

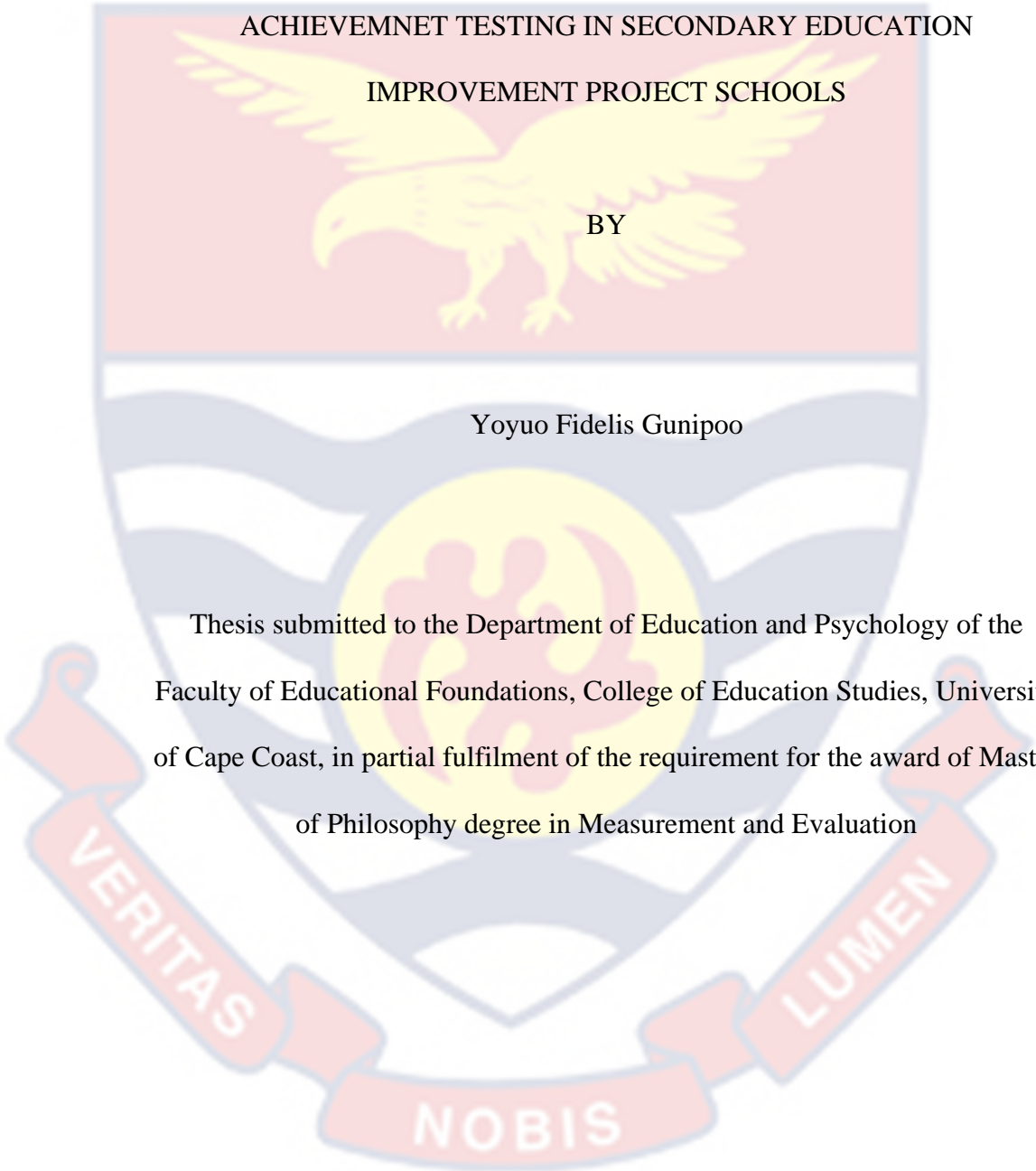
UNIVERSITY OF CAPE COAST

ASSESSMENT OF MATHEMATICS TEACHERS' PRACTICE OF
ACHIEVEMENT TESTING IN SECONDARY EDUCATION
IMPROVEMENT PROJECT SCHOOLS

BY

Yoyuo Fidelis Gunipoo

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



December 2022

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.....

