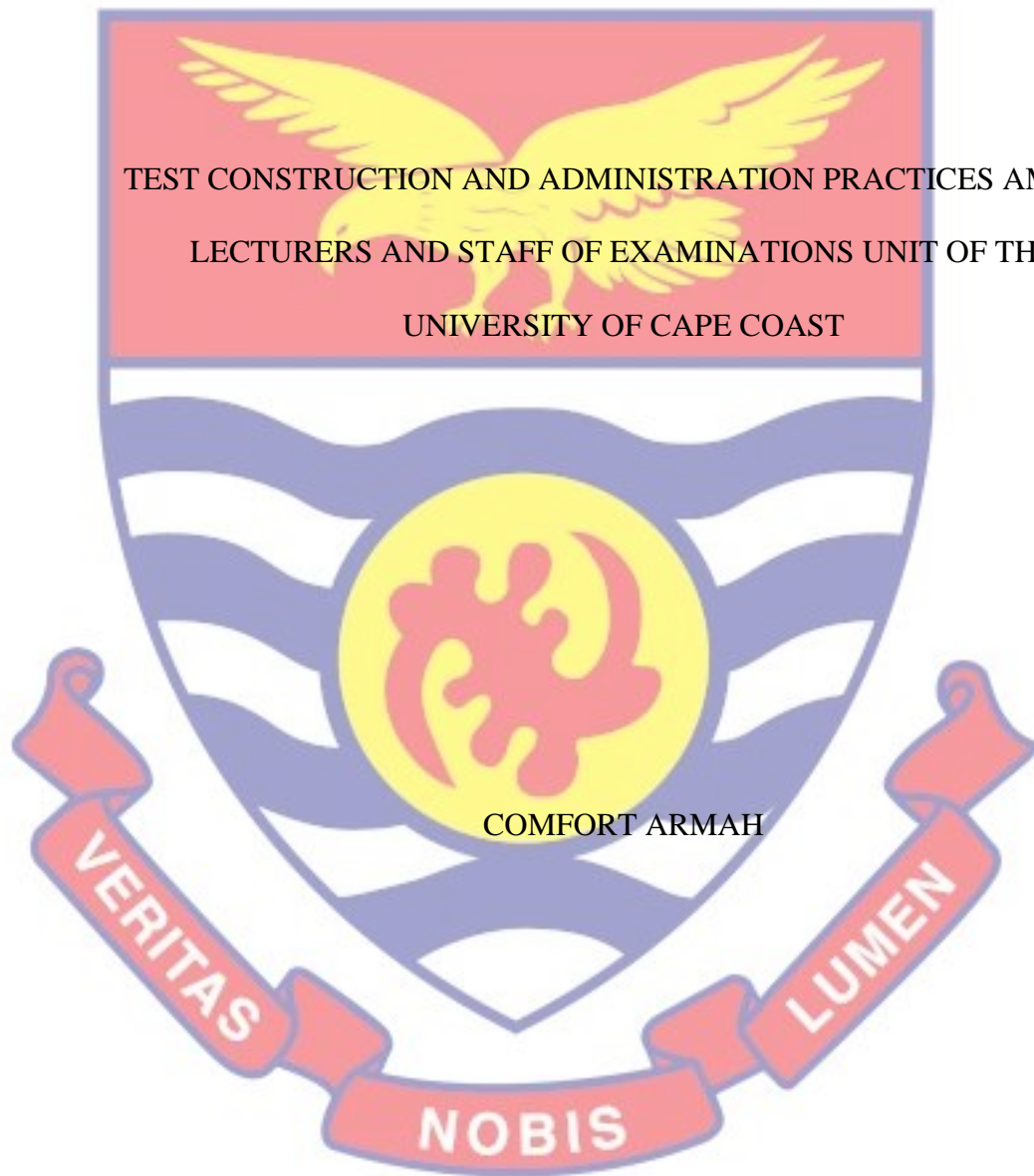


UNIVERSITY OF CAPE COAST



2018

UNIVERSITY OF CAPE COAST

TEST CONSTRUCTION AND ADMINISTRATION PRACTICES AMONG

LECTURERS AND STAFF OF EXAMINATIONS UNIT OF THE
UNIVERSITY OF CAPE COAST

BY

COMFORT ARMAH

This thesis submitted to the Department of Education and Psychology of the
Faculty of Educational Foundations, of College of Education Studies,
University of Cape Coast, in Partial Fulfilment of the Requirements for the
Award of Master of Philosophy Degree in Measurement and Evaluation

MARCH 2018

DECLARATION

Candidate's Declaration

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

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

Name:

Supervisors' Declaration

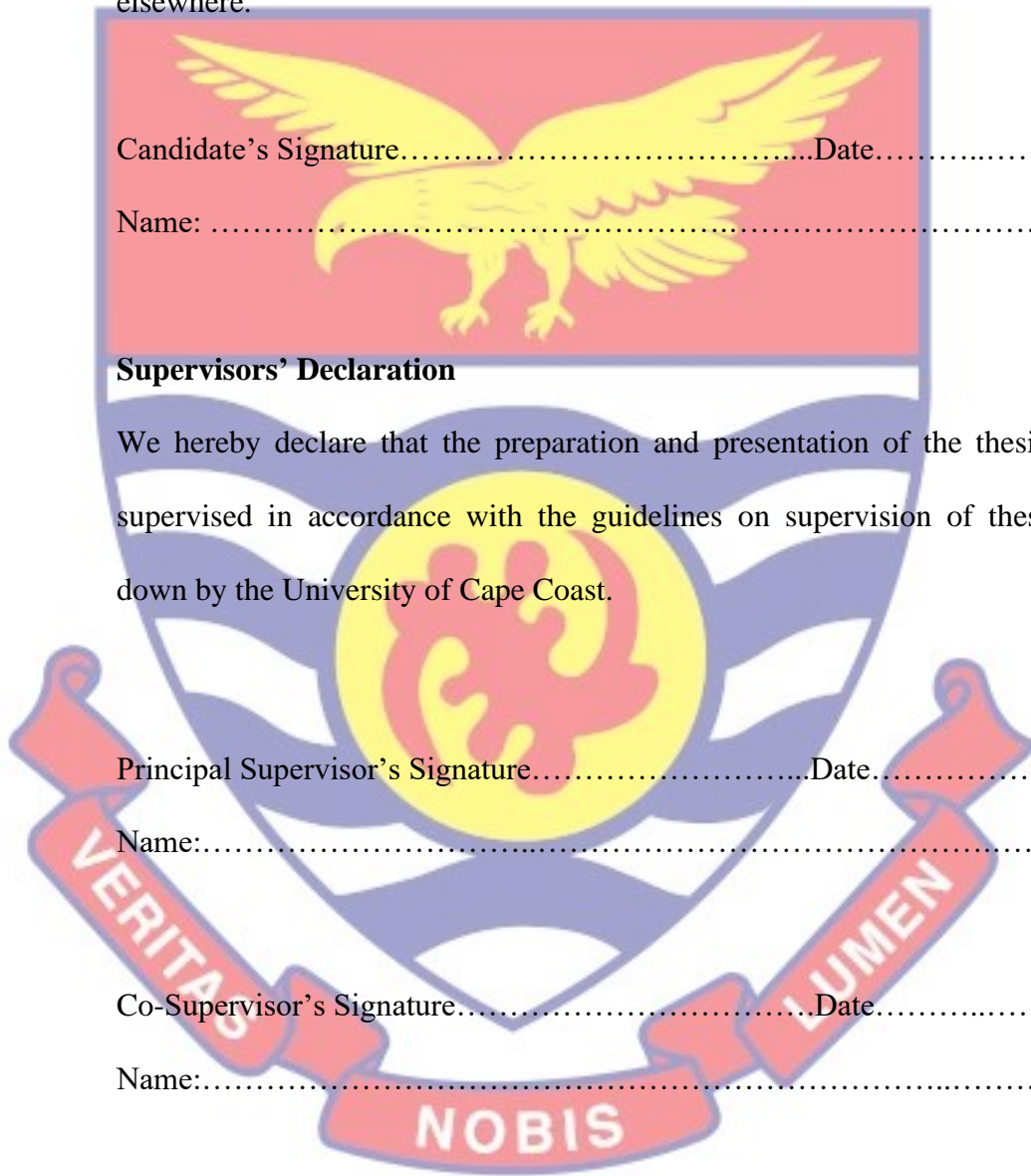
We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

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

Name:.....

Co-Supervisor's Signature.....Date.....

Name:.....



ABSTRACT

The study assessed the extent to which lecturers applied principles of test construction and administration in the University of Cape Coast. The descriptive survey design was used for the study. The census sampling procedure was used to include 131 lecturers in four faculties/schools and 40 Examinations Unit staff in the University of Cape Coast for the study. Seven research questions guided the study. A 67-item and 14-item questionnaires for lecturers and Examinations Units were used for data collection. The Cronbach's coefficient alpha for the lecturers and Examinations Unit questionnaires were 0.73 and 0.70 respectively. The results showed that, to a great extent, lecturers were aware of the test construction principles. Lecturers indicated that they defined the purpose of every test prior to construction, determined the test items format and prepared marking scheme as soon as test items were written. Again, it was found out that lecturers were aware of the test administration principles. Students were made aware of the rules and regulations governing the test during test administration. The study also showed that lecturers had difficulty in controlling large classes during quizzes and examinations. It is recommended that lecturers are given timely in-service training on test item construction and administration. Furthermore, university management should allocate funds for workshops, seminars and others for lecturers on test item construction and administration.

KEY WORDS

Test

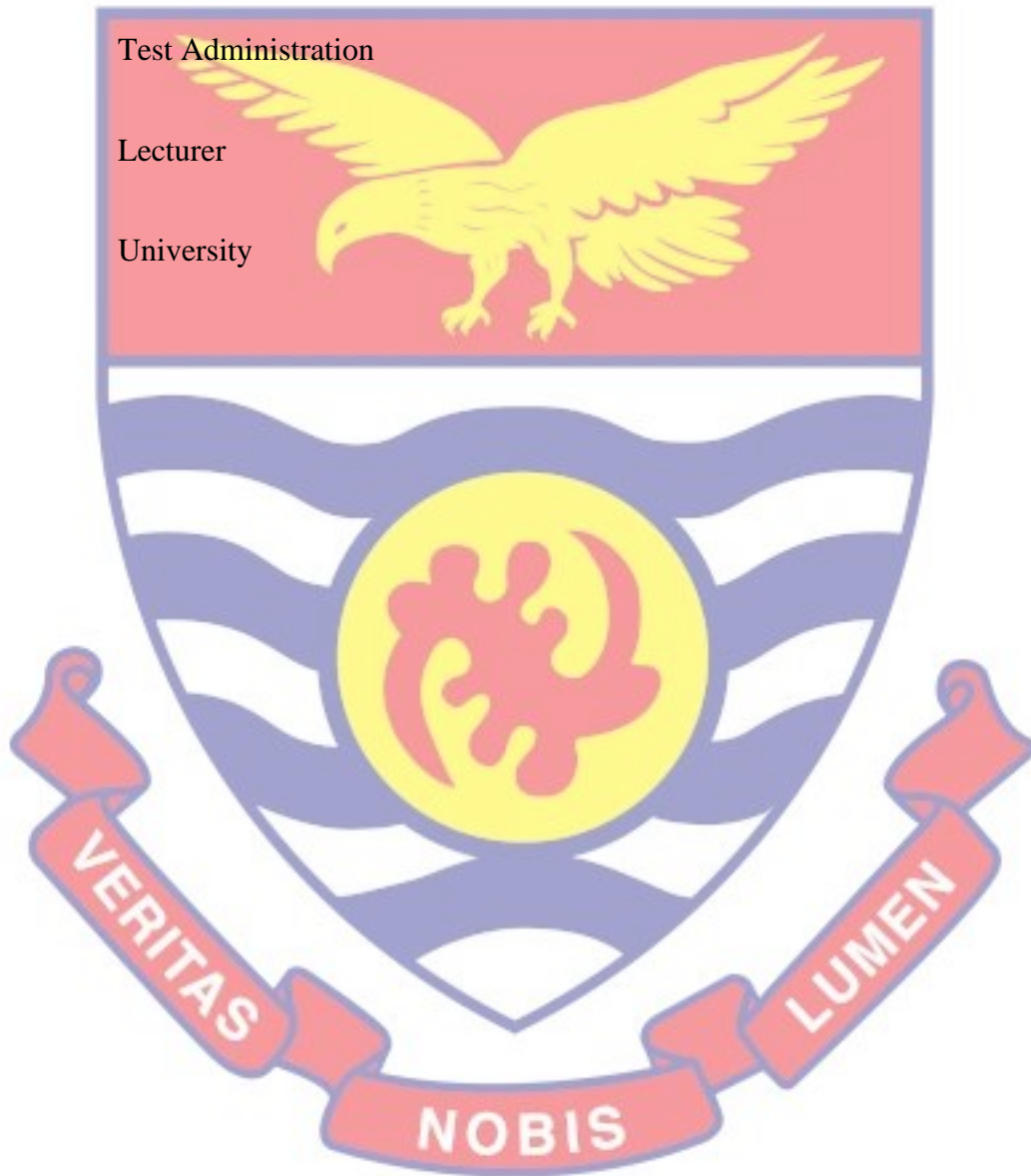
Testing Practices

Test Construction

Test Administration

Lecturer

University



ACKNOWLEDGEMENTS

I first and foremost thank the Almighty God for the successful completion of this study. My profound gratitude also goes particularly to my principal supervisor Professor Y. K. A. Etey, for his fatherly love, encouragement, guidance and support throughout the study and the entire M.

Phil programme.

I am greatly indebted to Dr. Kenneth Asamoah-Gyimah, my co-supervisor, for his expert advice and suggestions during this research work. I have learnt a lot through my interaction with him.

My extreme thanks go to my husband, Mr Nasir Yacub Entsie who always believed in me and provided the necessary encouragement and expression of faith in my ability to complete what I had begun.

I also express my sincere appreciation to all the lecturers of the Department of Education and Psychology, especially Professor F. K. Amedahe for the diverse ways he contributed to my completion of the M. Phil programme.

My appreciation also extends to Abraham Yeboah, Aaron Edusei, Abraham Gyamfi and Emmanuel Boakye for their help and encouragement in diverse ways. My last appreciation goes to all the lecturers in the University of Cape Coast, who participated in the study.

DEDICATION

To my dear husband and children



TABLE OF CONTENTS

	Page
DECLARATION	ii
ABSTRACT	iii
KEY WORDS	iv
ACKNOWLEDGEMENTS	v
DEDICATION	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xii
CHAPTER ONE: INTRODUCTION	
Background to the Study	1
Statement of the Problem	5
Purpose of the Study	7
Research Questions	8
Significance of the Study	8
Delimitations	9
Limitations	10
Definition of Terms	10
Organization of the Study	10
CHAPTER TWO: LITERATURE REVIEW	
Introduction	12
History of Testing	12
Construction of Achievement Tests	19
Principles for Constructing Matching Type Test Items	26
Principles for Writing Short Answer or Completion Items	28

Principles for Constructing Essay-Type Tests	30
Principles of Constructing Multiple Choice Items	33
General Principles of Constructing Test Items	35
Administration of Classroom Achievement Tests	37
The Challenges of Test Administration	43
Validity	44
Kinds of validity evidence	45
Factors affecting validity	48
Reliability	50
Methods of estimating reliability	50
Factors affecting reliability	53
Empirical Review	55
Test construction	55
Test administration	59
Summary of Literature Review	61
CHAPTER THREE: RESEARCH METHODS	
Research Design	63
Study Area	65
Population	65
Sampling Procedure	66
Data Collection Instruments	69
Validity of the Instrument	71
Pre-testing of the Instrument	71
Reliability of Instrument	72
Ethical Consideration	73

Data Collection Procedures	74
Data Processing and Analysis	74
Research question one	74
Research question two	75
Research question three	75
Research question four	75
Research question five	75
Research question six	75
Research question seven	76
Chapter Summary	76
CHAPTER FOUR: RESULTS AND DISCUSSION	
Description of Sample	77
Distribution of lecturers by gender	77
Research Question One	80
What is the extent of lecturer’s awareness of the principles of test construction in the University of Cape Coast?	80
Research Question Two	82
What is the extent of lecturer’s awareness of the principles of test administration in the University of Cape Coast?	82
Research Question Three	84
What is the extent to which lecturers apply the principles of test construction in setting classroom test items?	84
Research Question Four	88
What is the extent to which lecturers apply the principles of test administration in setting classroom test items?	88

Research Question Five	92
What is the extent to which the Examinations Unit of the University Cape Coast apply the principles of test administration?	92
Research Question Six	95
What are the major challenges encountered by lecturers in constructing tests in the University of Cape Coast?	95
Research Question Seven	97
What are the major challenges encountered by lecturers in test administration in the University of Cape Coast?	97
Source: Field Survey, Armah (2017)	98
Discussion of Research Findings	99
Lecturers' Awareness of Test Construction Principles	99
Lecturers' Awareness of Test Administration Principles	100
Lecturers' Application of Test Construction Principles	102
Lecturers' Application of Test Administration Principles	103
Examinations Unit Application of Test Administration Principles	104
Challenges in Test Construction	105
Challenges in Test Administration	106
Chapter Summary	107
CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS	
Overview of the Study	108
Summary of Key Findings	109
Conclusions	110
Recommendations	111

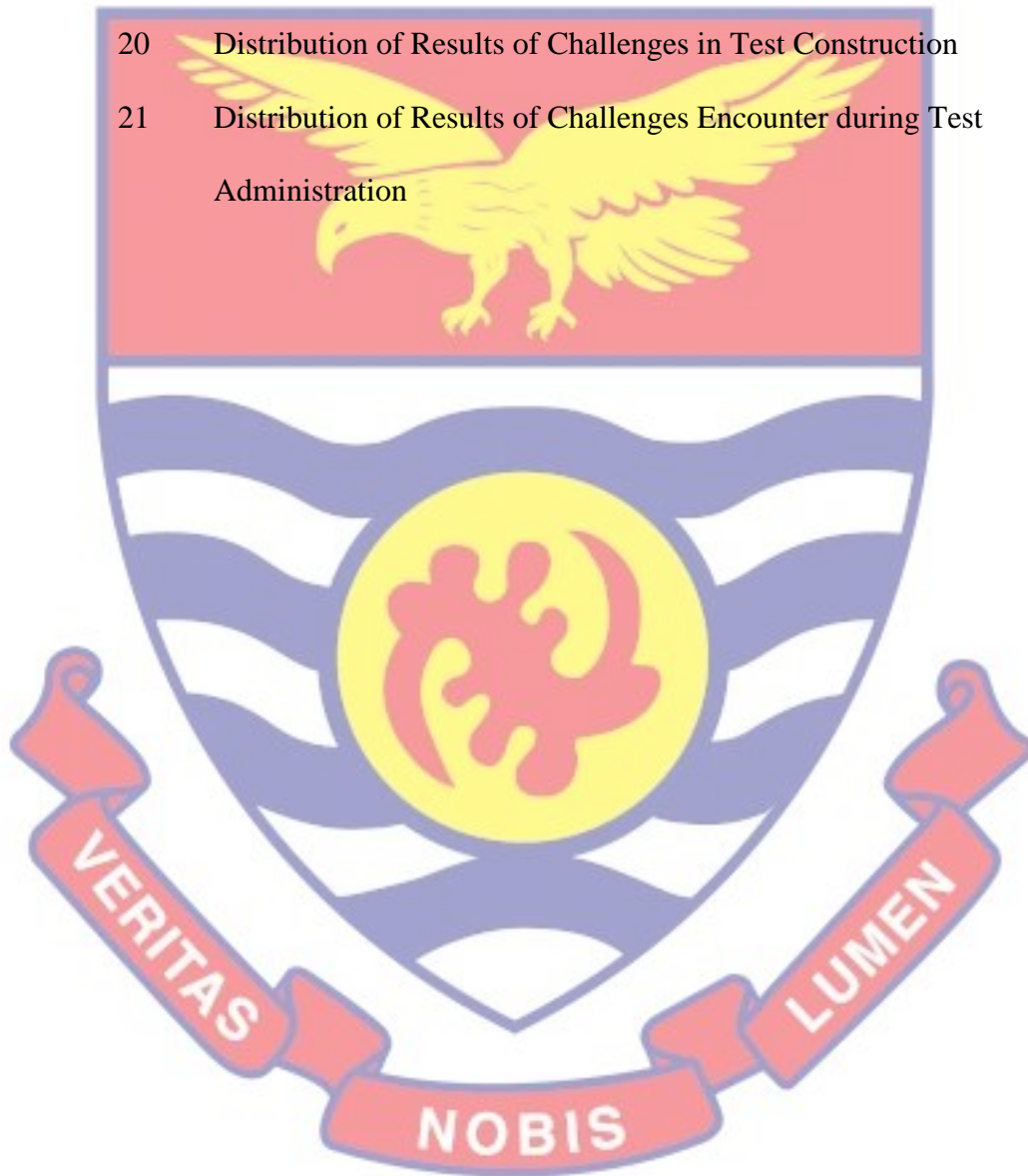
Suggestions for Further Research	112
REFERENCES	113
APPENDICES	124
APPENDIX A: QUESTIONNAIRE FOR LECTURERS	124
APPENDIX B: QUESTIONNAIRE FOR EXAMINATIONS UNIT	132



LIST OF TABLES

Table	Page
1 Colleges and Schools/Faculties under Them	67
2 Distribution of Sampled Faculties/Schools	69
3 Summary of the Reliability Coefficient of the Items	73
4 Distribution of Lecturers' by Gender	78
5 Teaching Experience of Lecturers	78
6 Distribution of Lecturers in the Selected Schools/Faculties	79
7 Distribution of Examinations Unit Workers by Gender	79
8 Working Experience of Examinations Unit Workers	80
9 Analysis of Results of Lecturers Awareness of Test Construction Principles	81
10 Analysis of Results of Lecturers Awareness of Test Administration Principles	83
11 Descriptive Statistics on Lecturers' Application of Test Construction Principles	85
12 One-Sample t-Test of Lecturers Application of Test Construction Principles	86
13 One-Sample t-Test of General Application of Test Construction Principles	88
14 Descriptive Statistics on Lecturers' Application of Test Administration Principles	89
15 One Sample t-Test of Lecturers Application of Test Administration Principles	90
16 One-Sample t-Test of General Application of Test Administration Principles	91

17	Descriptive Statistics of the Extent to which Examinations Unit	92
18	One Sample t-Test of the Extent to which the Examinations Unit Apply Test Administration Principles	93
19	One-Sample t-Test of General Application of Test Administration Principles in Examination Unit	95
20	Distribution of Results of Challenges in Test Construction	96
21	Distribution of Results of Challenges Encounter during Test Administration	98



CHAPTER ONE

INTRODUCTION

Test construction and administration practices play a crucial role in teaching and learning processes. Assessors' ability to adhere to test construction and administration practices would render test results of the students to be reliable and valid. The soundness of a particular test item makes such items devoid of ambiguities and other factors which could affect test validity. For one to know how lecturers construct and apply test administration principles in teaching and learning, it is imperative to assess the level of lecturers' awareness of test construction and administration practices.

