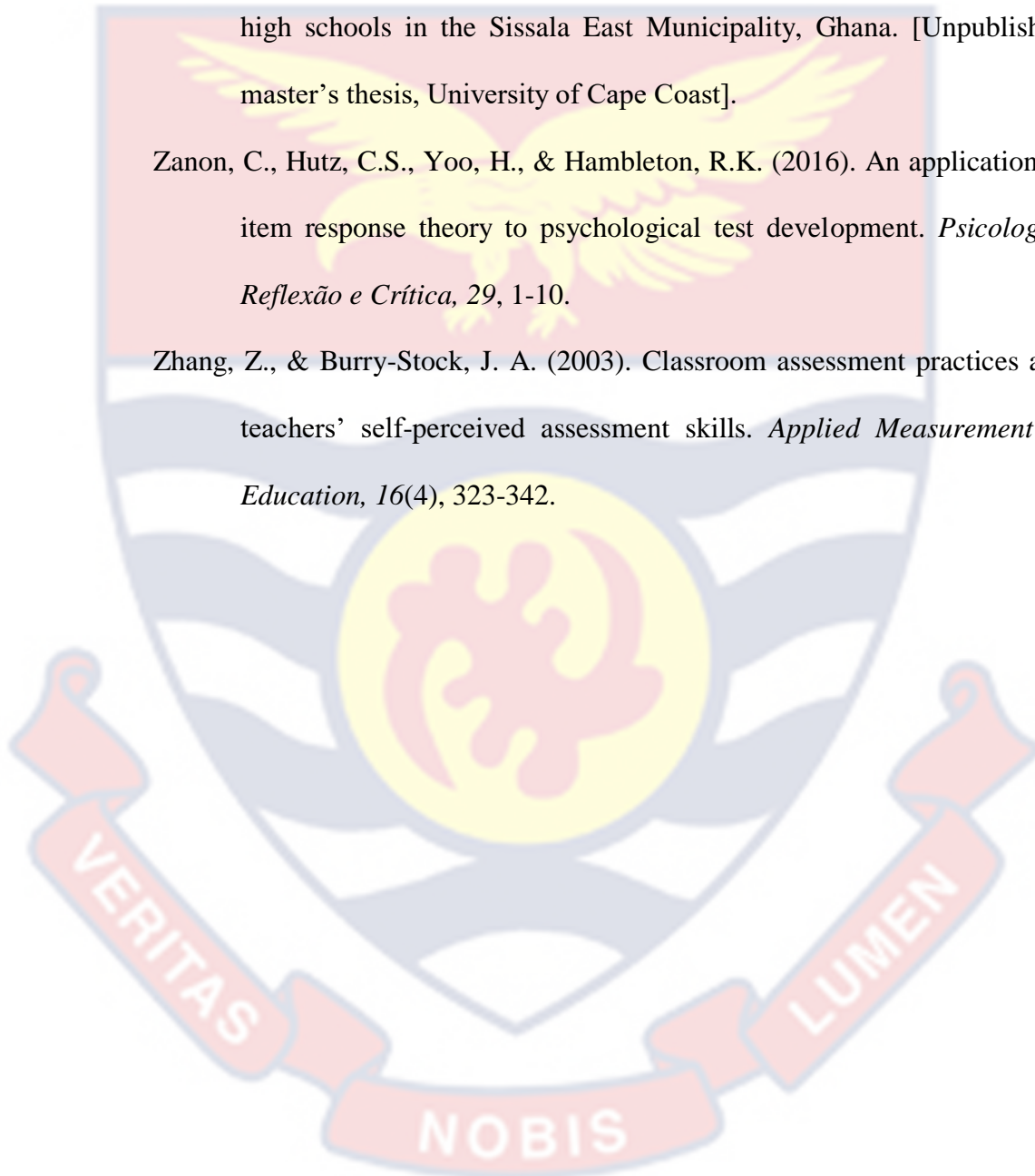
- Sparks, D. (1999). Assessment without victims: An interview with Rick Stiggins. *Journal of Staff Development*, 20(2), 1-4.
- Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–293
- Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right – using it well*. Assessment Training Institute.
- Tamakloe, E. K., Amedahe, F. K., & Atta, E. T. (2005). *Principles and methods of teaching*. Ghana Universities Press.
- Tamakloe, E. K., Atta, E. T., & Amedahe, F. K. (1996). *Principles and methods of teaching*. Black Mask Ltd.
- Traub, R. E., & Fisher, C. W. (1997). On the equivalence of constructed responses and multiple-choice tests. *Applied Psychological Measurement*, 1, 355-370.
- Ujah, E. U. (2001). *Development and validation of an introductory technology achievement test*. [Unpublished master's thesis, University of Nigeria, Nsukka].
- William, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade, & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). Routledge.
- Yeboah, A. (2017). *Adequacy and relevance of an assessment course: A follow-up study of graduate teachers in Ghana*. [Unpublished master's thesis, University of Cape Coast].