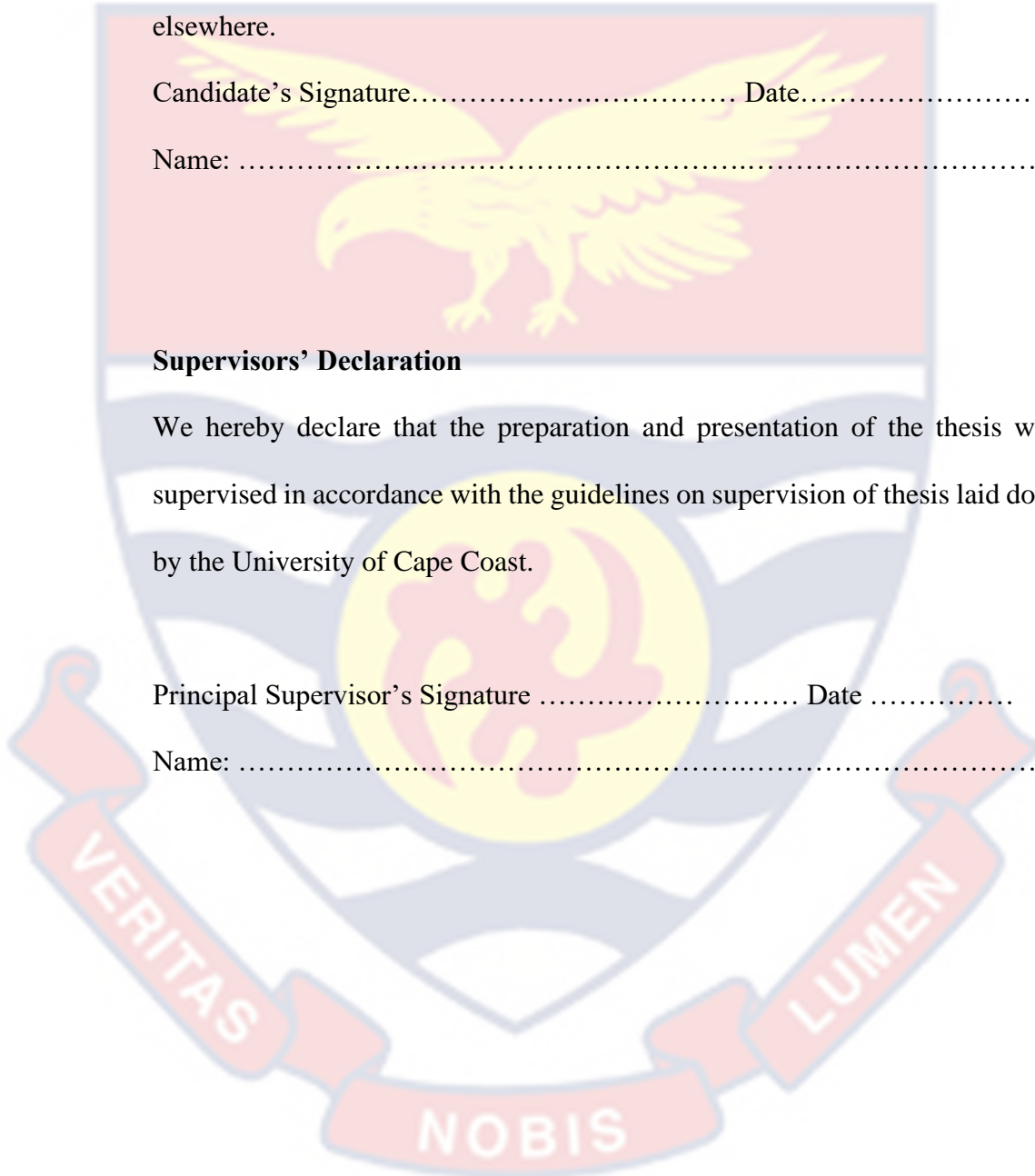
Name:

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:



ABSTRACT

The general purpose of the study was to assess Mathematics teacher's practice of achievement testing in Secondary Education Improvement Project (SEIP) schools in Upper West Region of Ghana. The study employed a quantitative approach where data was collected using adopted questionnaire. The quantitative data were analyzed using inferential statistics - ANOVA - and descriptive statistics - means, standard deviations, frequencies, and percentages. It became clear that mathematics teachers of Secondary Education improvement project schools hold positive view about achievement testing. For instance, most of them held the view that achievement testing establishes what students have learned, most believed that achievement test measures students' higher order thinking, again, majority held the view that achievement test results are trustworthy, also mathematics teachers in SEIP schools used various kinds of test strategies to assess students' outcome in schools. The results gave statistical evidence that there were significant differences in mean scores of the years of teaching experiences and the challenges encounter in practicing achievement testing. The study also established that there was no statistically significant difference in the influence of gender with regard to practicing of achievement testing. To strengthen teachers' knowledge on achievement testing, I recommend that more effective and efficient workshops and in-service training should be organized to teachers' secondary education improvement project schools with regard to achievement test practices.