Background to the Study

Assessment is the process of obtaining information that is used for making decisions about students, curricula and programmes, and educational policy (Etsey, 2012). It includes the full range of procedures used to gain information about students. Stiggins (1991) indicated that teachers spend much of their instructional time in assessment related activities. Tests constitute an integral part of the assessment process in education. According to Anamuah-Mensah and Quagrain (1998), classroom tests have been regarded as one of the most tangible clues and the most crucial yardstick in determining the attainment of the objectives of any learning experience in the classroom. Quagrain (1992) stated that achievement tests provide needed information for evaluation. Without evaluation there cannot be feedback and knowledge of results. Without knowledge of results there cannot be any systematic improvement in learning. Since school learning programmes are expected to provide students with the capability to complete various tasks successfully,

one way of assessing each student's learning is to give a number of these tasks to be done under specified conditions.

Mehrens and Lehmann (2005) define a test as a task or series of tasks which are used to measure specific traits or attributes in people. Tests can be standardized or teacher-made (classroom). According to Asamoah-Gyimah (2002), "classroom or teacher-made tests are frequently used as a major evaluating device of students' progress in schools" (p. 2). Hardly can one foresee or hypothesize an educational system where the student is not put under classroom or teacher-made tests. Etsey (2002) define teacher-made/classroom tests as tests constructed by classroom teachers (lecturers) for specific uses in each classroom and are closely related to a particular objective. They are usually tailored to fit the teacher or lecturer's instructional objectives.

The content of the teacher-made classroom test is determined by classroom teacher (lecturer). The field of teacher education has been moving forward in re-conceptualizing the goals of teacher education based on new visions of effective teaching (Reynolds, 1989). As more emphasis is being placed on student performance and teacher accountability, measurement and assessment are becoming increasingly important to all educators. It is critical therefore that lecturers possess not only comprehensive knowledge of subject matter but also the ability to assess the learning of the subject matter.

According to Andrews and Barnes (1990), the most important service that can be rendered to classroom teachers in a measurement course is teaching the techniques of constructing, administering and scoring classroom or teacher-made tests based on the basic principles in measurement and

evaluation. A large amount of the testing done by classroom teachers including lecturers of universities must necessarily be the measurement of the immediate results of daily lessons and larger units of work which are not suited to measurement by standardized tests.

The rationale for teacher's assessment is linked with the constructive model of learning. In this model, it is important to understand what the student knows and how he/she articulates it in order to develop the knowledge and understanding (Gipps, 1992).

In Ghana, the standardised achievement, aptitude, and intelligence tests that are found in the developed countries such as the United States of America (USA), Canada and Great Britain are to a large extent non-existent. The tests that are conducted by the West African Examinations Council (WAEC) at the terminal points of the educational system cannot be said to be standardised since they do not meet all the standard characteristics of standardised achievement tests (Oduro-Okyireh, 2008). Examples of the WAEC conducted tests are the Basic Education Certificate Examination (BECE) and the Senior Secondary School Certificate Examination (SSSCE). Linn and Gronlund (1995) outlined the features of standardised achievement tests, which include the following:

1. The test items should be of a high technical quality. They have been developed by educational and test specialists, tried out experimentally (pretested) and selected on the basis of difficulty, discriminating power and relationship to a clearly defined and rigid set of specifications.

2. Directions for administering and scoring should be precisely stated so that the procedures are standard for different users of the test.
3. Norms based on national samples of students in the grades where the test is intended for use should be provided as aids in interpreting the scores.
4. Equivalent and comparable forms of the tests should be usually provided as well as information concerning the degree to which the tests are comparable.
5. A test manual and other accessory materials should be included as guides for administering and scoring the test, evaluating its technical qualities, and interpreting and using the results.

In Ghana, standardised achievement tests are not found in the classroom. As a result of this situation, testing of achievement in the Ghanaian educational system is primarily through the use of informal classroom or teacher-made tests. But for a very effective and efficient instructional programme in any educational system, there is the need for both standardised achievement tests and teacher-made tests, in the sense that by way of their respective functions, one complements the other.

In the Ghanaian education system, where standardised achievement tests are non-existent and all the information needed for important instructional decisions are provided by informal classroom tests, there is the need for lecturers to always ensure that they follow the approved standard principles in the construction and administration of their tests. This way, they would be striving to make their test scores more reliable so that the uses to which the test scores will be put will be as sound and appropriate as possible.

Achievement testing in the Ghanaian educational system as described above is worth noting. This is because this very vital educational exercise to a large extent has become the sole duty of the classroom teacher. As to whether teachers are sufficiently prepared and professionally well equipped to perform this duty as expected, is also a matter of concern. As indicated by Amedahe (1989), irrespective of pre-service training, teachers in Ghana construct, administer, score and interpret the results of classroom achievement tests. According to him, while some teachers have received in their college courses, pre-service instruction concerning the construction, administration, scoring of tests and the interpretation of test results, others have not. The likely effects of this state of affairs seem to be seen in the testing practices, specifically construction and administration of test items of many University lecturers in Ghana today.

It is, therefore, in the light of the above that the study sought to investigate how lecturers in the University of Cape Coast, Cape Coast construct and administer test items. The study would bring to the lime light lecturers' knowledge, skills and competencies in test item construction and administration.

Statement of the Problem

The practice of assessment in the University of Cape Coast is based on the premise that lecturers in each department have had a course or training in the course they are handling. This premise assumes that lecturers in each department construct, administer and score classroom or teacher-made tests based on the basic principles of testing in measurement and evaluation. One question which comes to mind is “have all these lecturers been trained in test

construction and administration?” Studies by Amedahe (1989) and Quagrain (1992) revealed that most Ghanaian teachers had limited skills for constructing the objective and essay type tests, which are the most frequently used instruments in our schools. This is because one does not necessarily need to read Education (as a programme) to become a lecturer. Amedahe (2000) stated that “teacher-based tests may be made of a number of factors, notable among them are, training in assessment techniques, class size and a particular school’s policy in assessment standards with implications on validity and reliability of the assessment results” (p. 112-113).

Educators like Borich (2011) and Sarita (2001) emphasize the importance of having a clear articulation of what to assess using tests: that is, the knowledge, skill, and/or behaviour the stakeholders would like to measure. This helps the test developer to determine how to measure the intended learning outcomes. This means, in order to develop a test that meets appropriate standards, teachers and test developers should have a clear understanding of the intended leaning outcomes and the associated subject matter in a detailed and sufficiently precise manner to support the development of test items. As noted by Sarita (2001) test development is less challenging when the construct is the objective and the specific subject matters are concrete and discrete.

Assessment instruments must be such that the result is reliable and valid. To yield valid and reliable results, assessment instruments should be selected or constructed appropriately. In Ghana, classroom teachers specifically University of Cape Coast teachers depend on their little knowledge in constructing test items. The results of these assessment

instruments may tend to have low reliability and validity as well. Such assessment results cannot be used for making sound decisions. Amedahe (1989) noted that teachers (lecturers) do not apply any statistical item analysis to their test items to ensure their quality before administering them. Also, test specification tables are not taken into consideration by teachers when constructing test items thereby rendering the validity of the test items rather low (Amedahe, 1989). It is therefore, necessary to conduct this study to investigate how tests are constructed and administered by lecturers in the University of Cape Coast, Cape Coast.

Purpose of the Study

The general purpose of the study was to assess the extent to which lecturers applied principles of test construction and administration in the University of Cape Coast. Specifically, the study sought to:

1. assess the extent of lecturer's awareness of the principles of test construction in the University of Cape Coast,
2. assess the extent of lecturer's awareness of the principles of administration in the University of Cape Coast,
3. assess the extent to which lecturer's apply the principles of test construction in writing classroom test items,
4. assess the extent to which lecturer's apply the principles of test administration in the University of Cape Coast,
5. examine the extent to which the Examinations Unit of the University Cape Coast apply the principles of test administration,
6. investigate the challenges lecturers encounter during test construction in the University of Cape Coast and

7. investigate the challenges lecturers encounter during test administration in the University of Cape Coast, Cape Coast.

Research Questions

The study was guided by the following research questions.

1. What is the extent of lecturer's awareness of the principles of test construction in the University of Cape Coast?
2. What is the extent of lecturer's awareness of the principles of test administration in the University of Cape Coast?
3. What is the extent to which lecturers apply the principles of test construction in writing classroom test items?
4. What is the extent to which lecturers apply the principles of test administration in the University of Cape Coast?
5. What is the extent to which the Examinations Unit of the University Cape Coast apply the principles of test administration?
6. What are the major problems encountered by the lecturers in constructing tests in the University of Cape Coast?
7. What are the major challenges encountered by lecturers in test administration in the University of Cape Coast?

Significance of the Study

The findings of the study would be beneficial to the University of Cape Coast, lecturers, examination officers and other equally important stakeholders in education. The University of Cape Coast would realise the number of lecturers who have received training in educational measurement and evaluation and for that reason, lecturers that are aware of the principles of test item construction and administration. This would make management of the

University of Cape Coast to take a decision whether to advise lecturers who do not have training in educational measurement and evaluation to take a course or organize seminars and workshops for them.

Lecturers who are aware of the principles of test item construction and administration would be confident and happy with their work since they are doing what is expected of them. However, those who are not aware of principles of test item construction and administration would be challenged to enroll on educational measurement and evaluation courses thereby acquiring knowledge about test item construction and administration. The knowledge would aid them construct test items which would be reliable.

Examination officers and stakeholders would be confident in the test items lecturers would construct. This is because lecturers would use the right principles underpinning test item construction and administration during the construction and the administration of test items. The study results would also be useful for certification of students at the first-degree level, employability skills of students and to employers of labour.

Delimitations

The study focused on principles that lecturers in the University of Cape Coast employed when constructing test items. Principles underpinning test item administration by lecturers in the University of Cape Coast were also considered. However, the study did not consider principles regarding scoring and interpretations of test results.

Moreover, for proximity and familiarity of the geographical location of the study, the study was confined to the University of Cape Coast. Specifically, the study was confined to four colleges in the University of Cape

Coast. The respondents for the study were lecturers who have taught for not less than two years in the University of Cape Coast. This is because this category of lectures would have acquired skills and competences in their profession.

Limitations

The questionnaire that was employed for the study is a self-report measure and for that matter, respondents could give responses that might not reflect the actual situation on the ground.

In addition, there were some significant problems that were encountered during this research which had the tendency of affecting the results of the study, including unfavourable weather conditions and heavy down pour affected some of my schedules with the respondents.

Definition of Terms

Test construction: It is the procedure for writing test items.

Test administration: It is the procedure for distributing written test items to students for them to respond to them.

Practice: It is the process whereby lecturers adhere to test items construction and administration principles in universities.

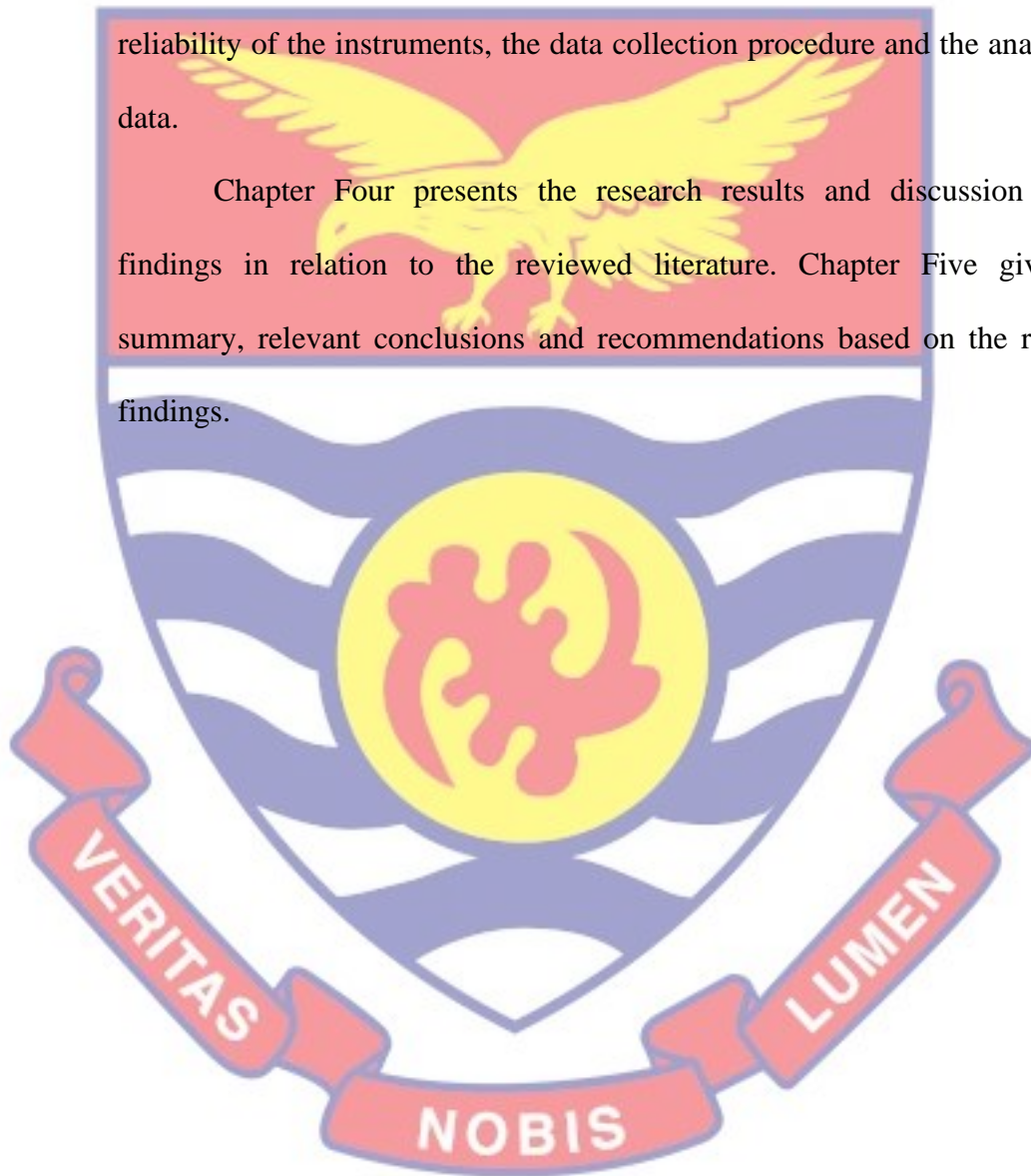
Lecturer: Teachers at the universities.

Organization of the Study

This thesis was organised into five Chapters. The study starts with Chapter One which presents the introduction and provides a background to the study and then discusses key research issues such as statement of the problem, research objectives, research questions, significance of the study, delimitations, limitations, as well as the organisation of the study.

Chapter Two of this study centres on the literature related to the study. The literature entails both conceptual framework and theoretical review. Chapter Three describes the methodology adopted for the study. It examines the research design, population, the sample and sampling procedures for the study, the research instruments, pilot testing procedure, the validity and reliability of the instruments, the data collection procedure and the analysis of data.

Chapter Four presents the research results and discussion of the findings in relation to the reviewed literature. Chapter Five gives the summary, relevant conclusions and recommendations based on the research findings.



CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter reviewed available related literatures and research findings on test construction principles and its challenges in measurement and evaluation as a whole and test administration practices and the challenges encountered. Specifically, the literature was reviewed in line with history of testing, principles of test construction and administration, validity and reliability. Empirical studies were also reviewed to support the findings of this study.

History of Testing

Individuals who do not learn from the past are often forced to repeat the mistakes of the past. Therefore, in reviewing literature on the state of the art of the assessment of human cognitive abilities, it is appropriate to look back on some of the services that have molded the development of these measures of intellectual ability with the view to understanding why they have the form and substance they do have.

The attempts to measure human cognitive abilities can be traced to a time early in the history of imperial China (Anastasi, 1982; Cunningham, 1986). According to Flanagan, Genshafts and Harrison (1997), Anastasi (1982) and Cunningham (1986), because the Chinese had no hereditary aristocracy, they developed a standardised civil service testing programme as far back as 2200 B.C and this programme lasted for about 4000 years. It was, however, discontinued when Alfred Binet displayed his scale for measuring intelligence in 1905. Researchers have pointed out that the tests covered the

examinee' knowledge of civil law, military affairs, agriculture, revenue and geography and that civil servants were tested every three years for the purposes of initial appointments and continuance in employment (Anastasi, 1982; Cunningham, 1986)

The historical development of educational measurement dates back to 1840s. For instance, the rules for conducting written exams, establishment of examination boards, practical mental tests, identification of factors of intelligence, use of objective classroom tests, scholastic aptitude tests, the development of test scoring machine, taxonomy of educational objectives are some of the reforms up to 1960s (Smith, 2005).

In England, civil service ability testing was adopted during the middle portion of the 19th century (Cunningham, 1986; Flanagan et al., 1997). Cunningham (1986) noted that the Chinese method of selecting government employees was used as a basis for the establishment of the Indian civil service. Yeboah (2017) also stated that the major concern of educational measurement in the first half of the twentieth century was the need to devise fair ways of discriminating between individuals when distributing access to the scarce and unchanging economic good of education.

Additionally, students in European schools were given oral exams until well after the 12th century, when paper began replacing parchment and papyrus. During the 16th century the Jesuits started using tests for the evaluation and placement of their students.

In USA, testing began in the later part of the 19th century, (Cunningham, 1986; Flanagan et al., 1997). It was pointed out that following the successful use in England of the Chinese method of selecting government

employees, the method was adopted in the USA. Cunningham pointed out further that the first civil service was established in 1883.

Formal testing in schools (paper and pencil tests) began with the introduction of paper in the 12th century, (Cunningham, 1986). According to Cunningham (1986), assessment by means of written tests was first used by the Jesuits at St Ignatio. He noted that the development of academic tests was pioneered in Britain, particularly in the University of London. Under its initial charter, testing and awarding of degrees were recognised as a legitimate basis for decision making. It is worth noting however, that, prior to this period, academic testing (oral testing) in schools had already begun. As stated by Anastasi (1982) among the ancient Greeks, testing was an established adjunct to the educational process. Tests were used to assess physical as well as intellectual skills. Oduro-Okyireh (2008) pointed out that the Socratic method of teaching with its interweaving of testing and teaching has much in common with today's programmed learning. On the account of Anastasi (1982) and Amedahe (1989), from their beginnings in the Middle Ages, European universities relied on formal examinations in awarding degrees and honours. These examinations, however, were largely oral.

Test development like many other aspects within psychology and education is a product of many contributors and disciplines throughout history. Notable among the early thinkers were the following personalities. Charles Darwin (1809 – 1882), a trained physician and later a clergyman, published the book-The Origin of the Species in 1859. He was an important factor in the increased acceptance of individual differences (Cunningham, 1986; Flanagan et al., 1997).

British mathematician and physicist, Sir Francis Galton (1822–1911), is generally recognised as the founder of formal testing. Galton's most important contributions were his emphasis on individual differences, which is the corner stone of the field of psychological measurement, his initial attempts to establish norms and standard scores, and his laying of the foundation for the development of the correlation coefficient. Credit for coining the term —mental test in 1890 however, is given to the American psychologist, James Mckeen Cattell (1860–1944). Galton and Cattell worked together to propel the field of mental testing forward in large and definable units (Anastasi, 1982; Cunningham, 1986; Flanagan et al., 1997, as cited in Oduro-Okyireh, 2008).

A mathematician and statistician of the first order, Karl Pearson (1857–1936), who was a student of Galton needs to be acknowledged. He derived the mathematical underpinnings of regression (then referred to as reversion), correlation, and covariation of observable phenomena in a manner that allowed Galton to make inferences about unobservable phenomena (Anastasi, 1982; Cunningham, 1986; Flanagan et al., 1997). It is worthy of noting that the correlation coefficient of Galton and Pearson continues to be used as the basis for reliability and validity coefficients in educational and psychological testing today (Oduro-Okyireh, 2008).

Alfred Binet (1857–1911), a French psychologist, developed the first intelligence test that measured high level mental functioning called the Binet-Simon test in 1905 together with Theodore Simon (1872–1961). Louis Terman is credited with the modification of the Binet-Simon tests in 1916 and coming out with the Stanford-Binet test which was the first well-standardised and carefully developed intelligence test. With the ongoing development in the

field of measurement at the time, the use of the Stanford-Binet test as an individual intelligence test declined after the introduction of the Wechsler tests developed by Wechsler in 1939 (Flanagan et al., 1997).

In the United Kingdom, examinations have their origin with the emergence of competitive entry to the civil service and the universities.

Mental testing started with attempts to identify more precisely those children requiring the provision of special education provision, and it was developed in the context of “clever plus” selection for grammar school (Akplu, 1989).

Tyler (1986) also stated that the major concern of educational measurement in the first half of the twentieth century was the need to devise fair ways of discriminating between individuals when distributing access to the scarce and unchanging economic good of education.

During World War I, Arthur Otis (1886 – 1964), under the tutorship of Louis Terman in 1917, developed the first group tests of intelligence which were used to screen recruits for intellectual fitness. Arthur Otis is further credited with the design and introduction of multiple-choice and other objective-test items (Oduro-Okyireh, 2008). It is worth noting that achievement testing in Ghanaian schools today involves the use of multiple-choice and other objective type tests.

Discoveries, innovations and development continued in the field of educational measurement over the years and by 1945 many of the theories and principles used in educational testing today had been developed (Amedahe, 1989).

Achievement Tests

Achievement tests in general are the universal way of inferring students' abilities, feelings, attitudes, opinions and achievements of what one has studied within a given period of instruction (Nitko, 2004). This format of assessing students is primarily "paper and pencil". Achievement tests are generally classified as objective and essay tests (Ebel, 1979; Etsey, 2008; Osterlind, 2006; Amedahe, 2008; Nitko, 2001; Brown, 1976). An objective test requires a respondent to give a brief response which is usually not more than a sentence long. Basically, two types of objective tests are used in schools and colleges. These are the selection and the supply types. The selection type consists of the multiple choice type, true and false type and matching types. The supply type has variations as completion, fill-in the blanks and short answers with the students supplying their own responses.

Each of the types of objective tests has its own advantages and disadvantages. The multiple choice and supply type are used to measure single units of learning outcomes. Knowledge and cognition can be measured together with comprehension (Nitko, 2001).

Supply/completion type is used to assess lower-order thinking skills such as recall and comprehension of information. Relatively simple learning outcomes are assessed using this type of test. The true or false item is used to measure students' ability to identify the correctness of statements of facts, definitions and principles. With true or false item, utility is placed primarily in assessing knowledge of factual information. True or false items are difficult to prepare (Salvia & Ysseldyke, 2001). True or false items are made up of four types, namely simple true or false with only two choices, (true or false);

complex true or false comprising three choices (true or false and opinion); compound true or false consisting of two choices (true or false) plus a conditional completion response; and multiple true (Salvia & Ysseldyke, 2001).

Objective type tests do not provide opportunities for bluffing and are best suited for measuring lower-level behaviours like knowledge and comprehension, economy of time in scoring and it is amenable to item and statistical analysis and scores are not affected by extraneous factors such as likes and dislikes (Etsey, 2008; Amedahe, 2010). The objective type test is however difficult to construct as it needs more time in planning and crafting and also cannot be used to assess complex learning outcomes satisfactorily like essay type test especially in synthesis, analysis and evaluation outcomes.

The second classification of test types is the essay type test. An essay test is a test that gives freedom to the learner to compose his/her own response using his/her own words (Etsey, 2008). Usually, the tests consist of relatively fewer items as compared to the objective type test but each item demands an extended response. There are two types of essay-tests, the restricted response type and the extended response type. The restricted response type limits the respondents to a specified length and scope of the response and the extended response type does not limit the respondent in the length and scope of the answer (Amedahe, 2010).