Yeboah, A. (2018). Competencies and skills in the assessment course: The perspective of the graduate teacher in Ghana. *The International Journal of Humanities and Social Studies*, 6(7), 221-227.

Zakariya, J. (2020). Teachers' testing practices of achievement test in junior high schools in the Sissala East Municipality, Ghana. [Unpublished master's thesis, University of Cape Coast].

Zanon, C., Hutz, C.S., Yoo, H., & Hambleton, R.K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica*, 29, 1-10.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342.





**APPENDICES**

**APPENDIX A****UNIVERSITY OF CAPE COAST****TEACHERS' QUESTIONNAIRE**

This questionnaire was designed to obtain data on how teachers construct, administer, score and interpret the results of their classroom achievement tests.

Such information would help to establish the testing practices of teachers in basic schools and ultimately promote learning.

It is hoped that you would provide honest answers to the items on the questionnaire. Please be assured that every information you share would be kept confidential and your anonymity is guaranteed.

**SECTION A****BACKGROUND INFORMATION**

**DIRECTION:** Please tick (✓) the option that describes your response(s) and write where applicable.

1. Gender: Female  Male

2. Highest educational qualification.

Teacher's Diploma

HND

University Degree (B.A., B.Sc.) B.Ed., M.Ed., M.A, M.Sc.)

University Degree (B. Ed)

Master's Degree (M.Ed., M.A., M.Sc., M.Phil.)

Other (Please specify).....

3. How many years (approximate) have you taught in the basic School?

Less than five (5) years.

Five (5) years or more.

4. Complete the following table to indicate the Form(s) and Subject(s) you teach.

Form(s)	Subject(s)
I.	
II.	
III.	

5. Have you ever taken a course in testing (i.e., educational measurement)?

No

Yes

### SECTION B

#### CONSTRUCTION OF CLASSROOM ACHIEVEMENT TESTS

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

In the construction of achievement test:	Always	Very Often	Sometimes	Never
6. I determine the purpose of a test before constructing test items				
7. I relate the instructional objectives of the subject matter to the test.				
8. I select the test format suitable for testing stated objectives.				
9. I use a table of specifications or test blueprint in determining the items on the test.				
10. I prepare more items than needed in the test or examination				
11. I write the test items in advance (at least two weeks) of the test date to permit reviews and editing.				

12. I prepare the marking scheme as soon as the test items are written.				
13. I review the test items after they have been set aside for a few days by reading over the items.				
14. I write clear and concise directions for the entire test and sections of the test.				
15. I copy test items directly from textbooks				
16. I evaluate the test to ascertain it's I. Clarity. II. Practicality. III. Fairness. IV. Validity. V. Efficiency.				

DIRECTIONS: Please tick (✓) the box that best describes your response(s), where applicable or write your answer in the space provided.

17. Do you usually do distracter analysis of your multiple-choice test items?

No

Yes

Not familiar with the term distracter analysis.

18. Do you usually estimate the difficulty level of your objective test items?

No

Yes

Not familiar with the term item difficulty.

19. Have you ever computed the discrimination index of your objective type tests?

No

Yes

Not familiar with the term discrimination index.

20. Which of the following do you consider in the arrangement of your objective test items?

- The type of item used.
- The difficulty level of the items (arranged in order of increasing difficulty).
- The learning outcomes being measured.
- The subject matter being measured.
- None of these.

21. How do you know that the time you allot to your essay tests is adequate? By using.....

- the number of items on the test to estimate the time.
- the time that you (the teacher) can take to complete the test.
- the time that about 90% of the students can take to complete the test.
- Other (Please specify).....

**SECTION C**

**ADMINISTRATION OF CLASSROOM ACHIEVEMENT TESTS**

DIRECTIONS: Please tick (√) the cell that indicates how frequently you engage your students in each of the following steps when preparing the students in advance for the test.