KEYWORDS

Mathematics teachers' practice

Improvement project schools

Achievement testing,

Test construction

Secondary Education.



ACKNOWLEDGEMENTS

The product of a graduate degree is never the result of one person's effort, but rather the culmination of the goodwill, sacrifice and support of a select few. My appreciation and thanks go to my supervisor, Prof Mark Owusu Amponsah whose scholarly insights and academic rigor helped me to develop a better understanding of the issues involved in this thesis. I am particularly grateful for his wonderful support, patience and kindness over the past years, which have been important to my studies. I am deeply grateful to all the staff of the Psychology and educational foundation studies for all the knowledge, support and skills they have impacted in me in diverse ways.

I am very grateful to all the various senior high school headmasters and teachers for their tolerance and the willingness with which they answered my questions. The fellowship of a few other friends has carried me through the programme and the writing of this thesis. My heartfelt love to my family especially my parents appreciation go to you for your depth of understanding for making the ultimate sacrifice and tolerating the disruption in your lives that enabled me to undertake and complete this programme.

Finally, I give the ultimate acknowledgement and thanks to my heavenly Father for having sustained me through life. He deemed it appropriate to instill in me the energy and intellect without which I would not have been able to carry out this research successful.

DEDICATION

To my mother Bodon who devoted his time and other resources to support me completed this program to my father Youra Yoyuo and the entire family.



TABLE OF CONTENTS

Contents	Pages
DECLARATION	ii
ABSTRACT	iii
KEYWORDS	iv
ACKNOWLEDGEMENTS	v
DEDICATION	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER ONE: INTRODUCTION	
Background to the Study	1
Statement of the Problem	10
Purpose of the Study	14
Research Questions	14
Research Hypotheses	15
Significance of the Study	15
Delimitations	17
Limitations	18
Definition of Terms	18
Organization of the Study	19
CHAPTER TWO: LITERATURE REVIEW	
Introduction	21