Extended-response type of essay questions has no bounds placed on the student as to the point(s) he/she has to discuss and the type of organization he/she will use. This type of question permits the student to demonstrate his/her ability to:

1. Call on factual knowledge
2. Evaluate factual knowledge
3. Organize ideas
4. Present ideas in a logical and coherent written fashion (Amedahe, 2010, p.25)

The extended response makes the greatest contributions at the levels of synthesis and evaluation of writing skills (style, quality). Skills such as good study habits are developed as learning materials are studied in whole. Essay type test is practical for testing a small number of students and best suited for testing higher-order behaviours and mental processes (Etsey, 2008; Amedahe, 2010).

However, essay type tests are difficult to score objectively and also provide opportunity for bluffing where students write irrelevant and unnecessary material. They are time consuming to both the teacher who scores the response and the student who does the writing. They are also susceptible to the halo effect where the scoring is influenced by extraneous factors such as the relationship between scorer and respondent. Lastly, items are an inadequate sample of subject content as several content areas are omitted when constructing essay tests type (Etsey, 2008; Amedahe, 2010).

Construction of Achievement Tests

Quality classroom-based assessment means adherence to standard procedures for test construction. Every classroom teacher is expected to possess and apply requisite skills in construction of good items for class assessments.

Based on available literature, the test construction principles that the study judged as most comprehensive and practicable in the classroom testing situation were those postulated by Tamakloe, Atta and Amedahe (1996) and Etsey (2004). These are in eight steps. The steps are:

1. Define the purpose of the test,
2. Determine the item format to use,
3. Determine what is to be tested,
4. Write the individual items,
5. Review the items,
6. Prepare the scoring key,
7. Write directions, and
8. Evaluate the test.

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

The second step of the planning stage is the determination of the item format to use. As stated earlier in the literature, the most common item formats in classroom achievement testing are the essay and the objective-types. According to Etsey (2004), it is sometimes necessary to use more than one item format in a single test. This is because depending on the purpose of the test, one item format cannot be used exclusively to measure all learning outcomes. According to Mehrens and Lehmann (1991), the choice of an

appropriate item format depends on factors such as the purpose of the test, the time available to prepare and score the test, the number of students to be tested, the skills to be tested, the difficulty level desired, the physical facilities available for reproducing the test, the age of the students and the teacher's skill in writing the different types of items.

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

To plan a test, one needs to prepare a two-way table, which is called a test blueprint or table of specification. The table covers the names of the major categories of taxonomy heads for the columns, while the row heading indicates the major topics of the subject matter to be tested (Sarita, 2005). In the body of the table, the "cells", formed by a combination of a particular taxonomy category and a particular subject-matter topic. Thus, the blue print serves as a double-entry classifying a scheme for specific objectives (Sarita,

2005). Gay (2011) also noted that a blue print is a table that gives the total number of items to be written with the content areas and levels of behavioural complexity taught. He further explained that the test blue print ensures that the test will sample learning across the range of content areas covered by the teacher and the cognitive and affective skills and processes one considers as important. He also suggests the steps in constructing a test blue print. These include:

1. Classifying each instrumental objective for the contents.
2. Recording the number of items to be constructed for each objective.
3. Totaling the items of each instructional objective totaling the number of items falling into each behaviour.
4. Completing the column and row percentages (Gay, 2011, p.36)

Sarita (2005) indicated that before planning and setting a classroom test, the following decisions must be taken into account: How the test is to be used, when to test, what emphasis to give to various aspects of achievement and whether the test should emphasize complex achievement. He further indicated that the context to be covered by the test should be stated, what level and distribution of difficulties are appropriate for the questions included in the test, what means and format to use in presenting the test to students and finally to test important areas of the content.

Test items are written at the planning stage. Tamakloe et al., (1996) and Etsey (2004) pointed out that whichever test item types that are being constructed must follow the basic principles laid down for them. There are, however, general guidelines that according to Mehrens and Lehmann (1991) and Etsey (2004), apply to all types of tests. This includes:

1. The table of specifications must be kept before the teacher and continually referred to as the items are written.
2. The test items must be related to and match the instructional objectives.
3. Well-defined items that are not vague and ambiguous must be formulated. Grammar and spelling errors must be checked. Textbook or stereotyped language must be avoided.
4. Excessive verbiage and complex sentences must be avoided.
5. The test items must be based on information that students should know or are expected to know.
6. More items than are actually needed in the test must be prepared in the initial draft. Mehrens and Lehmann (1991) suggested that the initial number of items should be 25% more while Amedahe (1989) has suggested 10% more items than are actually needed in the test.
7. Items of varying levels of difficulty must be used. This, however, depends on the purpose of the test.
8. The items and the scoring keys must be written as early as possible after the material has been taught.
9. The test items must be written in advance (at least two weeks) of the testing date to permit reviews and editing.

After the items, have been written, Tamakloe et al. (1996) call the next stage the item preparation stage. At this stage the test items must be reviewed and edited. Etsey (2004) suggested that the items must be critically examined at least a week after writing them. He emphasised that where possible, fellow teachers or colleagues in the same subject area should review the test items.

Reviewing and editing the items are for the purpose of removing or rewording poorly constructed items, checking difficulty level of items and checking the length of the test. All test items should be checked for technical errors and irrelevant clues.

When the reviews and editing are done, the test items can now be assembled. In assembling test items, the following points must be considered (Etsey, 2004; Kubiszyn & Borich, 1984; Mehrens & Lehmann, 1991; Tamakloe et al., 1996)

1. The items should be arranged in sections by item formats. The sections must progress from easier formats (true-false) to more difficult formats (interpretive exercises and essay).
2. Within each section or format, the items must be arranged in order of increasing difficulty. One way of achieving this is to group items in each format according to the instructional objectives being measured and make sure that they progress from simple to complex.
3. The items must be spaced and numbered consecutively so that they are not crowded and can easily be read.
4. All stems and options must be together on the same page and if possible, diagrams and questions must be kept together.
5. If a diagram is used for a multiple-choice test, the diagram must be placed above the stem.
6. A definite response pattern to the correct answer must be avoided.

According to Mehrens and Lehmann (1991), such a grouping has the advantage of helping the teacher to ascertain which learning activities appear to be most readily understood by students, those that are least understood and

those that are in-between. Mehrens and Lehmann noted that ordering items in ascending order of difficulty leads to better performance than either a random or hard-to-easy ordering. Although, empirical evidence is also inconclusive about the effectiveness of using statistical item difficulty as a means of ordering items, Sax and Cromack (2005) and other testing experts have recommended that for lengthy or timed tests, items should progress from the easy to the difficult if for no other reason than to instill confidence in the examinee, especially at the beginning. It should be noted however, that, the use of statistical item difficulty or item difficulty indexes by the classroom teacher seems impracticable to a large extent (Kubiszyn & Borich, 1984; Tamakloe et al., 1996). This is because statistical item difficulty data are always gathered after test administration or test try outs and teacher-made test items are usually not pre-tested. Mehrens and Lehmann (1991) however, recommended that subjective judgement must be relied on to determine difficulty level of items. They have stated that teachers could only categorise their items as “difficult, average or easy” (p. 71).

In addition to the above, Koksai (2004) indicated that in test construction, it is essential that the teacher asks the following questions: Is the task perfectly clear? Is there more than one possible correct answer? Can test-takers arrive at the correct response without having the skill supposedly being tested? Do test-takers have enough time to perform the task(s)?

Principles for Constructing Matching Type Test Items

Miller, Linn and Gronlund (2009) noted the following as principles for writing matching type items.

1. Use only homogeneous material in a single matching exercise. It is without a doubt the most important rule of construction and yet the one most commonly violated. One reason for this is that homogeneity is a matter of degree, and what is homogeneous to one group may be heterogeneous to another.

2. Include an unequal number of responses and premises and instruct the student that responses may be used once, more than once, or not at all. This will make all the responses eligible for selection for each premise and will decrease the likelihood of successful guessing. When an equal number of responses and premises are used and each response is used only once, the probability for guessing the remaining responses correctly is increased each time a correct answer is selected. The odds for correct guessing increase as the list of available responses decreases, and the final response, of course, can be selected entirely on the basis of this process of elimination.

3. Keep the list of items to be matched brief and place the shorter responses on the right. A brief list of items is advantageous to both the teacher and the student. From the teacher's standpoint, it is easier to maintain homogeneity in a brief list. In addition, there is a greater likelihood that the various learning outcomes and subject-matter topics will be measured in a balanced manner. Because each matching exercise must be based on homogeneous material, a long list will

require excessive concentration in one area. From the students' viewpoint, a brief list enables them to read the responses rapidly and without confusion. Approximately four to seven items in each column seems best.

4. Arrange the list of responses in logical order, place words in alphabetical order, and numbers in sequence. This will contribute to the ease with which the students can scan the responses in searching for the correct answers. It will also prevent them from detecting possible clues from the arrangement of the responses.

5. Indicate in the directions the basis for matching the responses and premises. Although the basis for matching is rather obvious in most matching exercises, there are advantages in clearly stating it. First, ambiguity and confusion will be avoided. Second, testing time will be saved because the student will not need to read through the entire list of premises and responses and then "reason out" the basis for matching. Special care must be taken when stating directions for matching items. Directions that precisely indicate the basis for matching frequently become long and involved, placing a premium on reading comprehension.

Also, Mehrens and Lehmann (1991) noted the following as principles for constructing matching type test.

1. If at all possible, have the response list consist of short phrases, single words, or numbers.
2. Each matching exercise should consist of homogeneous items.
3. Keep each list relatively short.

4. Avoid having an equal number of premises and responses.
5. Arrange the answers in some systematic fashion.
6. Avoid giving extraneous irrelevant clues.
7. Explain clearly the basis on which the match is to be made.
8. Maintain grammatical consistency.
9. Every response in one column should be a plausible answer to every premise in the other column.
10. Check suggestions for writing objective item.

Principles for Writing Short Answer or Completion Items

The short-answer item is subject to a variety of defects, even though it is considered one of the easiest to construct. The following suggestions were indicated by (Koksal, 2004; Sarita, 2005). These processes include:

1. Word each statement in specific terms with clear meanings so that the intended answer is the only a single word, brief phrase or number.
2. Word each item so that the blank or answer space is toward the end of the sentence.
3. Avoid copying statements verbatim from texts or classroom materials.
4. Omit important rather than trivial words.
5. Avoid “butchered” or “mutilated” sentences, use only one or two blanks in a completion sentence.
6. Keep the blanks of equal length and arrange the items so the answers are placed in a column at the right or left of the sentences.
7. State the precision, numerical units, or degree of specificity expected of the answer.
8. Word the items to avoid irrelevant clues or specific determiners.

Mehrens and Lehmann (1991) further noted the following as principles for constructing short-answers.

1. Word the item so that the required answer is both brief and specific. As indicated earlier, the answer to an item should be a word, phrase, number, or symbol. This can be easily conveyed to the students through the directions at the beginning of the test and by proper phrasing of the question. More difficult is stating the question so that only one answer is correct.
2. Do not take statements directly from textbooks to use as a basis for short-answer items. When taken out of context, textbook statements are frequently too general and ambiguous to serve as good short-answer items. When items are not taken verbatim from another source, it also forces the item writer to carefully consider the importance of the word or phrase being omitted as well as the other words or phrases in the sentence that might provide a clue to the answer.
3. A direct question is generally more desirable than an incomplete statement. There are two advantages to the direct-question form. First, it is more natural to the students, as this is the usual method of phrasing questions in daily classroom discussions. Second, the direct question is usually better structured and free from much of the ambiguity that creeps into items based on incomplete statements.
4. If the answer is to be expressed in numerical units, indicate the type of answer wanted. For computational problems, it is usually preferable to indicate the units in which the answer is to be expressed. This will clarify the problem and will simplify the scoring.

5. Blanks for answers should be equal in length and in a column to the right of the question. If blanks for answers are kept equal in length, the length of the blank space does not supply a clue to the answer.

Mehrens and Lehmann (1991) further noted the following as principles for constructing short-answers.

1. For computational problems, the teacher should specify the degree of precision and the units of expression expected in the answer.
2. Omit important words only.
3. Avoid excessive blanks in a single item.
4. Have the blanks occur near the end of the sentence.
5. Use the direct question rather than the incomplete statement.
6. Avoid giving irrelevant clues.
7. Write the item so that there is only one correct answer.
8. Don't take statements directly from the text, omit a word, and use the statement as a test item.
9. Word the item so that the required answer is brief and specific.
10. Check the suggestions for writing objective items.

Principles for Constructing Essay-Type Tests

Mehrens and Lehmann (1991) indicated the following as principles for constructing essay-type tests.

1. The essay question should be used only to assess those instructional objectives that cannot be satisfactorily measured by objective items.
2. Give adequate time and thought to the preparation of essay questions.
3. The question should be written so that it will elicit the type of behaviour you want to measure.

4. A well-constructed essay question should establish a framework within which the student operates.

- (a) Delimit the area covered by the question.
- (b) Use words that themselves give directions.
- (c) Give specific directions to the students to the desired response.
- (d) Indicate clearly the value of the question and the time suggested for answering.

Also, Miller, Linn and Gronlund (2009) maintained that the improvement of the essay question as a measure of complex learning outcomes requires attention to two problems: (1) How to construct essay questions that call for the desired student responses, and (2) how to score the answers so that achievement is reliably measured. They outlined principles such as:

1. Restrict the use of essay questions to those learning outcomes that cannot be measured satisfactorily by objective items. Other things being equal, objective measures have the advantage of efficiency and reliability. When objective items are inadequate for measuring the learning outcomes, however, the use of essay questions can be easily defended despite their limitations. Complex learning outcomes such as those pertaining to the organization, integration, and expression of ideas will be neglected unless essay questions are used. By restricting the use of essay questions to these areas, the evaluation of student achievement can be most fully realized.
2. Construct questions that will call forth the skills specified in the learning standards. Like objective items, essay questions should measure the achievement of clearly defined content standards or instructional outcomes. If the ability to apply principles is being measured, for example, the

questions should be phrased in such a manner that they require students to display their conceptual understanding or a particular skill. Essay questions should never be hurriedly constructed in the hope that they will measure broad, important (but unidentified) educational goals. Each essay question should be carefully designed to require students to demonstrate achievement defined in the desired learning outcomes.

3. Phrase the question so that the student's task is clearly defined. The purpose a teacher had in mind when developing the question may not be conveyed to the student if the question contains ambiguous phrasing.
4. Indicate an approximate time limit for each question. Too often, essay questions place a premium on speed because inadequate attention is paid to reasonable time limits during the test's construction. As each question is constructed, the teacher should estimate the approximate time needed for a satisfactory response. In allotting response time, keep the slower students in mind.
5. Avoid the use of optional questions. A fairly common practice when using essay questions is to give students more questions than they are expected to perform and then permit them to select a given number.

Also, Koksal (2004) and Sarita (2005) outlined steps in essay-type tests. These are:

1. Define behaviour the examinee is expected to exhibit or describe the process to be exhibited before beginning to write the essay question.
2. Ask questions that require the examinee to determine the ability to use essential knowledge and to do so in situations that are new or novel for

the examinee, rather than simplify recalling information from a textbook or a classroom.

3. Ask questions that are relatively specific and which require specific responses.
4. If a test includes several essay questions, be sure that they cover the appropriate range of topics and complexity of behaviour called for in the test blue-print, but it must be noted that the complexity of the questions is within the educational maturity level of the examinees.
5. Require all the examinees to answer the same questions: don't give optional questions where applicable.

Principles of Constructing Multiple Choice Items

The multiple-choice format is widely used in achievement tests of all types to assess a variety of learning outcomes. Multiple-choice tests cannot only measure factual recall, but can also measure the students' ability to reason and to exercise judgement. Before students can do this, they therefore need to adapt deep learning strategies in their studies to help them understand what they were taught.

Multiple-choice tests afford excellent content sampling, which generally leads to more content-valid score interpretations. They can be scored quickly and accurately by machines, clerks, teaching assistance, and even students themselves. This is so because the element of subjectivity in scoring is totally absent in the multiple-choice test. Multiple-choice items do not require students to write out and elaborate their answers. This minimizes the opportunity for less knowledgeable students to "bluff" or "dress up" their answers.

The general applicability and the superior qualities of multiple-choice test items are realized most fully when care is taken in their construction. This involves formulating a clearly stated problem, identifying plausible alternatives, and removing irrelevant clues to the answer. Miller, Linn and Gronlund (2009) suggested the following principles for writing multiple-choice items.

1. The stem of the item should be meaningful by itself and should present a definite problem.
2. The item stem should include as much of the item as possible and should be free of irrelevant material. This will increase the probability of a clearly stated problem in the stem and will reduce the reading time required.
3. Use a negatively stated stem only when significant learning outcomes require it.
4. All the alternatives should be grammatically consistent with the stem of the item.
5. An item should contain only one correct or clearly best answer.

Including more than one correct answer in a test item and asking students to select all the correct alternatives has two shortcomings. First, such items are usually no more than a collection of true—false items presented in multiple-choice format. They do not present a definite problem in the stem, and the selection of answers requires a mental response of true or false to each alternative rather than a comparison and selection of alternatives. Second, because the number of alternatives selected as correct answers varies from one student to

satisfactory scoring methods are more cumbersome than most teachers are likely to want to use or have to explain to students.

In addition, Mehrens and Lehmann (1991) suggested the following as principles for writing multiple-choice items.

1. The essence of the problem should be in the stem.
2. Avoid repetition of words in the options.
3. Avoid superfluous wordings. The stem should be concise and clear.
4. When the incomplete statement format is used, the options should come at the end of the statement.
5. Arrange the alternatives as simply as possible.
6. Avoid highly technical distracters.
7. All distracters should be plausible.
8. Avoid making the correct answer consistently longer than the incorrect ones.
9. Avoid giving irrelevant clues to the correct answers.

General Principles of Constructing Test Items

Sarita (2005) and Gay (2011) outlined some basic tips for designing test items which include:

1. Create new test items each time one teaches a course. Each time one teaches a course, one probably does so a little bit differently; different ideas and concepts are emphasized. Using old tests usually means one is not testing on what one has actually taught.
2. Leave yourself time to write the test. Do not write the test the night before it is to be taken. Give yourself one to two weeks to refine the test questions and format.

3. Create a bank of questions during the term. You do not need to wait until just before an exam to write the questions. If you pay attention to the questions and discussion in class, you will probably have ideas after

each class about what the students are comprehending as well as what they are questioning. Write test questions down as they come to you. This helps to ensure that you test all of what you have taught, covers what the students know and should know, and saves you work down the road.

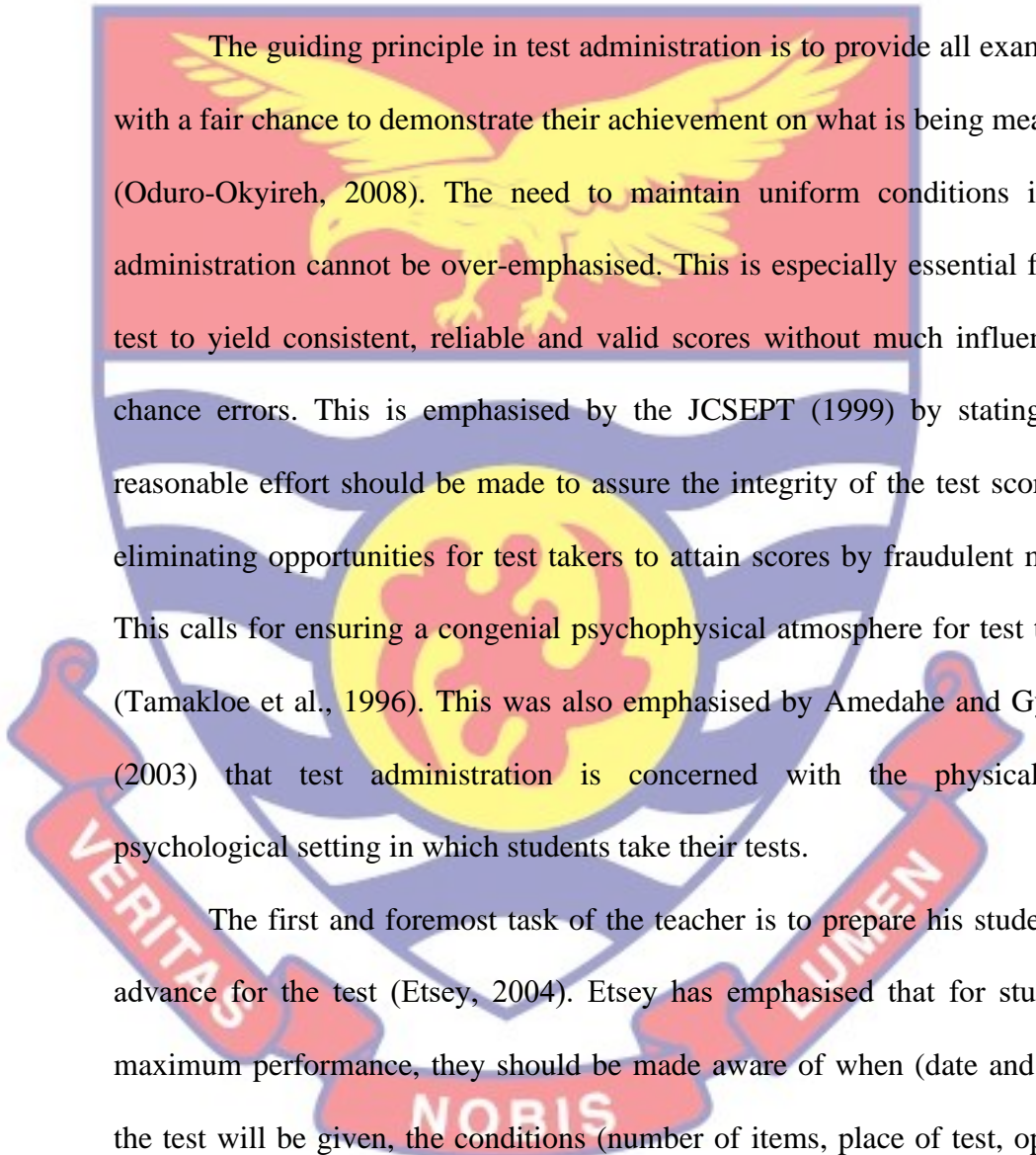
4. Pay attention to the layout of the examinations. The layout should be clear, crisp and easy to read. If you have different kinds of questions, group them together. If point values change, state what the value is for the question(s) as appropriate.

Also, Mehrens and Lehmann (1991) noted the following as general principles regarding construction of test items.

1. Define the instructional objectives.
2. Prepare a table of specifications.
3. Formulate well-defined questions.
4. Avoid excess verbiage (needlessly complex sentence).
5. The test item should be based on information that the examinee should know without having to consult a reference source
6. Avoid race and sex bias.
7. Prepare more items than one will actually need.
8. Write and key the test items as soon as possible after the material has been taught.

9. Prepare the items well in advance to permit review and editing.
10. Avoid specific determiners.
11. Insert some novelty into one test.
12. Avoid copying from textbooks.