In preparing students in advance for the test, I make them aware of :	Always	Very Often	Sometimes	Never
22. When (date & time) the test will be given.				
23. The conditions (number of items, place of test) under which the test will be given.				
24. The content areas (study questions, list of topics or				



learning targets) that the test will cover.				
25. The test formats (objective type or essay-type tests).				
26. The emphasis or weighting of content areas (i.e. value in points or content areas with higher marks)				
27. How the test will be scored and graded.				
28. The rules and regulations governing the conduct of the test				
29. The importance of the results of the test.				

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

Test administration principles.	Always	Very Often	Sometimes	Never
30. I do not give tests immediately before or just after a long vacation, or other important events.				
31. I ensure that the sitting arrangement allows enough space so that students will not copy from each other.				
32. I ensure adequate ventilation and lighting in the testing room.				
33. I use “Do Not Disturb. Examinations In Progress” sign when students are taking tests and examinations.				
34. During examinations, I expect and cater for all possible emergencies.				
35. I announce the remaining time (time left to complete test) at regular intervals during test administration.				

36. During examinations, I stand where I can view all students and move among the students once a while to check malpractices.				
37. I keep all remarks in the testing room to a minimum and make sure they are related to the test.				
38. I tell students the dire consequences of failure in the test they are taking.				
39. I ask students to work faster during the time of testing in order to finish on time.				

**SECTION D**

**SCORING OF TEST AND INTERPRETATION OF THE TEST**

**RESULTS**

DIRECTIONS: Please tick (√) the box that best describes your response(s) where applicable or write your answer in the space provided.

40. Which method do you use in scoring your essay tests?

Reading the whole essay through and based on your impression about the quality of the essay you award the marks (Holistic Method).

Using a marking scheme in which the ideas and points clearly stated are awarded the marks (Analytic Method).

Other (Please specify).....

41. How do you score your essay tests?

Scoring all the items answered by each student before proceeding to the next student (i.e., student by student).

Scoring one item for all students before proceeding to the next item (i.e. item by item).

Other (Please specify).....

	NO	YES
42. In scoring essay tests, do you constantly follow the marking scheme as you score?		
43. In scoring essay tests, do you randomly reshuffle the answer scripts after scoring each item?		
44. In scoring essay tests, do you score all responses to a particular question at a sitting without interruption?		
45. Do you score essay tests only when you are physically sound and mentally alert?		
46. Do you keep previously scored items out of sight when scoring the rest of the items?		
47. In scoring essay tests, do you provide comments on the answer scripts for students to aid learning?		
48. Do you score the answer scripts of essay tests with the names of the students known to you?		
49. In scoring essay tests, do you score the mechanics of expressions such as penmanship, general neatness, spelling etc., separately from subject matter correctness?		

50. Which method(s) do you use in interpreting the results of your tests?

(You may indicate more than one)

[ ] Describing a student's level of performance in relation with that of other members of the class (Norm-referenced interpretation). E.g. A student places 12th out of 50 students in a class.

[ ] Describing a student's level of performance in terms of the learning tasks he can do. (Criterion-referenced interpretation). E.g. A student can solve 14 out of 20 problems in Algebra.

Other (Please specify).....

51. In norm-referenced interpretation of test scores, i.e. describing a student's level of performance in relation with that of other members of the class, which method(s) do you use? (You can choose more than one).

Simple ranking of raw scores

Percentile ranks

Stanine system of standard scores

Any other (Please specify).....

52. In criterion-referenced interpretation of test scores, i.e. describing a student's level of performance in terms of the learning tasks he can do, how do you set the performance standard (cut-off score)?

Through a standard of 50 percent correct score.

Through setting an arbitrary standard and adjusting it upward or downward depending on the circumstance.