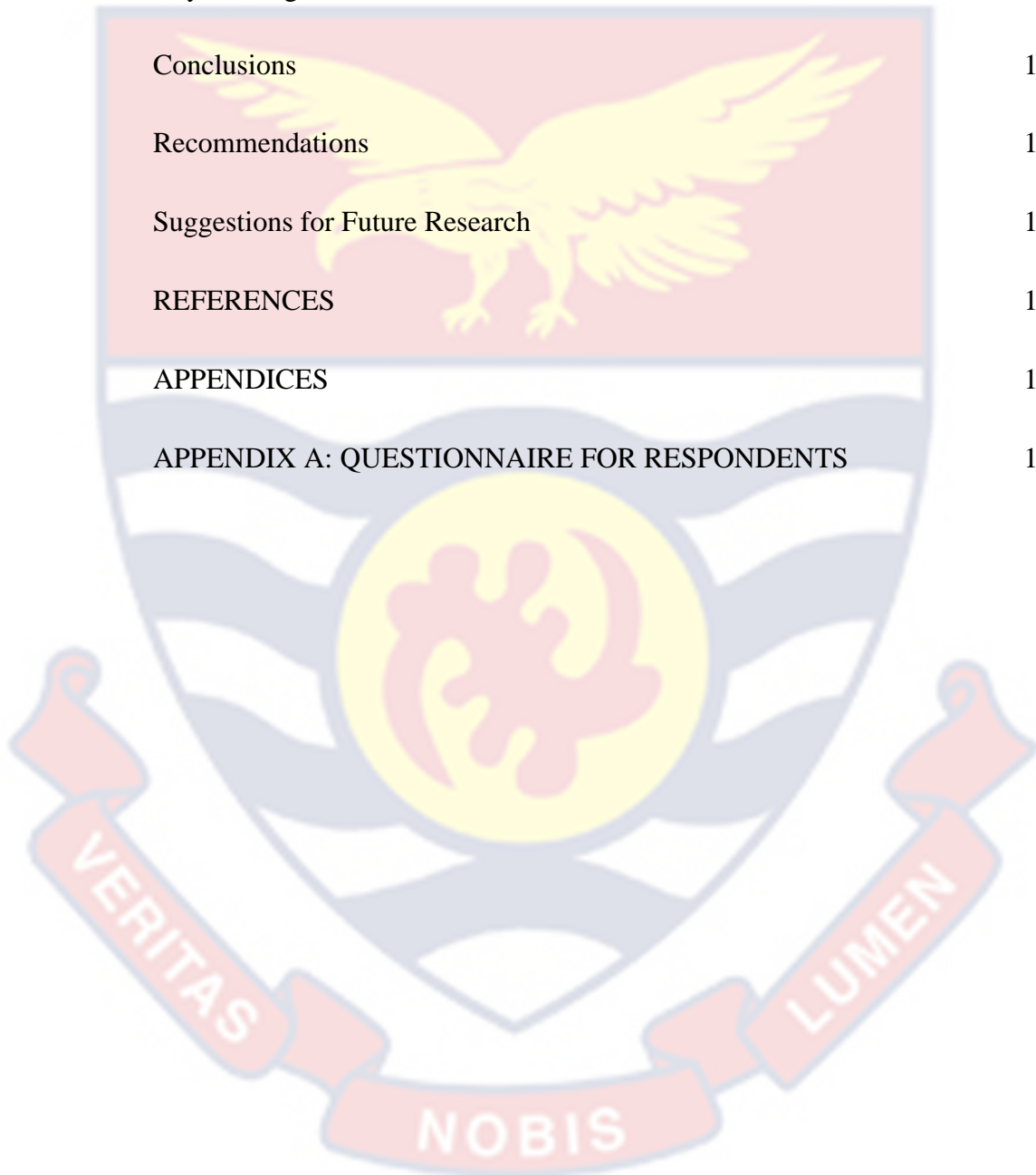
Constructivists' assessment	22
Examples of Constructivist Activities	23
Classical True Score Theory	24
Assumptions of the Classical True Score Theory	25
Conceptual Review	26
Concept of Teachers Test Practices	26
Teachers Years of Teaching Experience	32
Assessment Standards	34
Achievement Testing	36
Test	38
Teacher-Made Tests	42
Objective -Type Tests	44
Purposes of Achievement Test Results in Schools	48
Empirical Review	49
Teachers' views	50
Teachers Test Strategies	54
Challenges in practicing Achievement Test	55
Construction of Achievement Tests	57
Achievement Test Administration	64
Test Scoring	66

CHAPTER THREE: RESEARCH METHODS

Introduction	68
Research Design	68
Study Area	69
Population	70
Sampling Procedure	71
Data Collection Instrument	73
Instrument validity	74
Reliability of Instrument	75
Data Collection Procedures	76
Data processing and analysis	77
Chapter Summary	77
Data Processing and Analysis	78
Ethical considerations	78
CHAPTER FOUR: RESULTS AND DISCUSSIONS	80
Introduction	80
Presentation of Results	81
Description of Respondents	82
Analysis of the Research Questions	85

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction	111
Key Findings	112
Conclusions	113
Recommendations	114
Suggestions for Future Research	115
REFERENCES	116
APPENDICES	136
APPENDIX A: QUESTIONNAIRE FOR RESPONDENTS	136