Administration of Classroom Achievement Tests



The guiding principle in test administration is to provide all examinees with a fair chance to demonstrate their achievement on what is being measured (Oduro-Okyireh, 2008). The need to maintain uniform conditions in test administration cannot be over-emphasised. This is especially essential for the test to yield consistent, reliable and valid scores without much influence of chance errors. This is emphasised by the JCSEPT (1999) by stating that, reasonable effort should be made to assure the integrity of the test scores by eliminating opportunities for test takers to attain scores by fraudulent means. This calls for ensuring a congenial psychophysical atmosphere for test taking (Tamakloe et al., 1996). This was also emphasised by Amedahe and Gyimah (2003) that test administration is concerned with the physical and psychological setting in which students take their tests.

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

on the test, how the test will be scored and graded, and the importance of the results of the test.

The physical conditions that need to be in place to ensure maximum performance on the part of students include adequate work space, quietness in the vicinity, good lighting and ventilation and comfortable temperature (Etsey, 2004; Gronlund, 1985; Lindquist, 1996). Adequate work space is very essential for test administration because when tables and chairs are closely arranged together, students will not have the independence to work on their own. This will in no doubt lead to students copying from each other. In addition, tables provided for the examination must be conducive to the testing materials being used. For example, in practical geography examinations where topographical sheets are used, each student could use two tables or desks in order to get adequate work space (Tamakloe et al., 1996).

Noise and distraction in the testing environment should be kept at the barest minimum if not eliminated completely. Interruptions within and outside the testing room has the tendency of affecting student's performance (Mehrens & Lehmann, 1991; Tamakloe et al., 1996). Etsey (2004) has pointed out that it is helpful to hang a "Do Not Disturb Testing in Progress" sign at the door of the testing room to warn people to keep off. Good lighting is important in effective test administration. This facilitates students' reading of instructions and test items without straining their eyes, thereby working faster (Gronlund, 1985). "Good ventilation and comfortable temperature should be assured since their absence could create unrest or uneasiness in testees making concentration difficult" (Tamakloe et al., 1996, p. 215).

Etsey (2004), Amedahe (1989), Gronlund (1985) and Tamakloe et al., (1996) indicated that other basic physical conditions are that, all testing equipment must be in the room and readily available, and also, all possible emergencies during test administration must be expected and well catered for. The psychological conditions in test administration, on the other hand, include the position of the invigilator, timing of the test, threatening behaviours of invigilators, and interruption to give instructions and announcements.

A study on the examiner as an inhibiting factor, carried out by Bernstein (1953) and reported by Amedahe (1989) found out that, the presence of the examiner tended to inhibit the performance of those students who were nervous. The crux of the matter is that if the mere presence of the examiner or invigilator could affect the performance of students who are nervous, then there is no doubt that the position of the invigilator is very significant to the performance of students on examinations. Etsey (2004) has recommended that the invigilator should stand where all students could be viewed and move among the students once a while to check malpractices. Such movements should not disturb the students. He must be vigilant. Reading novels or newspapers, making of and listening to telephone calls, dozing off and chatting are not allowed.

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

Interruptions during testing, such as giving instruction, must be kept to the barest minimum and should always relate to the test. The time spent and time left to complete the test must be announced at regular intervals to enable students apportion their time to the test items. Where practicable, the time should be written on the chalkboard at 15-minute intervals until near the end of the test when it could be changed every five minutes. Further, students should start the test promptly and stop on time (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996). Teachers should always work at minimising test anxiety in students during testing. They should therefore, avoid, warning students to do their best because the test is important, telling students that they must work faster in order to finish on time, threatening dire consequences of failure in the test, and threatening students with tests if they do not behave (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996).

A study by Sarita (2005) indicated that when it is time for students to take the test, there are several things one should keep in mind to make the experience run as smoothly as possible. He noted that the following points should be considered.

1. Have extra copies of the test on hand, in case you have miscounted or in the event of some other problem;
2. Minimize interruptions during the exam by reading the directions briefly at the start and refraining from commenting during the exam unless you discover a problem;
3. Periodically write the time remaining on the board;

4. Be alert for cheating but do not hover over the students and cause a distraction.

There are also some steps that one can take to reduce the anxiety that students will inevitably feel leading up to and during an examination (Sarita, 2005). These include:

1. Have old exams on file in the department office for students to review.
2. Give students practice exams prior to the real test.
3. Explain, in advance of the test day, the exam format and rules, and explain how this fits with your philosophy of testing.
4. Give students tips on how to study for and take the exam – this is not a test of their test-taking ability, but rather of their knowledge, so help them learn how to take tests.
5. Have extra office hours and a review session before the test.
6. Arrive at the exam site early, and be there yourself (rather than sending a proxy) to communicate the importance of the event.

Discussing the basic principles related to the administration of classroom or teacher-made tests, Tyler (1991) pointed out that certain rules have been useful in actual practice in the administration of any testing programme and for that matter, could be used for administering any classroom or teacher-made test. These are:

1. Select the tests carefully, preferably in co-operation with a college assessment committee.
2. Order the tests well in advance of the date on which they are to be used.

3. Plan in detail for the administration of the tests. Choose examiners and proctors with great care.
4. Avoid overemphasis on the tests.
5. Seat the students in alternate chairs, if possible.
6. Make announcements slowly and clearly in a voice that is loud enough to be heard throughout the room.

7. Have the blanks on the front of the booklets filled out. Be sure to announce the date, how names are to be written, and other items that may need clarification.

8. Time the examination with extreme care, using a watch or a wall clock

9. Move about the room occasionally to see that students are working on the right part of the test, but do not stand gazing over a student's shoulder and do not constantly move nervously from student to student.

10. Stop the test immediately when the time is up and collect the answer booklets.

11. As soon as a certain test has been given, have all the students taking the test turn in their booklets promptly. Alphabetize and check the test papers against the class list.

12. Except in cases of illness, see that all absentees make up the test. This necessity will cause the tutor much trouble and worry, but it is unavoidable, for complete data are essential if the results of the test are to be used successfully in either teaching or guidance.

According to Mehrens and Lehmann (1991), during administration of test items, the following conditions must be established.

1. The physical conditions should be as comfortable as possible.
2. Test should not be given when examinees are doing pleasant things such lunch, recess.
3. Test should not be given immediately before or after long vacation or a holiday.
4. Teachers should make sure that students understand the directions and answer sheets during test takings.
5. Students should be monitored to minimise cheating.

The Challenges of Test Administration

Every event or activity is in one way or the other characterized by a challenge. Test administration as an activity is no exception. Sarita (2005) and Tyler (1999) identified the challenges and current problems in test administration in Ethiopia as follows: Large class size, poor resources and facilities, insufficient expansion of staffing, forged documentation, mark and grade inflation, cheating on tests, lack of academic honesty etc.

In Ghana, test administration is also bedeviled with a lot of challenges. These challenges are not quite distinct from the ones indicated by (Anamuah-Mensah & Quagrain, 1998). Poor resources and facilities create a major challenge for test administration. The large number of students in a particular class also create problems for assessors (Tyler, 1999). Another dominant challenge which is cheating cannot be left out. Most students do not prepare adequately before taking tests hence indulging in malpractices such as cheating in test (Etsey, 2004).

Validity

Validity is “the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses” of a test (AERA, APA, & NCME, 1999, p. 9). Validity according to Nitko (2001), is the soundness of one’s interpretation and uses of students’ assessment results.

This means that for teachers in the universities to produce valid results of their students, the students’ results must be supported with series of evidence.

The issue of validity concerns the interpretations and uses of measurement scores. The interpretations and uses of one’s assessment results are said to be valid only when the values implied by them are appropriate.

Essentially, the interpretation and uses one make of one’s assessment results are also valid when the consequences of these interpretations and uses are consistent with appropriate values. Here, when the values of the assessment are not in accordance with the consequence of the assessment then this principle is violated (Nitko, 2001).

A second important implication of the definition of validity is that validity is a matter of degree, it is not an “all-or-none” issue. That is, the validity of a test interpretation should be conceived in terms of strong versus weak instead of simply valid or invalid (Yeboah, 2017). For test users, validity should be a deciding factor in their choice of psychological tests. Although such choices are based on a number of practical, theoretical and psychometric factors, a test should be selected only if there is strong enough evidence supporting the intended interpretation and use.

A third important facet of validity is that, the validity of a test’s interpretation is based on evidence and theory. For a test user to be confident

in an interpretation and use of test scores, there must be empirical evidence supporting the interpretation and use. In addition, contemporary views on validity emphasize the importance of grounding the interpretation and use of a test in a defensible psychological theory.

Kinds of validity evidence

Specific forms of validity evidence include content-related evidence, criterion-related evidence and construct-related evidence. Under content-oriented evidence, when the rationale for the test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content must be described (Moss, 1992; Yeboah, 2017). It must also justify with reference to the intended population to be tested and the construct the test is intended to measure. Content-related evidence is also the extent to which the sample of items, tasks, or questions on a test is representative of the domain of content (Moss, 1992).

William (2001) argues that "content-related validity should be concerned not just with test questions, but also with the answers elicited, and the relationship between them" (p. 4). Here, William is advocating for content-related evidence to be extended to include the behaviour elicited actually corresponding to the intentions of the assessment task. He explains with an example. A test claiming to assess students' understanding of forces "would be invalidated if it turned out that the reading requirements of the test were so demanding that students with poor reading ability, but a sound understanding of forces, obtained low marks" (p. 4). Content-related evidence is, therefore, not only demonstrated by the degree to which samples of assessment tasks are representative of some domain of content. It is important for the behaviour

elicited by the test item not to have been influenced by factors that conceal the true ability or potential of the student. This could be an argument in support of school-based teacher assessment as the conditions of assessment can be arranged to provide ecological validity (Crooks, 2001). As Crooks (2001) pointed out, "the circumstances under which student performances are obtained can have major implications for the validity of the interpretations from an assessment" (p. 270). Issues such as low motivation, assessment anxiety, and inappropriate assessment conditions can all be threats to the validity of students' assessment results.

Tamakloe, Amedahe and Attah (2005) indicated that criterion-related validity entails the effectiveness of the test in predicting a student's performance in specified situations. Thus, criterion-related evidence serves as a basis for using the test scores to predict an individual's standing on a criterion measure of interest. This is achieved when the performance of the student on the test is checked against a standard measure, which is a direct and independent measure of the specific behaviour which the test is designed to predict. In other words, a test's criterion related validity is measured by how well it relates with some accepted criteria of the behaviour being measured. The literature distinguishes between two types of criterion-related validity evidence. These focus on whether the external criterion is obtained at the time of the administration of the test, or at some time in the future. Kinyua and Okunya (2014) described the two main types of criterion-related evidence (concurrent and predictive validity evidence) as follows:

Concurrent validity evidence refers to the extent to which an individual's current status on a criterion can be predicted from their prior

performance on an assessment instrument (Nitko & Brookhart, 2007). For example, a State of Anxiety Scale can be administered at the same time as the older and much more established Taylor Manifest Anxiety Scale so that if the results of the former are comparable to the later, the former test will have passed criterion validity test. Predictive validity evidence is where the performance of one test is used to predict the potential performance in another test. For example, the performance of a student in the West Africa Senior School Certificate Examination (WASSCE) used in predicting the Grade Point Average in the University during the first semester.

According to Deville (1991), “construct validity-evidence is directly concerned with the theoretical relationship of a variable (e.g. a score on some scale) to other variables. It is the extent to which a measure ‘behaves’ the way that the construct it purports to measure should behave with regard to established measures of other constructs” (p. 46). Also, Messick (1989) defines construct validity-evidence as an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.

The operationalizing of the construct involves developing a series of measurable behaviours or attributes that are posited to correspond to the latent construct. Defining the construct syntactically involves establishing assumed relationships between the construct of interest and other related constructs or behaviours (Benson, 1998; Crocker & Algina, 1986; Gregory, 1992).

Factors affecting validity

A number of factors affect the degree of validity of assessment results. These factors tend to reduce the degree of validity of the results affecting its use and appropriate interpretations. Amedahe and Asamoah-Gyimah (2013, p. 59) outlined some of these factors as:

1. Unclear directions: To help the student to respond meaningfully to test items, there is always the need to provide clear directions to them. Directions that do not clearly indicate to the testee how to respond to the tasks and how to record the responses will tend to reduce the validity of the results. This is because students may get confused over how to respond and how to record their responses that may in turn affect their performance.
2. Reading vocabulary and sentence structure: The sentence structure should not be too complex for the level of the students. When the vocabulary and sentences are too difficult and complicated for the students taking the assessment, it will result in the assessment measuring the student's reading comprehension ability rather than the student's achievement in a subject matter content. In this case, the use and interpretation of the test scores may have low validity.
3. Ambiguity of items: When test items are ambiguous, they can be interpreted in different ways. This can result in misinterpretation and confusion. This will intend reduce the validity of the students' results.
4. Inadequate time limits: Students need to be given adequate time to complete a test. Too short duration to complete tests may deprive testees the opportunity to reason and respond accordingly whiles too

long duration may also cause testees to finish very early and misbehave by sometimes cheating or changing the right answer and which therefore introduces a lot of biases to their results.

5. Poor construction of items: Poor construction can take the form of the items providing clues to the answer. In the selection type of test items

the clues may be provided by certain word(s) in the stem of the item (e.g, is/an) or grammatically inconsistent options. Poorly constructed items makes the tester deviate from the purpose of testing and this therefore reduce the validity of the results.

6. Nature of the group: Things being equal, the more heterogeneous the group, the higher the content-related validity coefficient.

7. The testing environment: It is another variable associated with the validity of teacher-made tests. If the testing environment is distracting or noisy or the test-taker is unhealthy, he or she will have a difficult time remaining consistent throughout the testing process (Griswold, 1990). Even though actions ought to be taken to ensure that the testing environment is comfortable, adequately lit with limited interruptions (Griswold, 1990), these factors are largely aspects of test administrative procedures that are external to the test itself. This is because even in contexts where the characteristics of the test taker and the environment are well taken care of, it emerges that individual difference in performance will still be recorded.

8. Length of test: Long tests do three things to help maintain validity. Firstly, they increase the amount of content that the student must address, ensuring a more accurate picture of student knowledge.

Secondly, long tests counteract the effects of faulty items by providing a greater number of better items. Third, long tests reduce the impact of student guessing.

Reliability

Reliability refers to the consistency of the scores obtained. That is, how consistent the scores are for each individual from one administration of an instrument to another and from one item to another. Reliability is a measure of how stable, dependable, trustworthy and consistent a test is in measuring the same thing each time (Worthen et al., 1993). Reliability is defined as the degree of consistency between two measures of the same thing (Hussain, Jamil, Siraji & Maroof, 2012). It is the degree to which assessment results would be similar under slightly different measurement conditions. For instance, if one assesses a student on a lesson taught, one hopes that, one would obtain almost the same score if one assesses the student one day later. Here, if one measures a person's level of achievement, one hopes that the scores will be similar under different administrators, using different scorers, with similar but not identical items.

Methods of estimating reliability

The methods used to estimate reliability differ in that they consider different sources of error (Liaquat, Asif, Siraji & Maroof, 2012). Many different approaches are used to estimate reliability, but the more common ones are:

1. Test-rest method
2. Split-half
3. Equivalent forms method

4. Measures of internal consistency (Liaquat, Asif, Siraji & Maroof, 2012, 71)

The test-retest method is a measure of stability and considers scores of students over a period of time. The same test is given to a group of students twice within an interval ranging from several minutes to years. The scores on the two administrations are correlated and the result is the estimate of the reliability of the test (Etsey, 2012). Here, there are various possible time intervals. The estimate of reliability will vary with the length of the interval, and therefore this interval length must be considered in interpreting reliability coefficients. Therefore, when stability reliability is reported in a test manual, the time interval between testing should always be specified, as well as some indication of the relevant intervening experiences. Here, any change in score from one setting to the other is treated as error. This presents the true performance of the student in question. Therefore, appropriate decisions can be taken for the student based on his/her performance. According to Webb (2002), stability estimate is often difficult to obtain and interpret in psychological measurements. That is, the very act of measurement causes the person to change on the variable being measured. The practice effects from the first testing, for instance, will probably be different across students, therefore, lowering the reliability estimate. In converse, if the interval is short, there may be a strong recall or memory effect. The equivalent-forms estimate of reliability is obtained by giving two forms (with equal content, means, and variances) of a test to the same group on the same day and correlating these results. With this procedure, one determines how confident one can generalize a person's score to what he would receive if he took a test composed of similar

but different questions. Again, any change in performance is considered error. In this case, instead of measuring changes from one time to another, one measures changes due to the specificity of knowledge. In the construction of equivalent forms of a test, the items should be of the same difficulty and of the same format (for instance, multiple choice, true or false and others).

Administrative instructions are expected to be the same for both tests (Liaquat, Asif, Siraji & Maroof, 2012). Equivalent form measure is used when ones' purpose is to make inferences about the knowledge one has in the subject-matter area (Liaquat, Asif, Siraji & Maroof, 2012). With this, one is less interested in how stable the knowledge is over time but more interested in whether one can infer or generalize to a larger domain of knowledge from a sample.

In a split-half method a single test is given to the students. The test is then divided into two halves for scoring. The two scores for each student are correlated to obtain the estimate of reliability (Liaquat, Asif, Siraji & Maroof, 2012). The test can be split into two halves in several ways. These include using odd-even numbered items, and also first and half-second half. To estimate what the reliability of the whole test would be, a correction factor is applied. The split-half method also gives the true performance of the student in the sense that, the student is assessed twice with the one assessment task at the same time.

The advantage of the split-half method is that only one form of the test needs to be administered. However, the Spearman-Brown prophecy formula assumes that the variances of the two halves are equal. When they are not, the estimated reliability of the whole test will be greater than that obtained by

other methods of internal consistency (Nitko, 2001). Therefore, one of the problems with the Split-half method is how to make the split and it can be surmounted by one making efforts necessary to construct two equivalent forms (except that only half as many items are needed).

The Cronbach alpha is another method of estimating internal consistency. Cronbach alpha is the average split-half correlation based on all possible divisions of a test into two parts. This internal consistency is used when test items are scored pass-fail or when more than one point is awarded for a correct response (Salvia & Yesseldyke, 2001). Kuder and Richardson KR-20 and 21 are more restricted method of estimating a test's reliability which is based on the average correlation between all possible split halves. Under KR 21, test items are scored dichotomously (that is, items that can be scored only right and wrong). Nitko (2001) indicated that internal consistency founded on the idea that the consistency with which students respond from one assessment task to the next provide the basis for estimating the reliability coefficient for the total scores.

Factors affecting reliability

Low internal consistency estimates are often the result of poorly written items or an extremely broad content area of measure (Crocker & Algina, 2008). However, other factors can equally reduce the reliability coefficient, namely the homogeneity of the testing sample, imposed time limits in the testing situation, item difficulty and the length of the testing instrument (Crocker & Algina, 2008; Mehrens & Lehman, 1991; DeVellis, 1991; Gregory, 1992). Group homogeneity is particularly influential when an assessor tries to apply a norm referenced test to a homogenous testing sample.

In such circumstances, the restriction of the range of the testing group translates into a smaller proportion of variance explained by the testing instrument, thus ultimately reducing the reliability coefficient (Crocker & Algina, 2008). It is essential to bear in mind the intended use of the instrument when considering these circumstances and deciding how to use an instrument (Crocker & Algina, 2008; Mehrens & Lehman, 1991; Gregory, 1992).

Imposed time constraints in a testing situation pose a different type of problem. That is, time limits ultimately affect a test taker's ability to fully answer questions or to complete an instrument. Clearly, this situation becomes problematic when the construct that an instrument intends to measure has nothing to do with speed competency (Crocker & Algina, 2008; Mehrens & Lehman, 1991; Gregory, 1992). The relationship between reliability and item difficulty addresses a variability issue once again. Eventually, if the testing instrument has little to no variability in the questions (i.e., either items are all too difficult or too easy), the reliability of scores will be affected (Crocker & Algina, 2008). Aside from this, reliability estimates can also be artificially deflated if the test has too many difficult items, as they promote uneducated guesses (Crocker & Algina, 2008; Mehrens & Lehman, 1991). Lastly, test length also affects the reliability estimate. Simply, longer tests yield higher estimates of reliability (Crocker & Algina, 2008).

The test length phenomenon can be best explained through an examination of the Spearman Brown prophecy equation, which indicates that as the number of items increase, there is a direct increase in the reliability estimate (Crocker & Algina, 2008). However, one must consider the reliability gains earned in such situations, as extremely long tests are not necessarily

desirable. For instance, if an assessor has an 80-item testing instrument with an internal consistency reliability coefficient of .78 and the Spearman-Brown prophecy indicates that one's reliability estimate will increase to .85 if assessor adds 25 items, one may consider that the slightly lower reliability estimate is more desirable than an excessively long instrument (Crocket &

Algina, 2008; Mehrens & Lehman, 1991; Gregory, 1992)

Empirical Review

Test construction

Studies in the teacher testing practices have been done over the years in Ghana and elsewhere. Amedahe (1989) conducted a research on the testing practices of secondary schools in the Central region of Ghana. The study adopted the descriptive sample survey design for its execution. The population of 213 for the study was made up of teachers in 18 secondary schools in the Central region of Ghana teaching English Language, History and Mathematics in Forms One, Three and Five. The study found out that:

1. Teacher's knowledge of concepts and terminologies used in measuring students' performance was weak.
2. Teachers did not follow the basic prescribed principles of classroom test construction.

Another study was conducted by Quagrain (1992) on teachers' competencies in the use of essay tests in the Ghana. The study was carried out in a sample of secondary schools in the Western region and the findings and summaries are as follows:

1. While some teachers reviewed their essays test items, others did not.
2. Majority of the teachers prepared their marking scheme after

examination while only a few prepared the scheme before the test is taken.

The study revealed that while some teachers followed prescribed practices such as teachers reviewing test items after construction, others did not. The study however, did not probe the reason why some of the teachers in the study put up practices contrary to those that have been advocated by testing experts. For instance, why does majority of the teachers in the schools ignore indicating the score point for each item on the question paper to guide students?

Etsey (2006) conducted a study to determine primary school teachers' standard of test construction and administration in Ghana and also whether length of service had any effect on the practice of assessment. Three hundred and forty-six primary school teachers, consisting of 174 females and 167 male teachers from two regions in Ghana participated in the study. The basic design used for the study was the cross-sectional descriptive survey and data was collected using a questionnaire.

The result showed that primary school teachers standard of test construction and administration was moderate. The main areas of strength in test construction and administration were (1) considering the purpose of the test before writing the test items (2) evaluating the test as a whole before submission, (3) preparing the scoring key in advance of testing data and (4) matching instructional objectives with the test formats. Major areas of weakness included, nonuse of test specification table, and not preparing more items than needed in the examinations to allow for review of assessment tasks. It was also found out that length of service did not have an effect on the

practice of test construction and administration by the teachers in the primary school in Ghana.

The result also showed that primary school teachers had very little knowledge of concepts used in assessment such as test specification table, taxonomy of educational objectives, item difficulty and item analysis. The above result presents just a little picture of the state of affairs in the primary schools. Another study that was conducted in the area of testing practices was by Magno (2003). This research determined the profile of the professors' level of appropriateness in test construction in University of Perpetual Help, Laguna in the Philippines. It was noted that the practical rules of test item construction is formulated on the basis of years of experience in preparing items and empirical evaluation of responses, i.e. the level of appropriateness in test construction correlated with the actual years of experience in teaching was investigated.

There were 33 college professors who were surveyed in this study. The survey determined their tendency to follow general principles, guidelines and procedure in test construction. The results indicated that majority of the professors (54.54%) had an average level of appropriateness in test construction with a mean value of 24.76. The results also revealed that there is no significant relationship between the level of appropriateness in test construction and the actual years of teaching experience. This is due to the adequate training and exposure that new professors had in test construction.