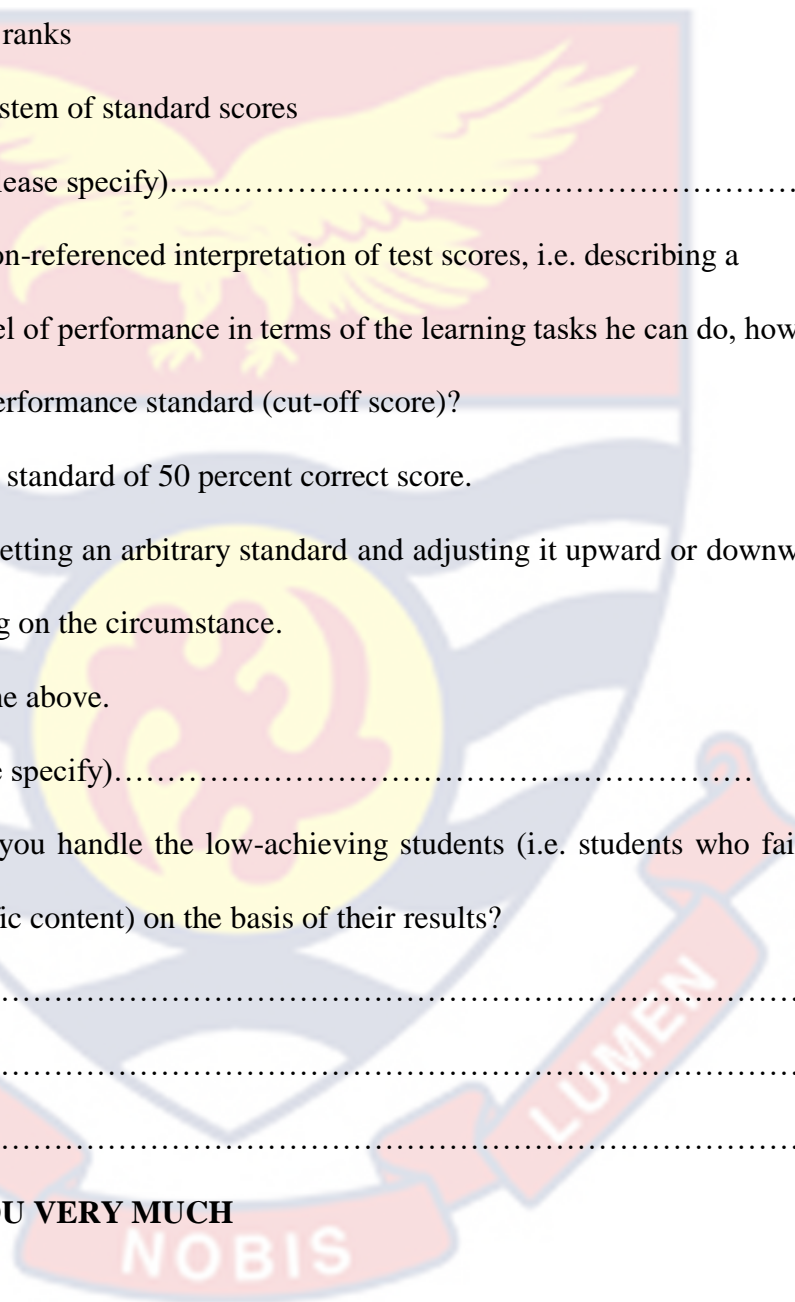
None of the above.

Other (Please specify).....

53. How do you handle the low-achieving students (i.e. students who fail to master specific content) on the basis of their results?

.....  
.....  
.....

**THANK YOU VERY MUCH**



**APPENDIX B****UNIVERSITY OF CAPE COAST  
DEPARTMENT OF EDUCATION AND PSYCHOLOGY  
INTERVIEW GUIDE FOR TEACHERS****Introduction**

I am grateful to you that you agreed to participate in this study.

**Rationale**

This interview would help in understanding basic school teachers' testing practices within the Greater Accra Region of Ghana. This would help to make recommendations for policy makers to aid teachers in basic schools in terms of their adherence to testing principles. The entire interview is likely to last for at most 20 minutes. I request to record this interview with your permission.

**Confidentiality**

You are assured that any information you provide during this interview will be kept privately and confidential. Your name will not be made to you even if you are quoted in the report.