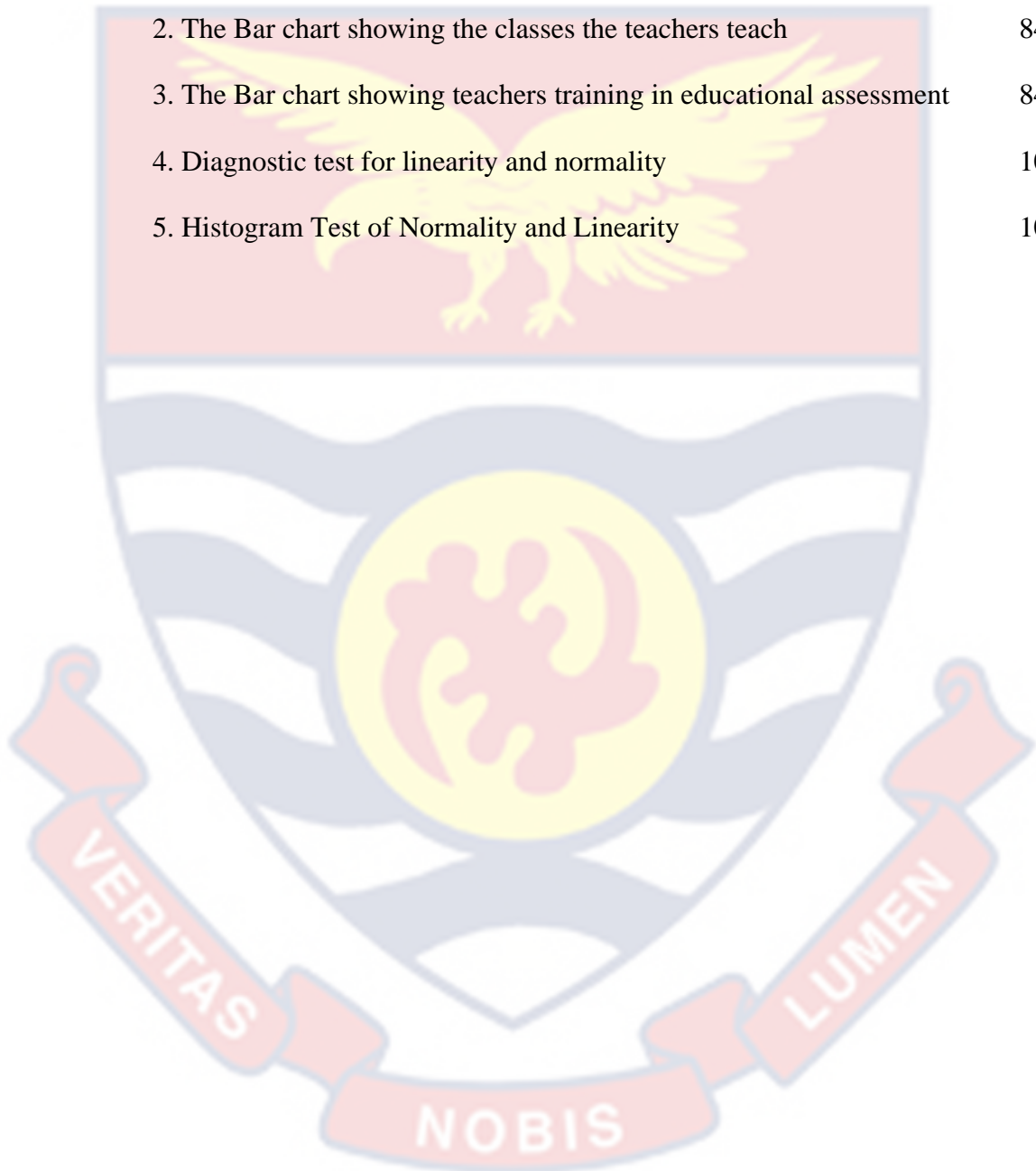


LIST OF TABLES

Table	Page
1: The enrolment of the sampled Mathematics teachers	73
2: demographic distribution of participants	81
3: Characteristics of Respondents	82
4: Mathematics Teachers View about Achievement Testing	86
5: Achievement test strategies	90
6: Results on challenges teachers encounter in practicing achievement testing	94
7: Results on how mathematics teachers in secondary education improvement project schools follow the principles of construction of test items.	97
8: Normality Test Results of the Variables	103
9: Homogeneity of Variances Test of the independent variables	105
10: Descriptive Statistics of the Study Variables (QoL*YiW)	106
11: Summary of One-way Analysis of Variance (ANOVA) Results	107
12: Post-Hoc Test among the years of teaching Experiences	108

LIST OF FIGURES

Figure	Page
1. The Bar chart showing the status (examiner or not examiner) of the teachers.	83
2. The Bar chart showing the classes the teachers teach	84
3. The Bar chart showing teachers training in educational assessment	84
4. Diagnostic test for linearity and normality	102
5. Histogram Test of Normality and Linearity	103



CHAPTER ONE

INTRODUCTION

Background to the Study

Teaching and classroom achievement testing are two important interrelated elements of instruction. The two are virtually inseparable and it is imperative that Mathematics teachers practice and follow the best principles in this regard. Many researchers in the area of achievement testing concentrate on the validity and reliability of achievement test items and seemingly neglect how the results can be used to improve students' learning.