Another study was conducted by Adam, Beyuo and Bediako (2007) to assess the Polytechnics lecturers' competencies and skills in test construction, administration and scoring. The study was conducted in the Western region of

Ghana using the Takoradi Polytechnic as a case study. The study found out that 80.6% of the lecturers had taken a course in Measurement and Evaluation. It was shown that majority of the lecturers in the Polytechnic were competent in test items construction. The analytic method of scoring was also the most dominant scoring method by the lecturers.

The response rate for the data collection was very low. Out of an expected coverage of seventy-two (72) lecturers, only forty-three percent (43%) response rate was recorded and therefore could not have been seen as a true reflection of practices in the Polytechnic of study. As relevant and heartwarming that the study revealed, just one (1) Polytechnic out of a total ten (10) in Ghana used for the study is not enough and therefore should not be used as a true representation of the situation in all the Polytechnics in Ghana.

Okyireh (2008) conducted a study on the testing practices of senior secondary school teachers in the Ashanti region of Ghana aimed at finding out whether pre-service training in testing contributes to competence in actual testing practices and also whether teachers followed the basic principles in their testing practices.

Cluster and simple random sampling techniques were adopted to select 265 teachers of Mathematics, Science and English Language in 26 Secondary schools for the study. The main instrument used for the data collection was the questionnaire.

The study showed that, to a great extent, teachers followed the basic principles in test construction, administration and scoring. Teachers were also reported from the study to be using both the norm-referenced and criterion-referenced approaches in their test-score interpretation. The study also

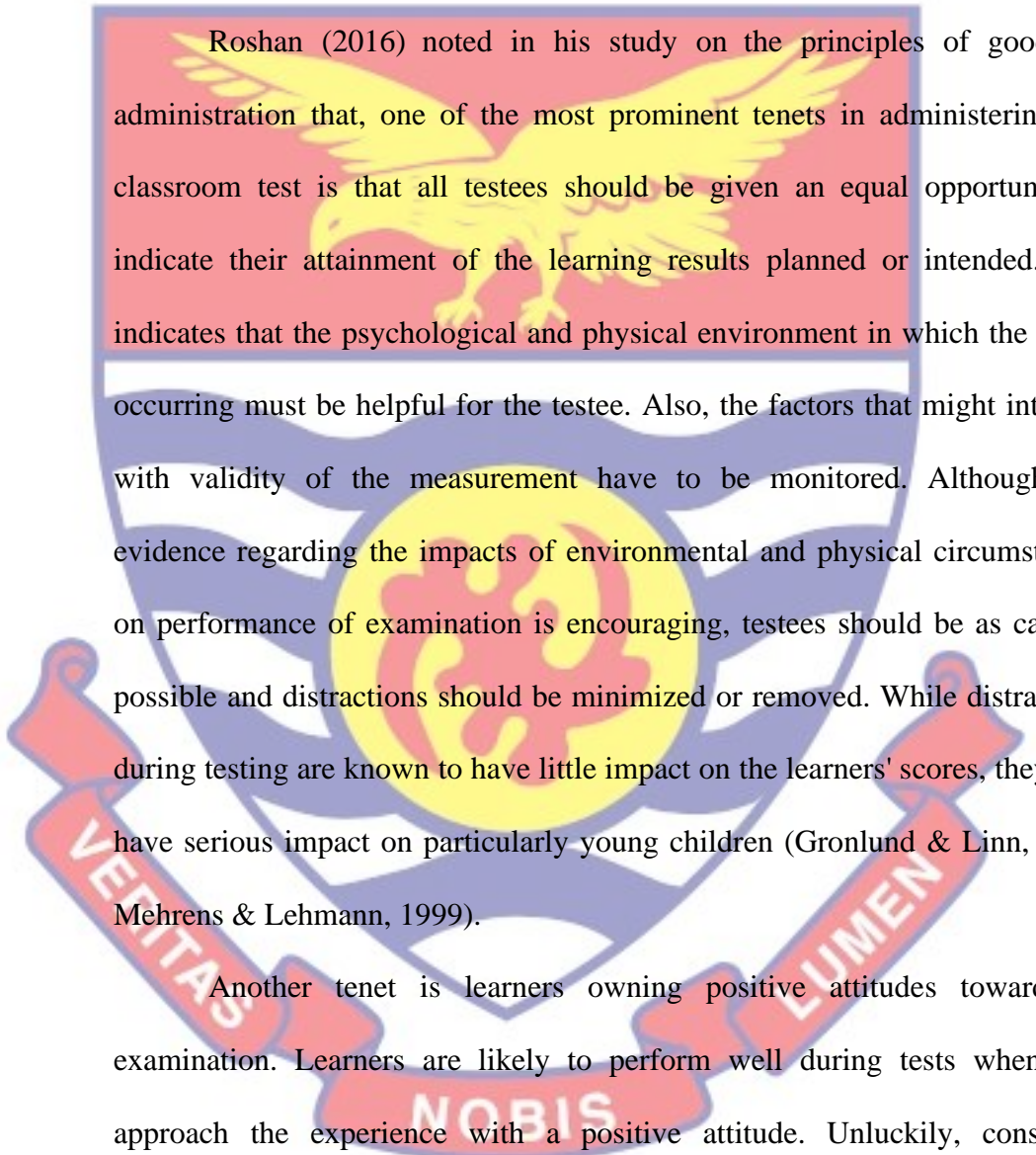
revealed that pre-service instruction in educational measurement had a positive impact on actual testing practices although the impact was quite subtle. Baker (2003) in Uganda also found in his study that teachers were aware of a variety of test construction skills.

Test administration

Oduro-Okyireh (2008) found in his study that teachers did not observe good psychological conditions when testing their students. These psychological conditions include position of the invigilator, timing of the test, threatening behaviours of invigilators, and interruption to give instructions and announcements (Etsey, 2004). This is actually a disturbing phenomenon since psychometricians such as Nunnally (1972) and Gronlund (1985) have asserted that poor psychological conditions such as asking students to hurry up in order to complete the test on time and other threatening behaviours of invigilators affect students performance in one way or the other, negatively. Oduro-Okyireh's findings were in line with findings of Tamakloe et al. (1996) and Etsey (2004) who noted in their study that test administration guidelines include the intent of providing examinees with a fair chance to demonstrate their ability on what is being measured.

There is, one issue of teachers, generally, not using the “Do Not Disturb Examinations in Progress” sign when students are taking examinations to ensure total silence in the vicinity. From Okyireh's point of view, this situation did not have any adverse effect on the conduction of examinations in the schools. This is because in the 20 observations carried out on the conditions under which students take their examinations, there were only four cases of intermittent noisy environments. Finally, the findings here did not

wholly confirm the finding of Amedahe (1989) that teachers generally observed good physical and psychological conditions when administering classroom achievement tests. This is because this study clearly indicates that teachers observed good physical conditions but did not observe a considerable number of good psychological conditions.



Roshan (2016) noted in his study on the principles of good test administration that, one of the most prominent tenets in administering any classroom test is that all testees should be given an equal opportunity to indicate their attainment of the learning results planned or intended. This indicates that the psychological and physical environment in which the test is occurring must be helpful for the testee. Also, the factors that might interfere with validity of the measurement have to be monitored. Although, the evidence regarding the impacts of environmental and physical circumstances on performance of examination is encouraging, testees should be as calm as possible and distractions should be minimized or removed. While distractions during testing are known to have little impact on the learners' scores, they may have serious impact on particularly young children (Gronlund & Linn, 1990; Mehrens & Lehmann, 1999).

Another tenet is learners owning positive attitudes towards an examination. Learners are likely to perform well during tests when they approach the experience with a positive attitude. Unluckily, constantly teachers fail to assist learners to develop positive attitudes toward examinations. Therefore, test anxiety among some learners is heightened in many instances during test taking (Mehrens & Lehmann, 1999; Linn & Miller, 2005).

Another study conducted by Rukundo and Magambo (2010) on the effective test administration in schools and specifically focused on principals and good practices for test administrators in Uganda revealed the following as test administration principles:

1. Examinees and parents had to be notified regarding the test date and time.
2. Candidates had to be reminded to bring materials necessary for the test.
3. All students with special needs (e.g. glasses and hearing aids) had to be considered before the start of the test.
4. All adequate invigilation had to be planned.
5. Examination administrators had to read appropriate test administration procedures such as timing, examination regulations and test modifications.
6. The rooms where the test would be conducted had to have adequate ventilation and lighting and have been properly arranged.
7. Seats had to be arranged in such a way that candidates cannot look at each other's work.

Summary of Literature Review

From the literature review, it was observed that most teachers-made tests contained many faults and most, function exclusively at the recall level. On item type, lecturers use all major item formats. Lecturers typically did not use test improvement strategies such as test blueprints and item analysis. In Ghana, the studies reviewed have shown that to a great extent, lecturers generally, did not follow the basic prescribed principles in test construction. Also, regarding test administration, it was observed that lecturers did not

practice good test administration principles. Furthermore, psychological conditions such as position of the invigilator, timing of the test, threatening behaviours of invigilators, and interruption to give instructions and announcements were not generally upheld.



CHAPTER THREE

RESEARCH METHODS

This chapter discusses the methodology that was adopted in carrying out the study. The methods and approaches as described in this chapter are under eight sub-sections. These are the research design, population, sample and sampling procedure, research instruments, pilot-testing procedure, validity and reliability of the instruments, data collection procedure and data analyses.

Research Design

The descriptive survey design was used for the study. This enabled me to collect information on the current state of lecturers' awareness of the principles of test construction and administration in the University of Cape Coast. The descriptive survey design seeks to explore and describe events as they are.

Descriptive design helped me to explore from lecturers their awareness level of principles of test construction, administration and the challenges they faced in achievement test construction and administration. I also elicited responses regarding the application of test administration principles by the Examinations Unit in the University of Cape Coast. The descriptive research design was deemed best for the study because, according to Cohen, Morrison and Manion (2004), in descriptive survey design, researchers gather data at a particular point in time with the intention of describing the nature of existing conditions or identifying standards against which existing conditions can be compared. Also, the descriptive survey helps to deal, essentially, with questions concerning what exists with respect to variables or prevailing conditions in a situation (Ary, Jacobs & Razavieh, 1990). As recommended by

Leedy and Omrod (2010), this method is suitable for purposes of making generalisations from a sample to a population so that inferences could be made about the characteristics, opinions, attitudes and past experiences of the population. Descriptive survey design provides a more accurate and meaningful picture of an event or phenomenon and seeks to explain people's

perception and behaviour on the basis of data gathered at a particular time (Frankel & Wallen, 1993). Comparatively, other research designs such as case study, experimental and historical surveys would not have been appropriate to use for this study looking at the stated objectives and the time period for the study (Babbie, 2005). I wanted to know from lecturers their awareness and how they use principles of construction and administration. So, the design was appropriate for me to elicit information from lecturers about test construction and administration.

But the design is not totally devoid of bias. The design makes use of deductive reasoning and sometimes getting a sufficient number of questionnaire which will be used is problematic. Confidentiality is also a primary weakness of the design. Often subjects are not truthful as they feel the need to skew their responses to a desired result of the study. Osuola (2001) in buttressing the points on the weaknesses of the descriptive research, pointed out that, "designing a quality investigation requires particular attention to two central factors: appropriate sampling procedures, and precision in defining terms in eliciting information" (p. 201). He further stated that, while descriptive research is a prerequisite for finding answers to questions, it is not in itself sufficiently comprehensive in providing answers and that it cannot also provide cause-and-effect relationships.

Notwithstanding these challenges, the design remains the best and most appropriate for this study.

Study Area

The University of Cape Coast is a prestigious public research university located in Cape Coast, Ghana. The university, which is five kilometers west of Cape Coast town, is on a hill overlooking the Atlantic Ocean. It operates on two campuses: the Southern Campus (Old Site) and the Northern Campus (New Site). Two of the most important historical sites in Ghana, Elmina and Cape Coast Castle, are only a few kilometers from the university.

The university was established in 1962 out of a dire need for highly qualified and skilled manpower in education. It was established to train graduate teachers for second cycle institutions such as teacher training colleges and technical institutions, a mission that the two existing public universities at the time were unequipped to fulfill. The university has since added to its functions the training of doctors and health care professionals, as well as education planners, administrators, and agriculturalists.

Population

Population refers to the large general group of many cases from which a researcher draws a sample and which is usually stated in theoretical terms (Neuman, 2003). Polit and Hungler (1998) defined a study population as that which reflects the entire aggregate of cases that meet designated set of criteria. According to Amedahe (2004), the target group about which a researcher is interested in gaining information and drawing conclusions is what is known as the population. It is a group of individuals who have one or more

characteristics in common that are of interest to the researcher. The population of the study comprised the four colleges and the staff of the Examination Unit in the University of Cape Coast. The choice of population is purposive. The purposive technique otherwise known as the judgmental technique involves choosing subjects who in the opinion of the researcher are thought to be relevant to the research issue.

According to Amedahe (2000), target population refers to the population that the researcher will ideally like to generalise. The target population of the study comprised of five (5) Colleges and the Examination Unit in the University of Cape Coast. But the accessible population comprised of College of Agric and Natural Sciences, College of Health and Allied Sciences, College of Humanities and Legal Studies and College of Education Studies. The total number of schools/faculties in the various colleges at the time of the study was fourteen (14). Within the four colleges, I targeted lecturers for the study and their total number was 134. The staff of Examination Unit who were fifty-two (52) were also targeted for the study.

Sampling Procedure

A sample denotes a small and representative proportion of the population. Sampling enables the researcher to study a relatively small number of units in place of the target population and to obtain data that is representative of the whole population (Sarantakos, 1998). Sampling involves the process of selecting a portion of the population to represent the entire population (Amedahe, 2000). Table 1 presents the colleges and school/faculties sampled for the study.

Table 1: Colleges and Schools/Faculties under Them

College	Schools/Faculties
College of Agric and Natural Sciences	School of Biological Sciences School of Physical Sciences School of Agricultural Sciences
College of Health and Allied Sciences	School of Medical Sciences School of Nursing and Midwifery School of Allied Health Sciences
College of Humanities and Legal Studies	Faculty of Arts Faculty of Social Sciences School of Business School of Law
College of Education Studies	Faculty of Science and Technology Education Faculty of Educational Foundations Faculty of Humanities and Arts Education School of Educational Development and Outreach
College of Distance Education	College of Distance Education

Source: Field Survey, Armah (2017)

Due to the nature of the population of the study, multi-stage sampling strategy was used to select the sample. Multi-stage sampling is a strategy used when conducting studies involving a very large population. This is when the entire population is divided into naturally-occurring clusters and sub-clusters, from which the researcher randomly selects the sample (Adane, 2013).

In the first stage, I employed the simple random sampling technique to select four (4) colleges out of the five (5) colleges in the University of Cape Coast. During the second stage, I employed simple random sampling technique (tables of random numbers method) to select schools/faculties for the study. Simple random (table of random numbers method) was used to select one school/faculty from each of the selected four colleges. The

schools/faculties selected for the study were (a) School of Biological Sciences, (b) School of Medical Sciences, (c) Faculty of Social Sciences and (d) Faculty of Educational Foundations. According to Amedahe (2002), simple random sampling technique is the process in which each element in the population has an equal, independent chance of being selected. This was done by using the table of random numbers. The sampling frame was first identified or constructed. After this, appropriate columns from the table of random numbers were selected. I then picked from the columns randomly and registered the name in the sampling frame corresponding to the numbers constituting the sample.

The third stage, after a school/faculty was randomly selected from each of the selected colleges, I then employed census procedure to select the participants of the study. The census procedure was used to obtain 134 lecturers for the study. According to Cooper and Schindler (2000), census involves the use of all members in any population of interest. They stated that a “census is feasible when the population is small” (p. 164). The simple random sampling procedure was used to select 40 Examination Unit staff out of 52 for the study. Participants here staff of Examination Unit and invigilation assistants from other departments in the University of Cape Coast. The sampled Faculties and Schools is presented in Table 2.

Table 2: Distribution of Sampled Faculties/Schools

Schools/Faculties	Census
School of Biological Sciences	29
School of Medical Sciences	32
Faculty of Social Sciences	52
Faculty of Educational Foundations	21
Examination Unit	40
Total	174

Source: Field Survey, Armah (2017)

Data Collection Instruments

The main instrument that was used to collect data for the study was the questionnaire. Two sets of questionnaires were used for the study. One was used for lecturers and the other for Examinations Unit in the University of Cape Coast. The instruments were self-designed based on the research questions. The choice of questionnaire was based on the assertion of Cohen, Manion and Morrison (2004), that, it is widely used and is also a useful instrument for collecting survey information, providing structured, numerical data and being able to be administered without the presence of the researcher. Shadish, Cook and Campbell (2002) also maintains that the purpose of the survey questionnaire is to elicit information about the characteristics or opinions of the respondents. The questionnaire has the advantages of allowing the researcher to collect data from a group of respondents at the same time and it is easy to score. The choice of questionnaire was based on the assertion of Shadish, et al. (2002) that, they are particularly advantageous whenever the sample size is large enough to make it uneconomical for reasons of time or

funds to observe or interview every participant. Further, questionnaire is easy to administer, friendly to complete and fast to score and therefore take relatively less time from researchers and respondents (Knowles, 2000).

Despite all these advantages, questionnaires are not able to bring in-depth knowledge of respondents about an issue as compared to instruments like interviews and others. The items on questionnaire might not have the same meaning to all respondents and this could affect the responses of respondents. Also, I was unable to probe for additional details from the respondents, using a questionnaire (Gay, Mills & Airasian, 2009; McMillan & Schumacher, 2001). The questionnaire comprised close ended items only. The questionnaire for lecturers comprised seven (7) sections; Section A, B, C, D, E, F and G (See Appendix A). The first section (Section A, items 1 to 3) centered on demographic data of the respondents. The second section, (Section B, items 4 to 14) centered on lecturers' awareness of test construction principles. The third section, (Section C, items 14 to 26) centered on lecturers' awareness of test administration principles. The fourth section (Section D, items 27 to 38) centered on lecturers use of test construction principles. The fifth section (Section E, items 39 to 50) focused on lecturers' use of test administration principles. The sixth section (Section F, items 51 to 60) covered challenges in test construction and lastly, section seven (Section G, items 61 to 67) covered challenges in test administration (See Appendix A).

The questionnaire for Examinations Unit had two sections. The first section (Section A, items 1 to 2) covered the demographic data of respondents and the second section (Section B, 3 to 14) covered the extent to which

Examinations Unit in the University of Cape Coast applied test administration principles (See Appendix B).

Items on the questionnaire were multiple-scored on a four-point Likert type scale. The items on the Likert type scale were scored ranging from one (1) = strongly disagree, (2) = disagree, (3) = agree and (4) = strongly agree.

The Likert type scale was chosen because according to Asamoah-Gyimah (2002), in measuring the views and impressions of teachers on an on-going practice, it is the simplest, but equally efficient approach when considered alongside social-distance scales, Thurstone scales and the scalogram analysis. It was adopted also to ensure effective analysis of the data even though it restricts free expression and perception of respondents in a study.

Validity of the Instrument

In order to enhance the validity of the study, the questionnaire was given to my supervisors in the Department of Education and Psychology in the University of Cape Coast (UCC) for expert judgement and assessment. This was to ensure content related evidence to the items. The items were examined to see whether they were related to the research questions and also comprehensively cover the details of the study. Based on their comments and suggestions, the questionnaires were fine-tuned to achieve the purpose of the study.

Pre-testing of the Instrument

A pre-test of the instrument was conducted to ascertain any need for revisions. Pre-testing of the instrument was carried out at University of Education, Winneba. The aim of the pre-testing was to improve the validity and reliability of the instruments. The participants (teachers) of the pre-test

were asked to complete the questionnaire and to provide comments or suggestions for revising any ambiguous items. They were also told to discuss openly with me any ambiguity, incoherence or incomprehension that they experienced about any aspect of the draft questionnaire. The final instrument for the study was produced after subsequent revisions in the wording of a few

items. For example, items 32 and 43 which read “I copy verbatim from textbooks” and “I inform students of test type” were reworded as “I copy directly from textbooks” and “I inform students of test format”. The necessary corrections were effected after the pre-testing.

Reliability of Instrument

The reliability (internal consistency) of the items for the main study was estimated using Cronbach’s co-efficient alpha. According to Cronbach (as cited in Ebel & Frisbie, 1991), co-efficient alpha can provide a reliability estimate for a measure composed of items of varying point values such as essays or attitude scales that provide responses such as strongly agree and strongly disagree with intermediate response options. To obtain the reliability of the instrument, Cronbach’s co-efficient alpha was used to estimate the internal consistency. Table 3 provides the summary of the reliability coefficient obtained for each of the study variables.

Table 3 : Summary of the Reliability Coefficient of the Items

Item Sections	R. Coefficient	No. of Items
Lecturers’ awareness of test construction principles	.87	11
Lecturers’ awareness of test administration principles	.75	12
Lecturers’ use of test construction principles	.88	12
Lecturers’ use of test administration principles	.87	12
Challenges in test construction	.91	10
Challenges of the test administration	.80	7
Examinations Unit use of test administration principles	.72	12

Source: Field survey, Armah (2017)

According to Fraenkel and Wallen (2000), the reliability coefficient should be at least 0.70 and preferably higher. This means that the reliability indexes obtained for the various sections of the instrument are appropriate for the study. The overall reliability index of the two questionnaires are 0.76 and 0.71 respectively. Therefore, the reliability obtained is justifiable for the study.

Ethical Consideration

Ethical clearance form was taken from Institutional Review Board in the University of Cape Coast. The form spelt out the purpose of the study, the need for individual participation, anonymity as well as confidentiality of respondents’ responses. Informed consent was sought from participants by explaining the purpose of the study to them.

Anonymity of respondents was highly considered in the study. This gave the participants the opportunity to have their identity concealed. Neither names nor any identifiable information from respondents were taken.

On the issue of confidentiality, effort was made to maintain confidentiality of the responses of the participants. Participants were told that

their responses would be kept confidential and that no one known to them would have access to the information provided.

Data Collection Procedures

I personally administered the questionnaire to all the 134 lecturers involved in the study in the 4 selected Colleges in the University of Cape Coast, Cape Coast. I used a period of three weeks to go to all the sampled Colleges to administer the questionnaire. Respondents were given a period of two weeks to respond to the questionnaire after which I went round again to the Colleges for collection. A total of 131 questionnaires were retrieved which represented 97.7% of the 134 questionnaire distributed.

Two days were used to collect data from the staff of the Examinations Unit in the University of Cape Coast, Cape Coast to elicit responses concerning the principles they followed in test administration. A total of 40 questionnaires which were given to the staff of the Examinations Unit were all retrieved.

Data Processing and Analysis

The responses to the questionnaires were first edited, coded and scored. The editing procedure was to check whether respondents followed directions correctly, and whether all items were responded to. Section A on the questionnaire was on some demographic data of the respondents. These responses were analysed using frequencies and percentages.

Research Question One

What is the extent of lecturer's awareness of the principles of test construction in the University of Cape Coast?

The scale of measurement used was the interval scale. The statistical tools used were frequencies and percentages.

Research Question Two

What is the extent of lecturer's awareness of the principles of test administration in the University of Cape Coast?

The scale of measurement used was the interval scale. The statistical tools used were frequencies and percentages.

Research Question Three

What is the extent to which lecturers apply the principles of test construction in setting classroom test items?