**Consent**

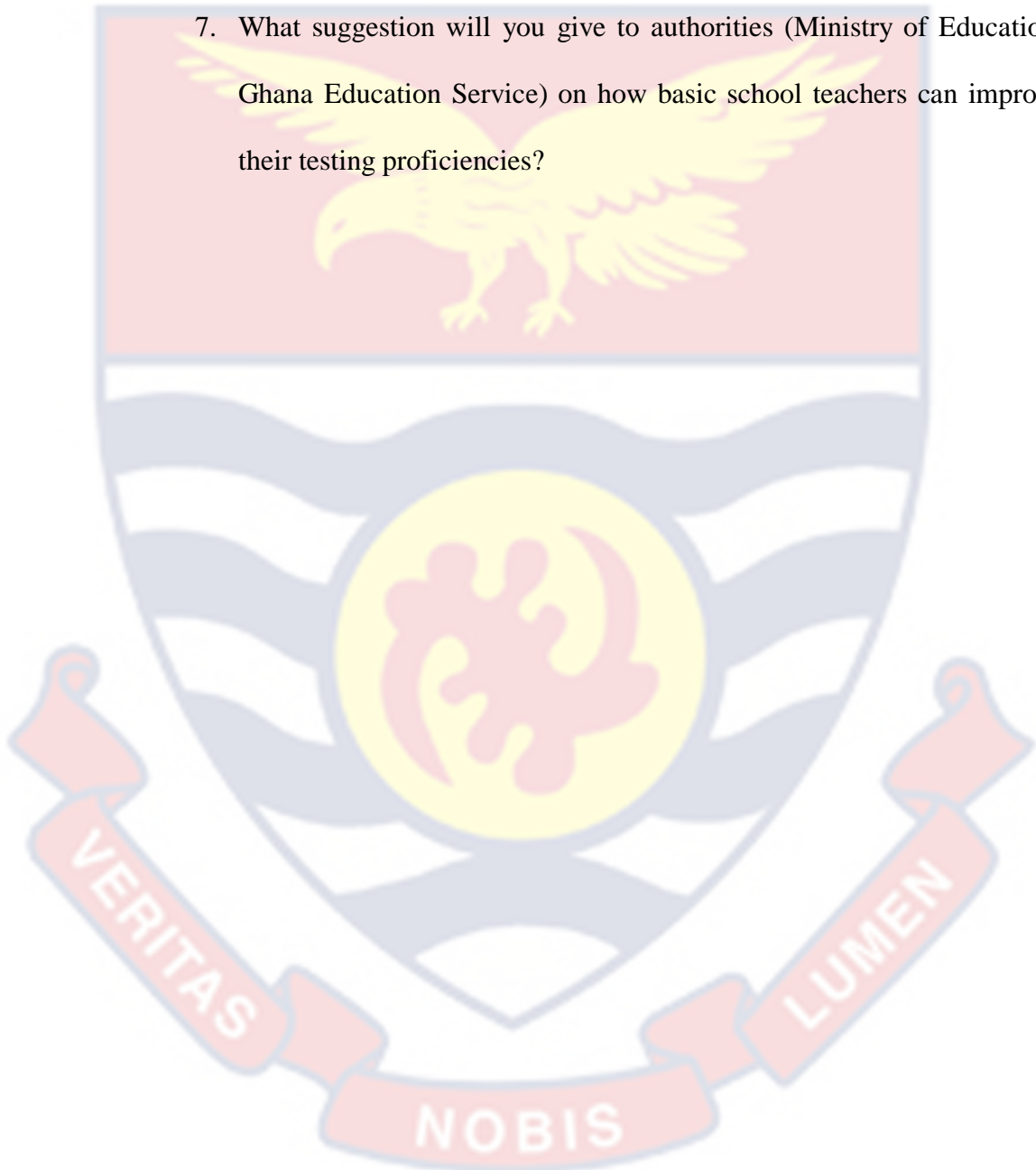
Please you have the right to be a part of the study and also have the choice to withdraw at any time.

**Ground rules**

Please be informed that no answer is wrong and so feel free to respond to the questions asked.

1. Please can you tell me about yourself (your age, the school you teach in, your highest qualification and the class level that you teach in)?
2. Generally, what do you know about test construction, administering and scoring?
3. What are some of the testing principles you follow when constructing achievement test items?
4. What are some of the testing principles you follow when administering achievement test items?

5. What are some of the testing principles you follow when scoring essay-type test items?
6. Do you think your education has prepared you to be able to construct, administer and score tests?
7. What suggestion will you give to authorities (Ministry of Education, Ghana Education Service) on how basic school teachers can improve their testing proficiencies?