Constructivists learning theory and those who believe in their dictates, propose the use of multiple forms of achievement testing rather than relying solely on traditional paper and pencil tests to form the foundation for teachers' achievement test practices. Constructivists suggest that all the domains of the student should be involved in learning and also encourage them to construct knowledge. The argument follows that proper achievement test practice is an influential factor to shape and modify classroom instruction which subsequently improves students' learning outcomes.

Different forms of achievement test strategies exist and are available to teachers to measure the knowledge of students in a content area and gauge their current level of learning outcomes. Performance, portfolios and traditional based achievement test formats are amongst the major types widely used (Brempong, 2019). Brempong (2019) posits that performance-based achievement tests require students to perform real-work tasks. This technique requires students to apply concepts they have learnt instead of merely knowing the concepts predetermined by the classroom teacher to be achieved. On the

other hand, traditional types of achievement tests measure general knowledge of learners and include test types such as short answer essays or constructed responses and multiple-choice questions (Brempong, 2019).

Portfolios represent a collection of students' products. Achievement tests can therefore be dichotomized into standardized achievement test and classroom (teacher-made) achievement tests (Armah, 2018). Further classification of achievement tests was made by Alrfooh (2012) into written - multiple choice tests and essay type tests-, oral tests and performance tests in the educational system. Studying in an educational system without being exposed to several educational and psychological procedures of achievement testing is impossible. Achievement testing must be carried out necessarily in an educational system in order to make purposeful and appropriate decisions on students, curricula and educational policies (Amedahe & Gyimah, 2014). Decisions about the student could be on managing classroom instruction, putting them into appropriate categories, providing guidance and counseling services, enlightening them on educational opportunities, and satisfying their competence. Decisions about curricula and programmes should be on their effectiveness - summative assessment - and proper ways of improving them - formative assessment. In Ghana, decisions about educational policies and their effectiveness are mostly determined at the national level (Oduro-Okyireh, 2008). It is worth mentioning, however, that in every educational system, including Ghana, educational assessments of which test predominates, provides the needed information for the necessary decisions (Oduro-Okyireh&Partey, 2014; Adom, Mensah, & Dake, 2020).

Tests are strategies or procedures mostly used to determine the ability of students to complete given tasks, demonstrate mastery of a skill and or knowledge of content (Adom, Mensah & Dake, 2020). However, it must be stated that measuring the psychological attributes of an individual cannot be done directly as in the case of height and weight. The existence of such psychological attributes or constructs can never be absolutely determined. The degree to which any attribute characterizes an individual can only be inferred from the observation of his or her behavior. Setting standards for subjectively measuring an individual's construct or traits quantitatively becomes more prudent. Test is therefore, an essential tool in the assessment process that can help quantify the degree to which a construct exist in an individual. A number of achievement test tools or techniques in the educational setting are used to collect information about student(s). These techniques include oral and written examinations, true-false and multiple-choice tests, matching tests, fill-in-the-blanks examinations, short answer tests and open ended questions(Kara & Celikler, 2015). It is worth noting that in the educational setting, teacher -made tests or classroom achievement tests dominate the other assessment tools (Quansah, Amoako, & Ankomah, 2019). According to Cizek (1998), achievement test in America plays a very crucial role in both elementary and secondary education. He contends that valuable decisions in both educational and occupational spheres rely heavily on achievement test. The situation in America is not different from Ghana; however, what goes into achievement tests is solely determined by the classroom teacher in Ghana (Adom, Mensah&Dake, 2020; Armah, 2018; Oduro-Okyireh, 2008).

In the classroom situation, students learn various contents and it is the basic requirement of teachers to test students on these content areas (Adom, Mensah and Dake, 2020). It is therefore critical that teachers possess not only comprehensive knowledge of their content areas but also the ability to appropriately test what students have achieved. It can therefore be inferred that the most important service teachers need is to teach them techniques of test construction, administration and scoring based on best principles of testing. According to Ebel (1972), the increasing use of educational tests is due to growing criticism of their practice. The quality of the test varies, some are particularly bad. He argued that educational tests can be socially harmful for variety of reasons. First, children are labeled as potentially tarnished and discouraged. Ebel (1972) argued that tests should not be judged by how reliable their later predictions are, but by how much performance is improved by encouraging and directing students and teachers effort. Another criticism is that achievement test encourages development of a single capacity, and reduces the diversity of talent within society. Brady (1997) suggested that, although achievement test practices is intended to support the curriculum, there is a challenge that it may come to dominate the curriculum because what is evaluated is taken as an indication of what is important. Ebel (1972) argues that tests generally lag rather than lead curricular change.