The scale of measurement used was the interval scale. The statistical tools used were the mean, standard deviation and a one-sample t-test.

Research Question Four

What is the extent to which lecturers apply the principles of test administration in setting classroom test items?

The scale of measurement used was the interval scale. The statistical tools used were the mean, standard deviation and a one-sample t-test.

Research Question Five

What is the extent to which the Examinations Unit of the University Cape Coast apply the principles of test administration?

The scale of measurement used was the interval scale. The statistical tools used were the mean, standard deviation and a one-sample t-test.

Research Question Six

What are the major challenges encountered by lecturers in constructing tests in the University of Cape Coast?

The scale of measurement used was the interval scale. The statistical tools used were frequencies and percentages.

Research Question Seven

What are the major challenges encountered by lecturers in test administration in the University of Cape Coast?

The scale of measurement used was the interval scale. The statistical tools used were frequencies and percentages.

Chapter Summary

The study was conducted to assess the extent to which lecturers apply principles of test construction and administration in the University of Cape Coast. The descriptive survey design was used for the study. Study respondents were lecturers and staff of the Examination Unit of the University of Cape Coast, Cape Coast. Data from lecturers' awareness of the principles of test construction and test administration was obtained using questionnaires. Data generated from the questionnaire were coded and analysed descriptively (means and standard deviations and frequencies and percentages). A one sample t-test was also conducted at 0.05 level of significance to ascertain significant level of each of the items. The limitation I encountered in this aspect of the study was unwilling attitudes of respondents towards me during the data collection. The next chapter presents the results and discussion of the study.

CHAPTER FOUR

RESULTS AND DISCUSSION

This chapter deals with the presentation and analysis of the data collected. The purpose of the study was to assess the extent to which lecturers applied principles of test construction and administration in the University of Cape Coast. The study specifically focused on the extent to which (a) lecturers are aware of the principles of test construction, (b) lecturers are aware of the principles of test administration, (c) lecturers apply the principles of test construction in setting classroom test items, (d) lecturers applied the principles of test administration, (e) the Examinations Unit of the University Cape Coast applied the principles of test administration, (f) lectures encountered challenges when constructing tests items and (g) lectures encountered challenges when administering tests items.

The methodology used in this study was the descriptive design. The instrument chosen for the data collection was the questionnaire. The statistical tools used for the analysis included frequency distributions, means, standard deviations and a one sample t-test analysis. The tests were conducted for significant differences at a significance level of 0.05.

Description of Sample

The study was carried out in the University of Cape Coast in the Cape Coast Metropolis, Central region of Ghana, with a sample size of 134 lecturers and 40 Examination Unit workers.

Distribution of lecturers by gender

Table 4 presents the distribution of lecturers by gender

Table 4 : Distribution of Lecturers' by Gender

Gender	Frequency	Percent (%)
Male	69	52.7
Female	62	47.3
Total	131	100.0

Source: Field Survey, Armah (2017)

Table 4 shows that 52.7% of the lecturers were males while 47.3% were females. Therefore, majority of the study participants were males.

Teaching experience of lecturers

Table 5 presents the distribution of teaching experience of lecturers who were involved in the study.

Table 5 : Teaching Experience of Lecturers

Teaching experience	Frequency	Percent (%)
1-5years	41	31.3
6-10years	42	32.1
Above 10years	48	36.6
Total	131	100.0

Source: Field Survey, Armah (2017)

Table 5 shows that 36.6% of the lecturers have taught for above 10 years while 31.3% have also taught for 1-5 years. The study revealed that teaching experience of lecturers involved in the study was evenly distributed.

Schools/Faculties of lecturers

Table 6 presents the schools/faculties distribution of lecturers involved in the study.

Table 6 : Distribution of Lecturers in the Selected Schools/Faculties

Schools/Faculties	Frequency	Percent (%)
School of Biological Sciences	29	22.1
School of Medical Sciences	32	24.5
Faculty of Social Sciences	51	38.9
Faculty of Educational Foundations	19	14.5
Total	131	100.0

Source: Field Survey, Armah (2017)

Table 6 shows that 38.9% of the lecturers were from the Faculty of Social Sciences. This was followed by 24.5% who were in School of Medical Sciences, 22.1% were in School of Biological Sciences and 14.5% were in Faculty of Educational Foundations. It can be concluded that majority of the study participants were from the Faculty of Social Sciences whilst few were from the Faculty of Educational Foundations.

Distribution of examinations unit workers by gender

Table 7 presents the distribution of examinations unit workers by gender

Table 7: Distribution of Examinations Unit Workers by Gender

Gender	Frequency	Percent (%)
Male	29	72.5
Female	11	27.5
Total	40	100.0

Source: Field Survey, Armah (2017)

Table 7 reveals that 72.5% of the Examinations Unit workers were males while 27.5% were females.

Working experience of examinations unit workers

Table 8 presents the distribution of working experience of Examinations Unit workers who were involved in the study.

Table 8 : Working Experience of Examinations Unit Workers

Working experience	Frequency	Percent (%)
1-5 years	34	85.0
6-10 years	2	5.0
Above 10 years	4	10.0
Total	40	100.0

Source: Field Survey, Armah (2017)

Table 8 shows that 85.0% of the Examinations’ Unit workers who were involved in the study have worked for 1-5 years. This was followed by 10.0% who have worked for above 10 years while 5.0% have also worked for 6-10 years. This means that majority of the Examinations’ Unit workers who were involved in the study have worked for 1 to 5 years.

Research Question One

What is the extent of lecturer’s awareness of the principles of test construction in the University of Cape Coast?

The purpose of research question one was to find out lecturers’ awareness level of the principles of test construction. Respondents were requested to respond to eleven (11) items. A three-point likert scale, ‘very aware’ (3), ‘aware’ (2), and ‘not aware’ (1) was associated with the test construction principles outlined on the questionnaire. A mean score of 2.0 was

obtained by summing up the weight of the responses divided by the total number of responses.

One hundred and thirty-one lecturers provided responses and the results are presented in Table 9.

Table 9 : Analysis of Results of Lecturers Awareness of Test Construction

Principles		
Item	M	SD
1. The purpose of every test should be defined before the test is constructed.	2.12	.881
2. The formats of the test items should be determined before the items are constructed.	2.24	.722
3. A marking scheme should be prepared as soon as test items are written.	2.12	.823
4. The test should adequately cover all aspects of students learning.	2.24	.722
5. The items constructed are to be reviewed at least a week after writing the item.	1.96	.863
6. Test specification table should be prepared for every test.	1.98	.794
7. Items are not to be copied verbatim from textbooks.	2.16	.711
8. More items must be prepared during the draft than needed in the test or examination.	2.15	.783
9. The items should match with the instructional objectives.	2.23	.742
10. Test items after construction are evaluated as a whole to find out whether they are simple and clear.	2.27	.645
11. Clear and concise directions are written for the entire test and sections of the test.	2.44	.609

Source: Field Survey, Armah (2017)

The results from Table 9 shows that lecturers were generally aware of test construction principles. Lecturers indicated that they were aware that clear and concise directions are written for the entire test and sections of test ($M=2.44$, $SD=.609$). It was seen that lecturers were aware that test items after construction are evaluated as a whole to find out whether they are simple and clear ($M=2.27$, $SD=.645$). The results of the study revealed that each of the lecturers were aware that the formats of test items should be determined before the items are constructed and test items should adequately cover all aspects of students learning respectively ($M=2.24$, $SD=.722$). It was found that lecturers were aware that test items should match with the instructional objectives when constructing them ($M=2.23$, $SD=.742$). It was found that lecturers were aware that items are not to be copied verbatim from text books ($M=2.16$, $SD=.711$). The results of the study therefore, revealed that majority of the lecturers involved in the study fair level of awareness of test construction principles.

Research Question Two

What is the extent of lecturer's awareness of the principles of test administration in the University of Cape Coast?

The purpose of research question two was to find out lecturers' awareness level of the principles of test administration. Respondents were requested to respond to eleven (11) items. A three-point likert scale, 'very aware' (3), 'aware' (2), and 'not aware' (1) was associated with the test administration principles outlined on the questionnaire. A mean score of 2.0 was obtained by summing up the weight of the responses divided by the total number of responses.

One hundred and thirty-one lecturers provided responses and the results are presented in Table 10.

Table 10: Analysis of Results of Lecturers Awareness of Test

Administration Principles

Item	M	SD
1. Students should be aware of all rules and regulation governing the test.	2.29	.680
2. The purpose of the test should be communicated to students.	2.27	.669
3. Students should be aware of the content areas the test will cover.	2.28	.686
4. When (date & time) the test will be given should be communicated to students.	2.19	.711
5. Students should be aware of test format.	2.19	.711
6. Eventualities such as power outages should be expected and prepared for.	2.14	.795
7. Hints should not be given to individual students who have problem with some of the items.	2.08	.823
8. Invigilators should not be walking around while students are writing the test.	2.15	.692
9. Students should not be told what will happen to them if they fail the test.	2.05	.768
10. Tests should not be written at any time during school hours.	2.11	.747
11. Announcements are given at regular intervals.	2.29	.680
12. Silence! Exams in progress sign.	2.37	.637

Source: Field Survey, Armah (2017)

Table 10 shows that in sum, lecturers involved in the study were aware of the test administration principles. It was revealed that lecturers were aware of test administration principles, “Silence! Exams in progress” (M=2.37,

SD=.637). The results indicated that lecturers were aware that announcements are given at regular intervals and also students were made aware of all rules and regulation governing the test (M=2.29, SD=.680) respectively. It was found out that lecturers were aware that students should be aware of the content areas the test will cover (M=2.28, SD, 686). Also, lecturers indicated that they were aware that communicating the purpose of the test to students is a principle of test administration (M=2.27, SD= .669). The results also showed that lecturers were aware that when (date & time) the test will be given should be communicated to students and students should be aware of test formats (M=2.19, SD=.711) respectively. The study therefore, revealed that majority of the lecturers involved in the study have fair level of awareness of test administration principles.

Research Question Three

What is the extent to which lecturers apply the principles of test construction in setting classroom test items?

The purpose of research question three was to find out the extent to which lecturers in the University of Cape Coast applied the principles of test construction in setting classroom test items. Respondents were requested to respond to twelve (12) items. A four-point likert scale, 'always' (4), 'very often' (3), 'sometimes' (2), and 'never' (1) was associated with the items outlined on the questionnaire.

One hundred and thirty-one lecturers provided responses and the results are presented in Tables 11, 12 and 13.

Table 11 : Descriptive Statistics on Lecturers' Application of Test

Statement	N	M	SD
1. I define the purpose of every test before constructing test items.	131	2.94	1.11
2. I determine the formats of test items before they are constructed.	131	2.85	.96
3. I prepare marking scheme as soon as I finish writing test items.	131	2.66	.91
4. I write the test items in advance (at least two weeks) of the test date to permit reviews and editing.	131	2.44	.97
5. I prepare a table of specification for test items.	131	2.47	1.05
6. I copy directly test items from textbooks.	131	2.53	.91
7. I write test items which match the instructional objectives.	131	2.48	.97
8. I review test items after construction.	131	2.40	.80
9. I prepare more items in my draft than needed in the test or examination.	131	2.24	.89
10. I write clear and concise directions for the entire test and sections of the test.	131	2.79	.83
11. I consider the knowledge level of test takers before constructing test items.	131	2.78	.96
12. I evaluate test items as a whole to find out whether the test items are simple and clear.	131	2.57	.88

Source: Field Survey, Armah (2017)

A one-sample t-test was conducted to test the statistical significance of the means at the 0.05 level of significance. A test value of 2.5 was used as a cut-off. The 2.5 cut-off point was obtained by summing the weight of the responses divided by the total number of responses. This is because in showing the current lecturers application of test construction principles, ideally lecturers are to either indicate always or very often on the likert scale of the application of test construction principles. The results are shown in Table 12 below.

Table 12 : One-Sample t-Test of Lecturers Application of Test Construction Principles

Items	t	df	Sig. (2-tailed)
1. I define the purpose of every test before constructing test items.	4.507	130	.000*
2. I determine the formats of test items before they are constructed.	4.121	130	.000*
3. I prepare marking scheme as soon as I finish writing test items.	1.969	130	.041*
4. I write the test items in advance (at least two weeks) of the test date to permit reviews and editing.	-.676	130	.501
5. I prepare a table of specification for test items.	-.375	130	.708
6. I copy directly test items from textbooks.	.430	130	.668
7. I write test items which match the instructional objectives.	-.225	130	.822
8. I review test items after construction.	-1.362	130	.005*
9. I prepare more items in my draft than needed in the test or examination.	-3.270	130	.001*
10. I write clear and concise directions for the entire test and sections of the test.	3.937	130	.000*
11. I consider the knowledge level of test takers before constructing test items.	3.311	130	.001*
12. I evaluate test items as a whole to find out whether the test items are simple and clear.	.946	130	.346

Source: Field Survey, (2017)

*Significant at 0.05

As shown in Table 12, lecturers involved in the study applied the following test construction principles. They:

1. defined the purpose of every test before constructing test items.
2. determined the formats of test items before they are constructed.
3. prepared marking scheme as soon as they finish writing test items.
4. reviewed test items after construction.
5. prepared more items in my draft than needed in the test or examination.
6. wrote clear and concise directions for the entire test and sections of the test.
7. considered the knowledge level of test takers before constructing test items.

All these seven items were statistically significant at the 0.05 level of significance.

An overall mean score was further computed for the general application of test construction principles by lecturers by transforming the individual items on applications into a general application of test construction principles. This was done by combining all the individual items into a single application termed general application of test construction principles. The cut-off point for the items were calculated by multiplying the number of items (12 items) by the cut-off point value (2.5). This gave a value of 30. The 30 was then compared with the overall mean score for lecturers' application of test construction principles. Table 13 presents the general application of test construction principles by lecturers.

Table 13 : One-Sample t-Test of General Application of Test Construction Principles

Application	Overall Mean	t	Df	Sig. (2-tailed)
General application of test construction principles	31.21	43.851	130	.000*

Source: Field survey, Armah (2017) *Significant at 0.05

As shown in Table 13 the general result was statistically significant at the 0.05 level of significance. The overall mean score was 31.21. The overall mean score for application of test construction principles is greater than the mean score for the items (30). This implies that in general lecturers applied test construction principles to a high extent.

Research Question Four

What is the extent to which lecturers apply the principles of test administration in setting classroom test items?

The purpose of research question for was to find out the extent to which lecturers in the University of Cape Coast applied the principles of test administration. Respondents were requested to respond to twelve (12) items. A four-point likert scale, ‘always’ (4), ‘very often’ (3), ‘sometimes’ (2), and ‘never’ (1) was associated with the items outlined on the questionnaire.

One hundred and thirty-one lecturers provided responses and the results are presented in Tables 14,15 and 16.

Table 14 : Descriptive Statistics on Lecturers' Application of Test**Administration Principles**

Statement	N	M	SD
1. I make students aware of all rules and regulations governing the test.	131	3.00	1.03
2. I communicate the purpose of the test to my students before they write it.	131	2.60	.89
3. I communicate the content areas that the test items will cover to students.	131	2.48	1.04
4. I conduct tests either immediately before or after holiday or special event.	131	2.28	1.09
5. I inform students of test format.	131	2.46	.96
6. I prepare for eventualities such as power outages.	131	2.21	1.09
7. No hint is given to individual students who have problems with some of the items.	131	2.40	1.11
8. I give announcement at regular intervals during test administration.	131	2.60	.90
9. Students are told what will happen to them if they fail the test.	131	2.37	.94
10. Tests are written at a time that will be convenient for students.	131	2.67	.92
11. I communicate how the test will be scored and graded to students.	131	2.65	.82
12. Silence! Exams in progress sign is mounted at exams center.	131	2.45	.86

Source: Field Survey, Armah (2017)

A one-sample t-test was conducted to test the statistical significance of the means at 0.05 level of significance. A test value of 2.5 was used as a cut-off. The 2.5 cut-off point was obtained by summing the weight of the responses divided by the total number of responses. This is because in showing the current lecturers application of test administration principles,

ideally lecturers are to either indicate always or very often on the likert scale of the application of test administration principles. The results are shown in Table 15 below.

Table 15: One Sample t-Test of Lecturers Application of Test Administration Principles

Statement	t	df	Sig. (2-tailed)
1. I make students aware of all rules and regulations governing the test.	5.554	130	.000*
2. I communicate the purpose of the test to my students before they write it.	1.323	130	.188
3. I communicate the content areas that the test items will cover to students.	-.210	130	.834
4. I conduct tests either immediately before or after holiday or special event.	-2.269	130	.025*
5. I inform students of test format.	-.499	130	.619
6. I prepare for eventualities such as power outages.	-3.010	130	.003*
7. No hint is given to individual students who have problems with some of the items.	-1.059	130	.292
8. I give announcement at regular intervals during test administration.	1.212	130	.228
9. Students are told what will happen to them if they fail the test.	-1.630	130	.106
10. Tests are written at a time that will be convenient for students.	2.148	130	.034*
11. I communicate how the test will be scored and graded students	2.073	130	.040*
12. Silence! Exams in progress sign.	-.659	130	.511

Source: Field Survey, Armah (2017)

*Significant at 0.05

As shown on Table 15, the following principles of test administration were applied. They:

1. made students aware of all rules and regulations governing the test,

2. conducted tests either immediately before or after holiday or special event.
3. prepared for eventualities such as power outages.
4. conducted tests are at times that were convenient for students.
5. communicated how the test will be scored and graded students.

An overall mean score was further computed on the general applications of test administration principles by lecturers by transforming the individual applications into a general applications of test administration principles. This was done by combining all the individual applications into a single applications termed general applications. The cut-off point for the items were calculated by multiplying the number of items (12 items) by the cut-off point value (2.5). This gave a value of 30. The 30 was then compared with the overall mean score for lecturers' application of test administration principles. Table 16 presents the general application of test administration principles.

Table 16: One-Sample t-Test of General Application of Test Administration Principles

Application	Overall Mean	t	df	Sig. (2-tailed)
General application of test administration principles	27.17	1.512	130	.062

Source: Field survey, Armah (2017) Not Significant at 0.05

As shown in Table 16 the general result was not statistically significant at the 0.05 level of significance. The overall mean score was 27.17. The overall mean score for application of test administration principles is less than the mean score for the items (30). This implies that lecturers' did not apply test administration principles.

Research Question Five

What is the extent to which the Examinations Unit of the University Cape Coast apply the principles of test administration?

The purpose of research question five was to find the extent to which the Examinations Unit of the University of Cape Coast applied the principles of test administration. Respondents were requested to respond to twelve (12) items. A four-point likert scale, ‘always’ (4), ‘very often’ (3), ‘sometimes’ (2), and ‘never’ (1) was associated with the items outlined on the questionnaire.

Forty Examinations Unit workers provided responses and the results are presented in Tables 17, 18 and 19.

Table 17 : Descriptive Statistics of the Extent to which Examinations Unit

Statement	N	M	SD
1. Students are made aware of all rules and regulations governing the test during administration.	40	1.73	.88
2. The purpose of the test is communicated to students during administration.	40	2.60	1.01
3. Tests are conducted either immediately before or after holiday or special event.	40	2.23	.73
4. Eventualities such as power outage are prepared for during test administration.	40	3.35	.92
5. Poor ventilations are prepared for during test administration.	40	3.03	.97
6. No hint is given to individual students who have problem with some of the items.	40	2.63	.81
7. Announcements are given at regular intervals during test administration.	40	1.60	.71
8. Students are told what will happen to them if they fail the test.	40	2.83	.98
9. Tests are written at a time that will be convenient for students.	40	2.80	1.07
10. Invigilators walk around while students are writing the test.	40	1.35	.48
11. Students are asked to work faster during the time of testing in order to finish on time.	40	2.65	.86
12. Do Not Disturb. Examinations In Progress sign is used when students are taking tests and examinations.	40	1.90	1.01

Source: Field Survey, Armah (2017)

A one-sample t-test was conducted to test the statistical significance of the means at 0.05 level of significance. A test value of 2.5 was used as a cut-off point. The 2.5 cut-off point was obtained by summing the weight of the responses divided by the total number of responses. This is because in showing the current examinations unit workers application of test administration principles, ideally examination unit workers are to either indicate always or very often on the likert scale of the application of the test administration principles. The result is presented in Table 18.

Table 18 : One Sample t-Test of the Extent to which the Examinations Unit Apply Test Administration Principles

Items	t	df	Sig. (2-tailed)
1. Students are made aware of all rules and regulations governing the test during administration.	-5.591	39	.000*
2. The purpose of the test is communicated to students during administration.	.628	39	.534
3. Tests are conducted either immediately before or after holiday or special event.	-2.372	39	.023*
4. Eventualities such as power outage are prepared for during test administration.	5.835	39	.000*
5. Poor ventilations are prepared for during test administration.	3.410	39	.002*
6. No hint is given to individual students who have problem with some of the items.	.980	39	.333
7. Announcements are given at regular intervals during test administration.	-8.029	39	.000*
8. Students are told what will happen to them if they fail the test.	2.089	39	.043*
9. Tests are written at a time that will be convenient for students.	1.778	39	.083
10. Invigilators walk around while students are writing the test.	-15.057	39	.000*
11. Students are asked to work faster during the time of testing in order to finish on time.	1.098	39	.279
12. Do Not Disturb. Examinations In Progress sign is used when students are taking tests and examinations.	-3.766	39	.001*

Source: Field Survey, Armah (2107)

*Significant at 0.05

As shown in Table 18, the Examinations Unit applied the following test administration principles.

1. Students are made aware of all rules and regulations governing the test during administration.
2. Eventualities such as power outage are prepared for during test administration.
3. Tests are conducted either immediately before or after holiday or special event.
4. Poor ventilations are prepared for during test administration.
5. Announcements are given at regular intervals during test administration.
6. Students are told what will happen to them if they fail the test.
7. Invigilators walk around while students are writing the test.
8. Do Not Disturb. Examinations, In Progress sign is used when students are taking tests and examinations.

An overall mean score was further computed for the general applications of test administration principles by the Examinations Unit by transforming the individual applications into a general applications of test administration principles. This was done by combining all the individual applications into a single applications termed general applications. The cut-off point for the items were calculated by multiplying the number of items (12 items) by the cut-off point value (2.5). This gave a value of 30. The 30 was then compared with the overall mean score for Examinations Unit's application of test administration principles. Table 19 presents the general

application of test administration principles by the staff of the Examinations Unit.

Table 19 : One-Sample t-Test of General Application of Test

Administration Principles in Examination Unit

Application	Overall Mean	t	df	Sig. (2-tailed)
General application of test administration principles	32.23	11.612	39	.002*

Source: Field survey, Armah (2017) *Significant at 0.05

As shown in Table 19 the general result was statistically significant at the 0.05 level of significance. The overall mean score was 32.23. The overall mean score for application of test administration principles is greater than the mean score for the items (30). This implies that generally Examinations Unit’s staff applied the test administration principles.

Research Question Six

What are the major challenges encountered by lecturers in constructing tests in the University of Cape Coast?

The purpose of research question six was to find out the challenges lecturers’ face when constructing tests. Respondents were requested to respond to ten (10) items. A four-point likert scale, ‘strongly agree’ (4), ‘agree’ (3) ‘disagree’ (2) and ‘strongly disagree’ (1) was associated with the challenges encountered by lecturer constructing test items outlined on the questionnaire. The responses were categorized into two main divisions: ‘Agree’ and ‘Disagree’.