## APPENDIX C

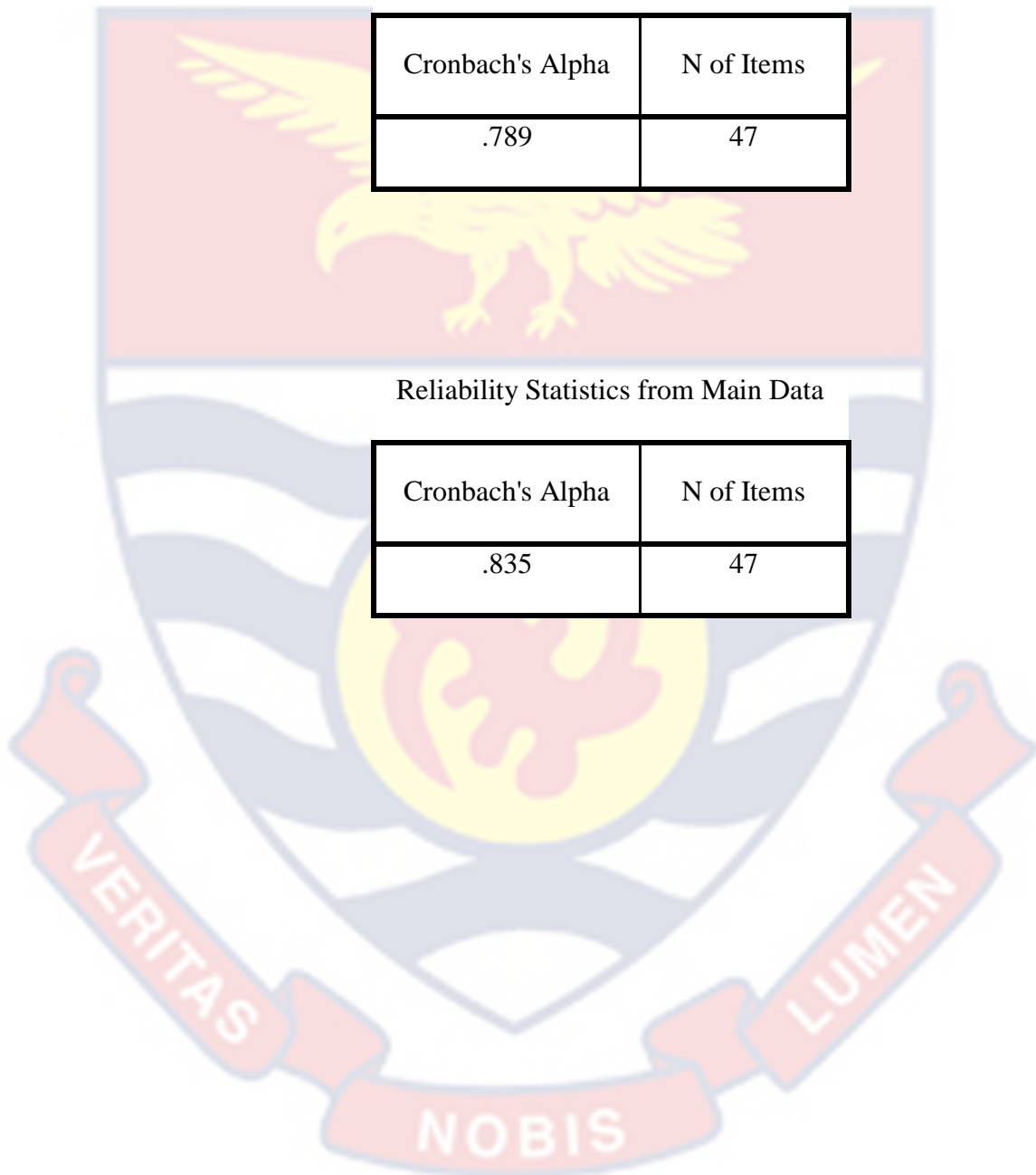
## RELIABILITY OUTPUT

## Reliability Statistics from Pilot Test

Cronbach's Alpha	N of Items
.789	47

## Reliability Statistics from Main Data

Cronbach's Alpha	N of Items
.835	47



APPENDIX D

INTRODUCTORY LETTER

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

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



UNIVERSITY POST OFFICE  
CAPE COAST, GHANA

Our Ref:

Your Ref:

18<sup>th</sup> June, 2020

TO WHOM IT MAY CONCERN

Dear Sir/Madam,

THESIS WORK  
LETTER OF INTRODUCTION  
MR. MOSES KABU KUBI OCANSEY

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

Mr. Ocansey is researching on the topic:

**“BASIC SCHOOL TEACHERS’ CLASSROOM TESTING PRACTICES IN THE GREATER ACCRA REGION.”**

We would be most grateful if you could provide him the opportunity and assistance to take data for the study. Any information provided would be treated strictly as confidential.

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

Thank you.

Yours faithfully,

Gloria Sagoe (Ms.)  
Chief Administrative Assistant  
For: HEAD



APPENDIX E

ETHICAL CLEARANCE