One hundred and thirty-one lecturers provided responses and the results are presented in Table 20.

Table 20: Distribution of Results of Challenges in Test Construction

Challenges	Agree		Disagree	
	Freq	%	Freq	%
1. I find it difficult to construct test items of equal difficulty.	85	64.8	46	35.2
2. I find it difficult to determine the formats of the test items before the items are constructed.	97	74.1	34	25.9
3. Time constraints do not permit the preparation of marking scheme before test construction.	96	73.3	35	26.7
4. I find it difficult to construct test items to cover all learning outcomes.	90	68.7	41	31.3
5. Constructing new test items is time consuming.	93	70.9	34	29.1
6. I find it difficult to construct test items which are fair and practical to all students.	89	68.0	42	32.0
7. I find it difficult to construct good distractors for multiple choice items.	88	67.2	43	32.8
8. I do not know how to prepare table of specification.	81	61.8	50	38.2
9. I copy directly test items from textbooks.	88	67.1	43	32.8
10. I find it difficult to prepare more test items than needed in the test or examination.	86	65.7	45	34.3

Source: Field Survey, Armah (2017)

The results from Table 20 showed that lecturers involved in the study encountered challenges when constructing test items. The results showed that 74.1% of the lecturers found it difficult to determine the formats of the test items before the items are constructed. It was also revealed that 73.3% of the

lecturers indicated that time constraints did not permit the preparation of marking scheme before test construction.

The study further found that 70.9% of the lecturers indicated that constructing new test items is time consuming. Also, 68.0% mentioned that they found it difficult to construct test items which are fair and practical to all students and 67.2% of the lecturers indicated that they found it difficult to construct good distractors for multiple choice items and also they copied directly test items from textbooks. Also, 65.7% indicated that they found it difficult to prepare more test items than needed in the test or examinations. It was revealed that 64.8% of the lecturers mentioned that they found it difficult to construct test items of equal difficulty. It can be observed that lecturers who participated in the study encountered challenges when constructing test items.

Research Question Seven

What are the major challenges encountered by lecturers in test administration in the University of Cape Coast?

The purpose of research question seven was to find out the challenges lecturers' face when administering tests. Respondents were requested to respond to seven (7) items. A four-point likert scale, 'strongly agree' (4), 'agree' (3) 'disagree' (2) and 'strongly disagree' (1) was associated with the challenges encountered by lecturer administration test items outlined on the questionnaire. The responses were categorized into two main divisions: "Agree" and "Disagree".

One hundred and thirty-one lecturers provided responses and the results are presented in Table 21.

Table 21: Distribution of Results of Challenges Encounter during Test

Administration

Challenges	Agree		Disagree	
	Freq	%	Freq	%
1. Difficulty in controlling large classes during quizzes and examinations.	113	86.2	18	13.8
2. Difficult to fix a date that will be convenient for all students.	114	87.0	17	13.0
3. Some students are tense when invigilators are going around.	107	81.7	24	18.3
4. Shortage of test materials such as answer booklet.	93	71.0	38	29.0
5. Mostly I find it difficult getting a venue for conducting quizzes.	98	74.8	33	25.2
6. I encounter poor lighting conditions when students are taking tests.	91	69.4	40	30.6
7. I encounter poor ventilation conditions when students are taking tests.	83	63.6	48	36.6

Source: Field Survey, Armah (2017)

The results from Table 21 showed that 87.0% of the lecturers found it difficult to fix a date that is convenient for all students during test. Also, 86.2% indicated that they found difficulty in controlling large classes during quizzes and examinations.

The study further revealed that 81.7% of the lecturers indicated that some students were tense when invigilators were going around and this is a challenge for them. It was also found that 74.8% mostly found it difficult

getting a venue for conducting quizzes. The study results also revealed that 71.0% of the lecturers faced shortage of test materials such as answer booklet during test administration. It was seen that 69.4% encountered poor lighting conditions when students were taking tests and 63.6% also encountered poor ventilation conditions during examinations. It can be observed that lecturers who participated in the study encountered challenges when administering test.

Discussion of Research Findings

The findings of study are discussed in relation with the following.

1. Lecturers' Awareness of Test Construction Principles
2. Lecturers' Awareness of Test Administration Principles
3. Lecturers' use of Test Construction Principles
4. Lecturers' use of Test Administration Principles
5. Examinations Unit use of Test Administration Principles
6. Challenges in Test Construction
7. Challenges in Test Administration

Lecturers' Awareness of Test Construction Principles

The purpose of research question one was to find out from lecturers their level of awareness of test construction principles. The findings from the study revealed that in general, lecturers who participated in the study are aware of the test construction principles. Some of the test construction principles indicated by lecturers who participated in the study include defining the purpose of every test, determining the test items format, preparing marking scheme as soon as test items are written, test items covering adequately all aspects of students learning, preparing test specification table for every test, not copying items from textbooks verbatim and others. Lecturers awareness of

these test construction principles would aid them to know the kind of test items to construct to assess their students. The findings of the study would also help lecturers to educate other colleagues who may not be privy to the principles regarding test construction principles. The findings from the study is supported by previous findings of Magno (2003) who in his study found that lecturers were aware of general principles, guidelines and procedure in test construction. The results further indicated that majority of the lecturers who participated in his study were aware of test construction principles.

The findings also corroborate with research findings of Oduro-Okyireh (2008). In the literature, Oduro-Okyireh pointed that teachers were aware of the basic principles in test construction such as defining the purpose of test, preparing test specification table and others. Kuntasal (2001) in his study revealed that 80.6% of the lecturers were aware of principles of test construction and as a result write direction for their test clearly. Etsey (2004) further found that assessors' awareness of the test specifications table would help them to match the course content with the instructional objectives. With the total number of items on the test in mind, the specification table helps to avoid overlapping in the construction of the test items, helps to determine the weighting of learning outcomes with respect to content areas, and makes sure that justice is done to all aspects of the course, thereby helping to ensure the content validity of the test.

Lecturers' Awareness of Test Administration Principles

The purpose of research question two was to find out from lecturers their level of awareness of test administration principles. The findings from the study showed that lecturers involved in the study were aware of the test

administration principles. Lecturers who participated in the study indicated the following as some of the test administration principles they are aware of; students are made aware of the rules and regulations governing the test, the purpose of the test is communicated to them, students are made aware of the content areas the test will cover, when the test will be given, preparing for eventualities such as power outages and silence! Exams in progress sign. Lecturers awareness of the test administration principles would help them to prepare students for examinations since students would know the nature and when the test would be taken.

The findings of the study are line with the findings of Etsey (2004) who found out in his study that teachers' awareness of test administration principles would first and foremost helps them to prepare their students in advance for the test. He further emphasized that for students' maximum performance, they should be made aware of when (date and time) the test will be given, the conditions (number of items, place of test, open or closed book) under which the test will be given, the content areas (study questions or list of learning targets) that the test will cover, the emphasis or weighting of content areas, the kinds of items (objective-types or essay-types) on the test, how the test will be scored and graded, and the importance of the results of the test.

The results of the study further corroborate research findings of Tamakloe et al., (1996). Tamakloe et al., revealed that teachers were aware of the test administration principles. The awareness level would aid teachers to ensure quietness in the vicinity, good lighting and ventilation. Adequate work space is very essential for test administration because when tables and chairs are closely arranged together, students will not have the independence to work

on their own. This will in no doubt lead to students copying from each other. In addition, tables provided for the examination must be conducive to the testing materials being used. For example, in Practical Geography examinations where topographical sheets are used, each student could use two tables or desks in order to get adequate work space (Tamakloe et al., 1996).

It was moreover shown that lecturers awareness of test administration principles would help them to put, “Silence! Exams in progress” sign post at vantage point to inform outsiders. This would prevent people from coming near the exams center hence ensuring peaceful atmosphere for the examinations. Etsey (2004) pointed out that it is helpful to hang a “Do Not Disturb Testing in Progress” sign at the door of the testing room to warn people to keep off. Good lighting is important in effective test administration. This facilitates students reading of instructions and test items without straining their eyes, thereby working faster (Oduro-Okyireh, 2008). Oduro-Okyireh also indicated that the awareness of test administration principles would help teachers to position invigilators well.

Lecturers’ Application of Test Construction Principles

The purpose of research question three was to find out from lecturers who participated in the study how they applied test construction principles. Lecturers indicated how they applied test construction principles. These include defining the purpose of every test before constructing test items, determining the formats of test items, preparing marking schemes, reviewing test items after construction, preparing more items in draft than needed in the test writing clear and concise directions for the entire test and sections.

The application of test construction principles by lectures could make them construct test items which are standard and of high reliability. This could also help to know the strengths and weaknesses of students in the class by writing clear and concise directions for the entire test and sections of the test during test items construction. These findings confirm the findings of Etsey (2004) which indicated that teachers' competencies and skills in writing clear and concise directions for test will help students to respond to test items effectively. This would prevent students from asking invigilation assistants questions regarding the tests and hence avoid students cheating in the test.

The findings further support the finding of Tamakloe et al. (1996) who indicated that reviewing and editing the items are for the purpose of removing or rewording poorly constructed items, checking difficulty level of items, checking the length of the test. This implies that test items which are faulty would be identified and corrected accordingly. Hence, ensuring smooth and effective test construction.

Lecturers' Application of Test Administration Principles

The findings for research question four which sought to find out lecturers' application of test administration principle indicated that in general, lecturers who participated in the study applied few test administration principles when assessing their students. The study revealed that lecturers made students aware of all rules and regulations governing the test, tests were written at a time that was convenient for students and how the test would be scored and graded were communicated to students prior to test administration.

However, the study revealed that lectures did not communicate the purpose of the tests to students before they were written, the content areas

were also not communicated to students. Similarly, the formats of the tests were not communicated, announcement were not given at regular intervals during test administration, and “Silence! Exams in progress” sign was not mounted. Lecturers inability to apply the test administration principles during test administration could make the test administration procedures to be characterized with errors such as cheating in the exams. This could make some students have undue advantage over the others by cheating. This could thereby make the performance of students to be associated with errors. These findings, however, contradict research findings of Oduro-Okyireh (2008) who indicated that teachers’ ability to apply principles in test administration helped them to provide all examinees with a fair chance to demonstrate their achievement on what is being measured. The need to maintain uniform conditions in test administration cannot be over-emphasised. This is especially essential for the test to yield consistent, reliable and valid scores without much influence of chance errors.

Examinations Unit Application of Test Administration Principles

The findings of research question five indicated that in general, workers at the Examinations Unit of the University of Cape Coast applied the principles of test administration. The study revealed that the purpose of the test was communicated to students before they wrote test, announcement was given at regular intervals during test administration, “Do Not Disturb. Examinations In Progress” sign was used when students were taking tests and examinations, all rules and regulations governing the test were communicated to students and tests were written at a time that was convenient for students.

The findings of the study confirm previous findings of Mehrens and Lehmann (1991) and Oduro-Okyireh (2008). In their studies, they found that interruptions within and outside the testing room have the tendency to affect students' performance. To avert situations, they suggested that teachers put in measures to reduce and if possible, eradicate interruptions. Etsey (2004) pointed out that it is helpful for teachers to hang a "Do Not Disturb! Testing in Progress" sign at the door of the testing room to warn people to keep off. Good lighting is important in effective test administration. This facilitated students reading of instructions and test items without straining their eyes, thereby working faster. Tamakloe et al., (1996) opined that good ventilation and comfortable temperature should be assured since their absence could create unrest or uneasiness in testees making concentration difficult

Challenges in Test Construction

The results from research question six indicated that lecturers involved in the study encountered challenges when constructing test items. Lecturers indicated the following as challenges they encountered when constructing test item; difficulty in constructing test items of equal difficulty, difficulty in determining the format of the tests items and difficulty in constructing test items to cover all learning outcomes, time consuming in constructing new test items and difficulty in constructing good distracter for multiple choice items.

These challenges indicated above could affect the test items constructed by lecturers. Lecturers could repeat test items on several occasions which would make students not to be serious with the course. It could also make the results of students who would be assessed with the items to be unreliable. Hence, lecturers need to have much knowledge with respect to test

item construction so as to curtail the challenges they face when constructing test items. Lecturers who did not have knowledge in educational measurement and evaluation could enroll on the programme. This could make them to be abreast with principles of test items construction. Lecturers could also give constructed test items to their colleagues' lecturers for further editing and review. This would help improve the items.

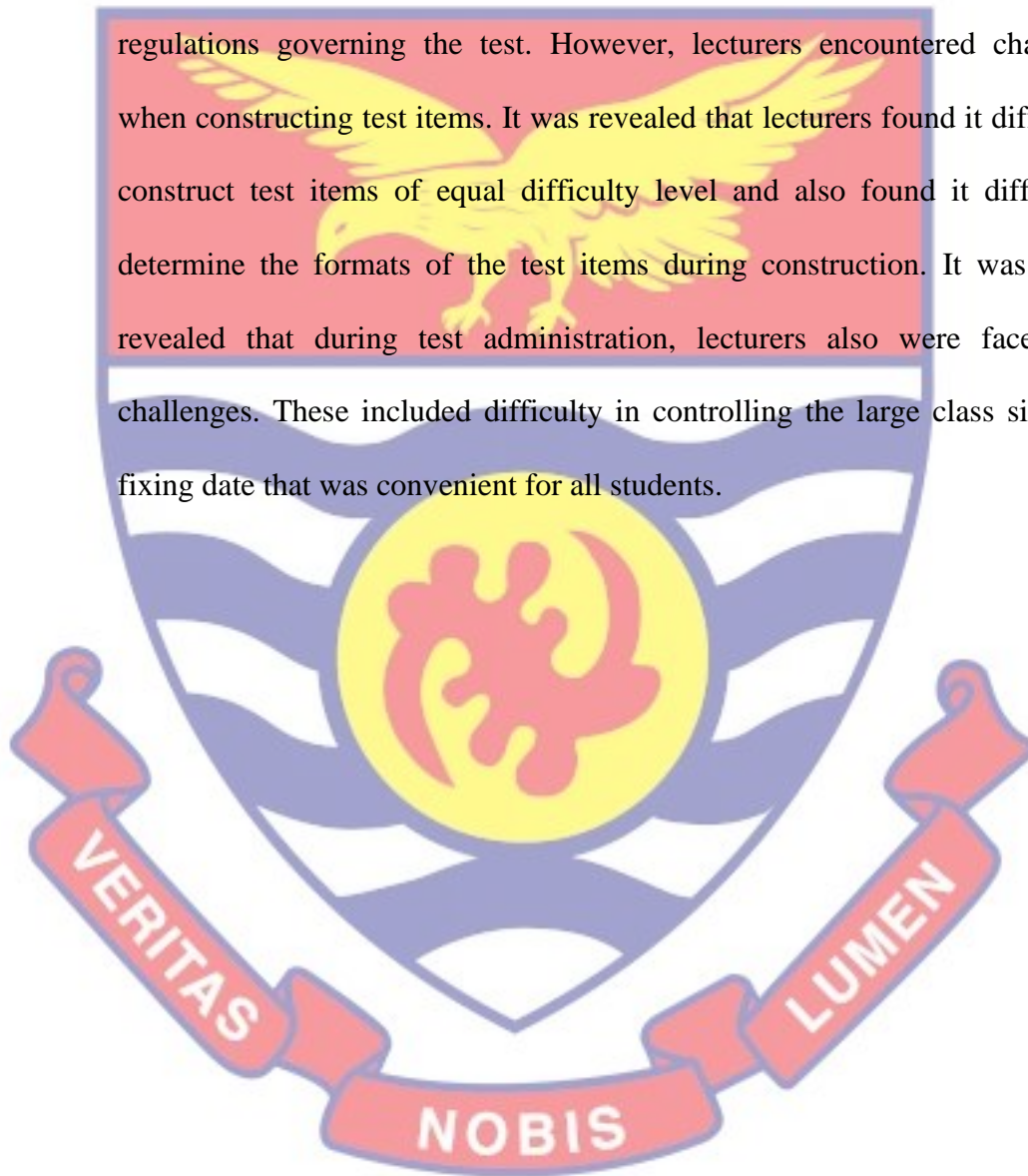
Challenges in Test Administration

The findings of research question seven showed that in general, lecturers who participated in the study faced challenges when administering test items. The study revealed the following as challenges lecturers faced; difficulty in controlling large classes during quizzes and examinations, difficulty in fixing a date that was convenient for all students, shortage of test materials such as answer booklet, poor lighting conditions when students were taking tests and poor ventilation conditions when students were taking test.

The study found that large class sizes made it difficult for lecturers to supervise students during quizzes and examinations. Students therefore took the opportunity to cheat which did not contribute to proper assessment procedures. Some unforeseen eventualities such as deaths made it difficult for lecturers to plan a particular date for quiz which is appropriate for all students. Moreover, some students sometimes had clashes in time table. These clashes sometimes made the slated time and date not to be convenient to some students. By implication the affected students either decided to write and forgo the other class or otherwise.

Chapter Summary

In summary, the study revealed that lecturers involved in the study were aware of the test construction and administration principles. It was shown that lectures defined the purpose of every test prior to construction. Also, during test administration, students were made aware all the rules and regulations governing the test. However, lecturers encountered challenges when constructing test items. It was revealed that lecturers found it difficult to construct test items of equal difficulty level and also found it difficult to determine the formats of the test items during construction. It was further revealed that during test administration, lecturers also were faced with challenges. These included difficulty in controlling the large class sizes and fixing date that was convenient for all students.



CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview of the Study

The purpose of the study was to assess the extent to which lecturers applied principles of test construction and administration in the University of Cape Coast. The Chapter is divided into four sections. The first section highlights the summary of main findings of the study. The second section discusses the overall study conclusions of the study. Recommendations for closing the gap are presented in section three. Finally, section four discusses the suggestions for further research in testing practices.

The study specifically focused on the extent to which (a) lecturers were aware of the principles of test construction, (b) lecturers were aware of the principles of test administration, (c) lecturers applied the principles of test construction in setting classroom test items, (d) lecturers applied the principles of test administration in setting classroom test items, (e) the staff of Examinations Unit of the University Cape Coast applied the principles of test administration, (f) the major challenges encountered by lecturers in constructing tests items and (g) major challenges encountered by lecturers in test administration.

The study was conducted in the University of Cape Coast in the Cape Coast Metropolis in the Central Region of Ghana. The study included surveys of 131 lecturers from four selected faculties/schools and 40 Examinations Unit workers in the University of Cape Coast.

Summary of Key Findings

The following are the main findings from the data analysis:

1. The study found that lecturers were fairly aware of the test construction principles. Lecturers involved in the study indicated that they defined the purpose of every test prior to construction, determined the test items format and prepared marking scheme as soon as test items were written. Lecturers also indicated that they constructed test items to adequately cover all aspects of students learning.
2. Lecturers were fairly aware of the test administration principles. They indicated that students were made aware of the rules and regulations governing the test, the purpose of the test is communicated to students and students were made aware of the content areas the test would cover during test administration.
3. The study showed that lecturers applied some principles of test construction. These included defining the purpose of every test before constructing test items, determining the formats of test items, preparing marking schemes and reviewing test items after construction.
4. It was found that lecturers who participated in the study applied few test administration principles. Lecturers indicated that they made students aware of all rules and regulations governing the test and test were written at a time that was convenient for student.
5. The findings showed that workers at the Examinations Unit of the University of Cape Coast applied the principles of test administration. They indicated that during examinations the purpose of the test was communicated to students before they wrote the test, announcement

was given at regular intervals during test administration and “Do Not Disturb. Examinations In Progress” sign was used when students were taking tests and examinations.

6. Lecturers indicated that they encountered challenges when constructing test item. These included, difficulty in constructing test items of equal level of difficulty, difficulty in determining the format of the test items, difficulty in constructing test items to cover all learning outcomes, constructing new test items was time consuming and difficulty in constructing good distracters for multiple choice items.
7. The study finally showed that lecturers had difficulty in controlling large classes during quizzes and examinations, difficulty in fixing a date that would be convenient for all students, shortage of test materials such as answer booklet, poor lighting conditions when students were taking tests and poor ventilation conditions when students were taking test.

Conclusions

On the application of the principle of test construction, lecturers generally reported they applied the principles when constructing test items. It could therefore, be concluded that, to a great extent, lecturers in the University of Cape Coast applied test construction principles in the sense that they were able to construct test item bearing in mind the purpose of the test and hence improving upon the quality of test items. The staffs of the Examinations Unit reported that they applied test administration principles. However, lecturers admitted that they encountered some challenges when constructing and

administering test items. This made lecturers repeat test items on several occasions. Also, large class sizes could make students cheat which does not contribute to proper assessment procedures. Hence, they need to have knowledge in educational measurement and evaluation so as to be abreast with test construction and administration principles.

Recommendations

From the findings from the study, it was evidently clear that lecturers who participated in the study were aware of test construction and administration principles. However, the application of test administration principles by lecturers was low. I therefore, provide the following recommendations to serve as motivation for lecturers regarding test item construction and administration.

I recommend that lecturers are given timely in-service training on test item construction even though they are doing quite well. The purpose of this is to help lecturers advance their knowledge with contemporary issues regarding test item construction. This would also help lecturers to adopt the contemporary procedures of constructing test items thereby, improving the reliability of the test results. I recommend that lecturers should be encouraged by the Academic Board of the University of Cape Coast not to hint students who have problems on test items.

It was revealed that lecturers did not write the test items in advance (at least two weeks) of the test date to permit reviews and editing, did not prepare table of specification for test items and some copied directly from textbooks. I recommend that lecturers should write test items in advance, prepare test

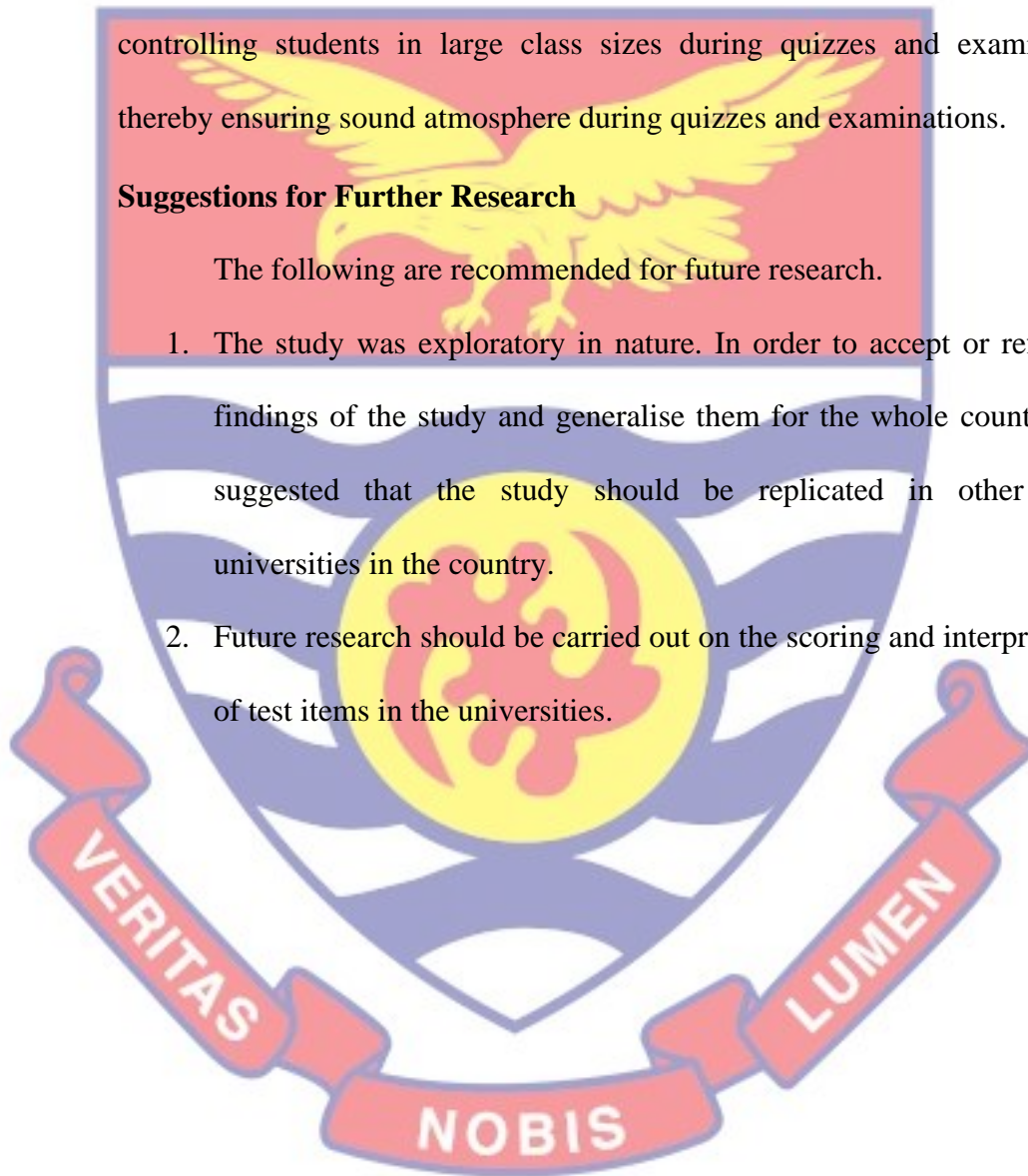
specification table and should reduce copying directly from textbooks. This would help to ensure objectivity and accuracy in their assessment results.

Large class sizes were indicated as a challenge. I therefore, recommend that the university management puts students in smaller class sizes. This would help lecturers to overcome most of the challenges such as difficulty in controlling students in large class sizes during quizzes and examinations thereby ensuring sound atmosphere during quizzes and examinations.

Suggestions for Further Research

The following are recommended for future research.

1. The study was exploratory in nature. In order to accept or refute the findings of the study and generalise them for the whole country, it is suggested that the study should be replicated in other public universities in the country.
2. Future research should be carried out on the scoring and interpretations of test items in the universities.



REFERENCES

- Adane, L. O. (2013). *Factors affecting low academic achievement of pupils in Kemp Methodist Junior High School in Aburi, Eastern region*. A Thesis submitted to the University of Ghana, Ghana.
- Akplu, H. F. (1989). Continuous assessment in schools and colleges. *Ghana Association of Science Teachers Journal*, 1(2), 1-3.
- Amedahe, F. K., & Asamoah-Gyimah, K. (2013). Introduction to Measurement and Evaluation. Cape Coast: CCE Publications.
- Amedahe, F. K. (1989). *Testing practices of secondary schools in the Central Region of Ghana*. Unpublished master's thesis, University of Cape Coast, Cape Coast, Ghana.
- Amedahe, F. K. (2000). *Continuous assessment*. Unpublished paper. University of Cape Coast, Ghana.
- Amedahe, F. K. (2002). *Fundamentals of educational research methods*. Mimeograph, University of Cape Coast (Unpublished).
- Amedahe, F. K. (2004). *Notes on educational research*. Unpublished. University of Cape Coast, Ghana.
- Amedahe, F. K., & Gyimah, K. A. (2003). *Measurement and evaluation*. Cape Coast, Ghana: Centre for Continuing Education.
- Amedahe, F.K. (2010). *Notes on educational research methods*. University of Cape Coast: Faculty of Education.
- Amedahe, F. K. (Ed.). (2008). *Teacher education conference proceedings*. University of Cape Coast: Adwinsa Publications (Gh.) Ltd.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Anamuah-Mensah, J., & Quagrain, K. A. (1998). Teacher competence in the use of essay -tests. A study of secondary schools in the Western region of Ghana. *The Oguaa Educator*. 12(1), 31-43.

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

Andrews, T. E., & Barnes, S. (1990). Assessment of teaching. In W. R. Houston (Ed.), *Handbook of research on teacher education: A project of the Association of Teacher Educators* (pp. 569 – 598). New York: Macmillan.

Ary, D., Jacobs, L. C., & Razavieh, A. (1990). *Introduction to research in education*. (4th ed.). Forth Worth: Holt, Rinehart & Winston Inc.

Asamoah-Gyimah, K. (2002). *An evaluation of the practice of continuous assessment in the secondary schools in the Ashanti Region of Ghana*. Unpublished University of Cape Coast, Ghana.

Babbie, A. (2005). Test fairness in traditional and dynamic assessment. *Theory and Practice in Language Studies*, 3(10), 1930-1938

Baker, E. L. (2003). *Testing and assessment: A progress report*. Educational Assessment 7: 1–12.

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

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.

Brown, M. (1976). *Exploring classroom assessment in mathematics: A guide for professional development*. National Council of Teachers of Mathematics: Reston

Cohen, L., Morrison, L., & Manion, K. (2004). *Research methods in education*. London: Routledge.

Cooper, C., & Schindler, E. (2000). *Learner-centred assessment*. Launceston: Global Learning Communities.

Crocker, L., & Algina, J. (1986). *Fundamental statistics in psychology and education*. USA: Lengage learning.

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

Crooks, T. (2001). The validity of formative assessment. *Educational Assessment Unit*. University of Otago, Dunedin, New Zealand. A paper presented at the British Educational Research Association annual Conference, University of Leeds, 13-15 September 2001.

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

DeVellis, R. F. (1991). *Scale development: Theory and applications* (Applied Social Research Methods Series, Vol. 26). Newbury Park: Sage.

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

Ebel, W. (1979). *Fundamental statistics in psychology and education*. (7th ed.). Tokyo: McGraw-Hill.

Etsey, Y. K. (2002). Pre-service teachers' knowledge of continuous assessment techniques in Ghana. *Journal of Educational Development and Practice*, 1(1), 1 - 8.

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

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

Etsey, Y. K. A. (2008). Notes on educational measurement and evaluation (unpublished lecture notes). University of Cape Coast, Ghana

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

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

Fraenkel, J. R., & Wallen, N. E. (1993). *How to design and evaluate research in education* (2nd ed.). Boston: McGraw Hill.

Fraenkel, J. R., & Wallen, N. E. (2000). *How to design and evaluate research in education*. (4th ed.). New York: McGraw-Hill, Inc.

Frey, B. B. (2007). *An introduction to quality test construction*. Retrieved from

<http://www.specislconnection.ku.edu/cgi-bin/cgiwrap/speccom/mainhp?cat=assessment§ion=main§ion=qualitytest/main>.

Gay, L., Mills, G., & Schumacher, W. (2006). *Educational research: competencies for analysis and applications*. New York, NY: Prentice Hall.

Gay, R. L. (2011). *Educational research: Competencies for analysis and application*. (4th ed.). New York: Macmillan Publishing Company.

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

Gregory, R. J. (1989). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn and Bacon.

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

Gregory, T. R. (2009). Grading and reporting in a standards-based environment: Implications for students with special needs. *Theory into Practice*, 48(1), 53–62

Griswold, P. A. (1990) Assessing relevance and reliability to improve the quality of teacher-made tests. *NASSP Bulletin*, 76, 18-24.

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

Gronlund, N. E. (1988). *How to construct achievement tests*. (4th ed.). New Jersey: Prentice-Hall, Inc.

Gronlund, N. E., & Linn, N. (1985). *Measurement and evaluation*. New York: Macmillan.

Gronlund, N. E., & Linn, R. L. (1995). *Measurement and assessment in teaching* (7th ed.). New Jersey: Merrill, Prentice-Hall.

Gyimah, K. A. (2002). *An evaluation of the practice of continuous assessment in the secondary schools in the Ashanti Region of Ghana*. Unpublished University of Cape Coast, Ghana.

Hussain, L., Jamil, H., Siraji, J., & Maroof, K. (2012). Development and Standardization of Intelligence Test for Children. *International Journal of Learning & Development*, 2(5), 190-202

Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association (JCSEPT), and the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.

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 Onlines J. Distance Educ.* 5(1), 23-36.

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

Leedy, D. P., & Ormrod, E. J. (2010). *Research: Planning and design* (9th ed.). Merrill, Upper Saddle River: Pearson Education Inc.

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., & Gronlund, N. E. (1995). *Measurement and assessment in teaching*. (7th ed.). New Jersey: Merrill, Prentice-Hall.

Linn, R. L., & Miller, M. D., (2005). *Measurement and assessment in teaching*. Upper Saddle River, New Jersey: Prentice Hall, Inc.

Linquist, T. (1996). *Seeing the whole through social studies*. NH: Heinemann.

Magno, E., (2003). Equating and linking of performance assessments. *Applied Psychological Measurement* 24, 325–337.

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology*. (4th ed.) New York: Holt, Rinehart and Winston Inc.

Mehrens, W. A., & Lehmann, I. J. (1995). *Measurement and evaluation in education and psychology*. (3rd ed.) New York: Harcourt Brace College Publishers.

Mehrens, W. A., & Lehmann, I. J. (1999). *Measurement and evaluation in education and psychology*. New York: Harcourt Brace College Publishers.

Mehrens, W. A., & Lehmann, I. J. (2005). *Research design explained* (6th ed.). New York: Holt, Reinehart and Winston.

Messick, S. (1989). Validity. (In R. L. Linn (Ed), *Educational measurement* (2nd ed. pp. 13-104). Phoenix: American Council on Education and Oryx Press.)

Miller, D. M., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and evaluation in teaching*, (12th ed.). New York: Merrill Press Edu Inc.

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

Neuman, W. L. (2003). *Social research methods: Qualitative and quantitative approaches*, (5th ed.). University of Wisconsin at Whitewater: A and B Publishers.

Nitko, A. J. (2001). *Educational tests and measurements* (3th ed.). Prentice Hall, Inc. Upper Saddle River, New Jersey.

Nitko, A. J. (2004). *Educational assessment of students*, (4th ed). Upper Saddle River, NJ: Pearson Education, Inc.

Nitko, A. J., & Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practices*, 18 (1), 5-13.

Nunnally, J. C. (2016). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

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, Cape Coast, Ghana.

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. (2001). *Introduction to research methodology*. Onitsha, Nigeria: Africana EEP Publishers Ltd.

Polit, D. F., & Hungler, B. P. (1998). *Essentials of nursing research: Methods, appraisal, and utilization* (8th ed.). Philadelphia: Wolter Kluwer/Lippincott Williams and Wilkins.

Quaigrain, A. K. (1992). *Teacher-competence in the use of essay tests: A study of secondary schools in the Western Region of Ghana.*

Unpublished Master's Thesis, University of Cape Coast, Ghana.

Rashan, D. (2016). Assessing business knowledge. In K. Martell, & T. Calderon, (Eds.), *Assessment of Student Learning in Business Schools: Best Practices Each Step of the Way*. Tallahassee, FL: Association for Institutional Research, Florida State University.

Reynolds, M. C. (Ed.). (1989). *Knowledge base for the beginning teacher*. New York: Pergamon Press.

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

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

Sarantakos, S. (1998). *Social research* (3rd ed.). NY Houndmills, Palgrave.

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

Sarita, K. (2011). *Theory and problems of statistics*. (2nd ed.). New York: McGraw-Hill Inc.

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

Shadish, W. R., Cook T. D., & Campbell D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Ujah, E. U. (2001). *Development and validation of an introductory technology achievement test*. Unpublished M.Ed. thesis, University of Nigeria, Nsukka.

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice*, 10 (1), 7 -12

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

Tyler, R. H. (1986). A review of the current state of educational measurement. *The New Era*, 67(3), 5-13.

Tyler, R. H. (1991). *Formal assessment in the classroom: The Ghana Education Service termly assessment plan*. Paper presented to the Quality Improvement in Primary Schools (QUIPS) Project. Funded by the USAID.

Webb, N. (2002). *Depth-of-knowledge levels for four content areas*. Wisconsin Center for Educational Research. New Jersey: Lawrence Erlbaum Associates.

William, D. (2001). An overview of the relationship between assessment and the curriculum, In D. Scott, (Ed.) *Curriculum and assessment* (Westport, CT, Ablex Publishing), 165–181.

Worthen, W. W., & Cole, N. S. (1993). *Gender and fair assessment*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

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, Cape Coast, Ghana.



SECTION B: LECTURERS' AWARENESS OF TEST

CONSTRUCTION PRINCIPLES

Indicate your level of awareness on the following activities regarding the principles of test construction by using VMA = Very much aware, A = Aware and NA = Not aware.

S/N	Item	VA	A	NA
4	The purpose of every test should be defined before the test is constructed.			
5	The formats of the test items should be determined before the items are constructed.			
6	A marking scheme should be prepared as soon as test items are written.			
7	The test should adequately cover all aspects of students learning.			
8	The items constructed are to be reviewed at least a week after writing the item.			
9	Test specification table should be prepared for every test.			
10	Items are not to be copied verbatim from textbooks.			
11	More items must be prepared during the draft than needed in the test or examination.			
12	The items should match with the instructional objectives.			
13	Test items after construction are evaluated as a whole to find out whether they are simple and clear.			

14	Clear and concise directions are written for the entire test and sections of the test.			
----	--	--	--	--

SECTION C: LECTURERS' AWARENESS OF TEST

ADMINISTRATION PRINCIPLES

Indicate your level of awareness on the following activities regarding the principles of test administration by using VMA = Very much aware, A = Aware and NA = Not aware.

S/N	Item	VMA	A	NA
15	Students should be aware of all rules and regulation governing the test.			
16	The purpose of the test should be communicated to students.			
17	Students should be aware of the content areas the test will cover.			
18	When (date & time) the test will be given should be communicated to students.			
19	Students should be aware of test format.			
20	Eventualities such as power outages should be expected and prepared for.			
21	Hints should not be given to individual students who have problem with some of the items.			
22	Invigilators should not be walking around while students are writing the test.			

23	Students should not be told what will happen to them if they fail the test.			
24	Tests should not be written at any time during school hours.			
25	Announcements are given at regular intervals.			
26	Silence! Exams in progress sign.			

SECTION D: LECTURERS' USE OF TEST CONSTRUCTION PRINCIPLES

Indicate the frequency of the following activities regarding how you use the principles of test construction in setting classroom test items by using A =Always, VO = Very Often, S = Sometimes, and N = Never.

S/N	Items	A	VO	S	N
27	I define the purpose of every test before constructing test items.				
28	I determine the formats of test items before they are constructed.				
29	I prepare marking scheme as soon as I finish writing test items.				
30	I write the test items in advance (at least two weeks) of the test date to permit reviews and editing.				
31	I prepare a table of specification for test items.				

32	I copy directly test items from textbooks.				
33	I write test items which match the instructional objectives.				
34	I review test items after construction.				
35	I prepare more items in my draft than needed in the test or examination.				
36	I write clear and concise directions for the entire test and sections of the test.				
37	I consider the knowledge level of test takers before constructing test items.				
38	I evaluate test items as a whole to find out whether the test items are simple and clear.				

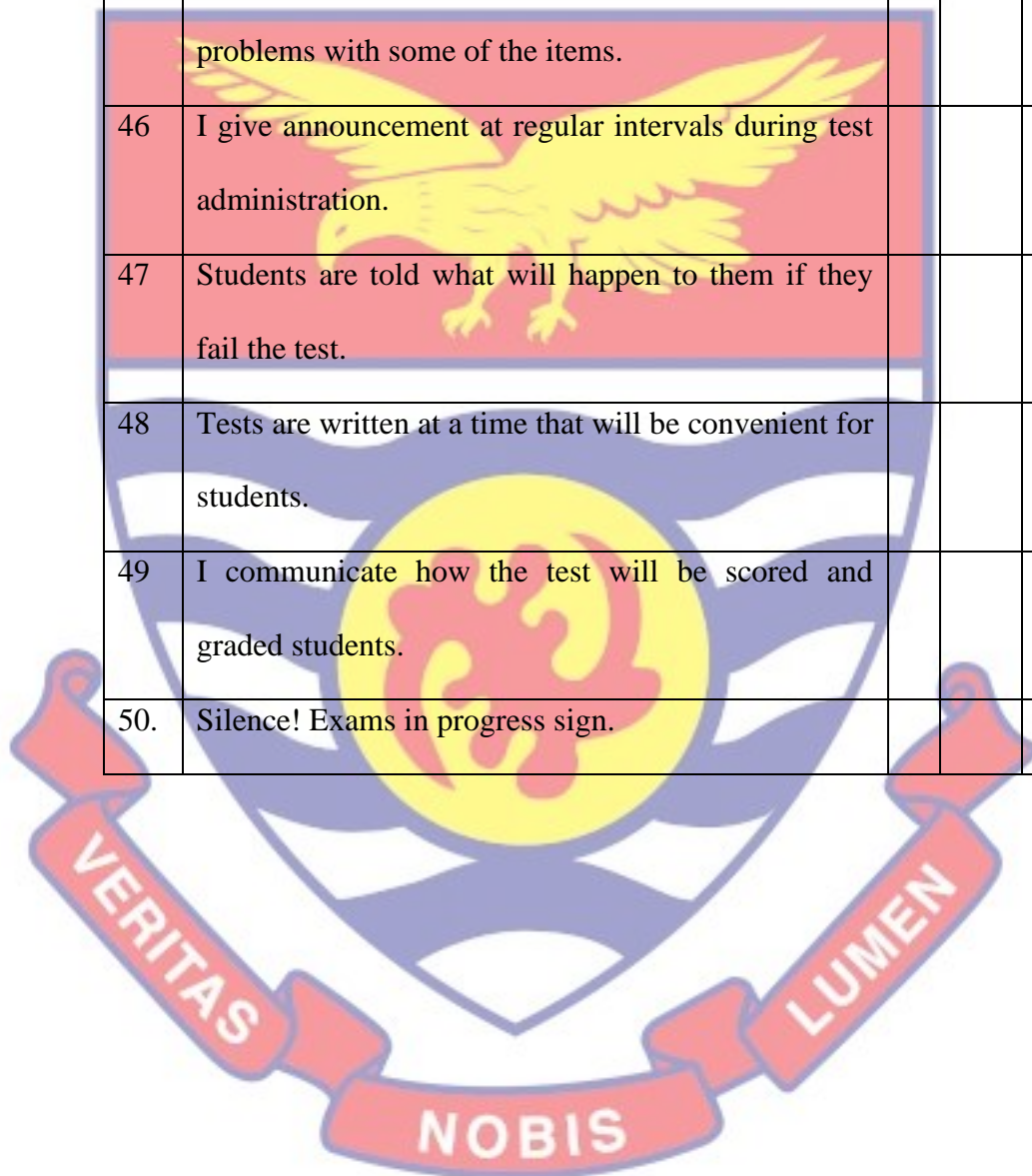
SECTION E: LECTURERS' USE OF TEST ADMINISTRATION

PRINCIPLES

Indicate the frequency on the following activities regarding how you use the principles of test administration in setting classroom test items by using A =Always, VO = Very Often, S = Sometimes, and N = Never.

S/N	Items	A	VO	S	N
39	I make students aware of all rules and regulations governing the test.				
40	I communicate the purpose of the test to my students before they write it.				
41	I communicate the content areas that the test items will cover to students.				

42	I conduct tests either immediately before or after holiday or special event.				
43	I inform students of test format.				
44	I prepare for eventualities such as power outages.				
45	No hint is given to individual students who have problems with some of the items.				
46	I give announcement at regular intervals during test administration.				
47	Students are told what will happen to them if they fail the test.				
48	Tests are written at a time that will be convenient for students.				
49	I communicate how the test will be scored and graded students.				
50.	Silence! Exams in progress sign.				



SECTION F: CHALLENGES IN TEST CONSTRUCTION

Indicate your level of agreement on the following activities regarding the challenges in applying the principles of test construction by using SA = Strongly Agree, A = Agree, D = Disagree, and SD = Strongly Disagree.

S/N	Items	SA	A	D	SD
51	I find it difficult to construct test items of equal difficulty.				
52	I find it difficult to determine the formats of the test items before the items are constructed.				
53	Time constraints do not permit the preparation of marking scheme before test construction.				
54	I find it difficult to construct test items to cover all learning outcomes.				
55	Constructing new test items is time consuming.				
56.	I find it difficult to construct test items which are fair and practical to all students.				
57	I find it difficult to construct good distractors for multiple choice items.				
58	I do not know how to prepare table of specification.				
59	I copy directly test items from textbooks.				
60	I find it difficult to prepare more test items than needed in the test or examination.				

SECTION G: CHALLENGES OF THE TEST ADMINISTRATION

Indicate your level of agreement on the following activities regarding the challenges in applying the principles of test administration by using SA = Strongly Agree, A = Agree, D = Disagree, and SD = Strongly Disagree.

S/N	Items	SA	A	D	SD
61	Sometimes difficulty in controlling large classes during quizzes and examinations.				
62	Sometimes difficult to fix a date that will be convenient for all students.				
63	Some students are tense when invigilators are going around.				
64	Shortage of test materials such as answer booklet.				
65	Mostly I find it difficult getting a venue for conducting quizzes.				
66	I encounter poor lighting conditions when students are taking tests.				
67	I encounter poor ventilation conditions when students are taking tests.				

APPENDIX B

UNIVERSITY OF CAPE COAST

COLLEGE OF EDUCATION STUDIES

FACULTY OF EDUCATIONAL FOUNDATIONS

DEPARTMENT OF PSYCHOLOGY AND EDUCATION

QUESTIONNAIRE FOR EXAMINATIONS UNIT

The purpose of this questionnaire is to assess the extent to which Examinations Unit in the University of Cape Coast apply principles of test administration. Your full participation will help make informed decisions about the principles of test construction and administration in universities. It would therefore be appreciated if you could provide responses to all items on the questionnaire. You are assured of complete confidentiality and anonymity of all information provided. Thank you in advance for your cooperation.

DIRECTIONS: Please tick [] where appropriate, and for others you may specify by writing.

SECTION A: DEMOGRAPHIC DATA OF RESPONDENT

1. Gender Male [] Female []

2. Working experience

1-5 years []

6-10years[]

Above 10 years []

**SECTION B: EXTENT TO WHICH EXAMINATIONS UNIT APPLY
TEST ADMINISTRATION PRINCIPLES**

Indicate the extent to which the following activities are applied in test administration by using A =Always, VO = Very Often, S = Sometimes, and N = Never.

S/N	Items	A	VO	S	N
3	Students are made aware of all rules and regulations governing the test during administration.				
4	The purpose of the test is communicated to students during administration.				
5	Tests are conducted either immediately before or after holiday or special event.				
6	Eventualities such as power outage are prepared for during test administration.				
7	Poor ventilations are prepared for during test administration.				
8	No hint is given to individual students who have problem with some of the items.				
9	Announcements are given at regular intervals during test administration.				
10	Students are told what will happen to them if they fail the test.				
11	Tests are written at a time that will be convenient for students.				

12	Invigilators walk around while students are writing the test.				
13	Students are asked to work faster during the time of testing in order to finish on time.				
14	Do Not Disturb. Examinations In Progress sign is used when students are taking tests and examinations.				