Assessment is a broad term that includes the use of various strategies and ways to investigate the extent to which students are achieving the predetermined learning aims and outcomes of a lesson (Mussawy, 2009). There are different kinds of assessment that can be used to test students' knowledge and see their current levels in specific subjects. Two major types of assessment widely used

are traditional types of assessment and performance-based assessment (Birenaum, & Feldman, 1998). Birenaum and Feldman, (1998) argued that traditional types of assessment tools are generally knowledge-based and include conventional types of tests such as multiple-choice questions, short answer essays or constructed responses and standardized tests whereas in performance-based assessments, students are required to perform a real task rather than merely selecting from options provided; and students are evaluated according to their performance outcomes and the extent to which those outcomes are about the rubrics or feedback tools. An achievement testing requires students to exhibit the status of their learning through a demonstration of mastery (Poikela, 2004).

Nobody can study in an entire educational system without being exposed to a wide range of educational and psychological assessment procedures. This is because frequently in an educational setting, decisions have to be made about students, curricula and programmes, and educational policies. According to Nitko (1996), decisions about students include managing classroom principles, placing students into various types of programmes, assigning them to appropriate categories, guiding and counseling them, selecting them for educational opportunities, and credentialing and certifying their competence. Decisions about curricula and programmes include decisions about their effectiveness (summative assessment) and about ways to better them (formative assessment). It is worth knowing, however, that educational measurements, of which in the Ghanaian educational system, tests exist, provide some of the needed information for these types of decisions (Ministry of Education, 2010). According to the standard for Educational and Psychological testing,

National Council on Measurement in Education (NCME, 2014) a test is a procedure in which a sample of an examinees character in a specified domain is acquired and subsequently evaluated and scored using a standardized process. However, it must be noted that the psychological traits of an individual cannot be measured directly like the measurement of height or weight. The existence of such a psychological construct can never be confirmed. The degree to which any trait characterizes an individual can only be assumed from observation of his or her character.

It becomes more prudent if one can quantitatively relate the subjective judgments of individuals about the estimated amount of constructor trait that exists in a person by establishing standards for such measurement.

A test is an important tool that helps to quantify such attributes which aids one to make a value judgment about the extent to which such constructs might probably exist in an individual. A large number of evaluations techniques are being used to collect information about students. These include formal and informal observation of students, paper-and-pencil tests, a student's performance on homework, laboratory work, and projects during oral questioning and analysis of students' records (NCME, 2014). Teachers in the educational setting would want to predict the extent to which their students are characterized by the knowledge they have imparted to them within a given period. All the domain of such construct might not be known by a single test. Nevertheless, a well-constructed test could sample to a large extent a reasonable amount of the construct on which value judgment could be made from. Educators and teachers must also be aware that a test itself is subject to errors that adversely could affect its use in deciding for students. Tom and Gary (2003)

further suggested that test misuse and abuse can occur when users of test results are unaware of the factors that can influence the usefulness of the test scores. Among the major factors are the technical sufficiency of a test and its validity and reliability. The technical insufficiencies might emerge from factors such as test appropriateness for testing, the content validity evidence, the appropriateness of the test strategies used, years of teaching experience, views of teachers, test construction and administration guidelines and gender influence of mathematics teachers are factors that may have affected achievement test practices, among others. It must also be noted that even when a test is technically sufficient, misuse and abuse can occur because technical adequacy does not ensure that test scores are accurate or meaningful. When students' achievement levels are not properly measured and interpreted, the teachers and school administrators are not able to provide the right educational opportunities and support each student needs. Testing provides results on which educational decisions are made. These decisions may be the ones that require information about the success of learning programmes or about students who have reached particular levels of skill and knowledge (Izard, 2005). Reliable and valid information about student achievement is widely understood to be vital for effective instruction, as it enables teachers to give appropriate results and adapt their instruction to match student needs. The importance of achievement test practices cannot be underestimated; first tests are used to investigate academic progress. Again, it is used for selection decisions, sometimes; an institution decides whether some teachers are qualifying for award and promotions. Those not acceptable are rejected and are no longer the concern of the institution (Cronbach, 1960; Nitko, 2001; Amedahe & Asamoah

Gyimah, 2003). An educational institution often uses test results to provide part of the information on which selection decisions are based. Tests and teaching are interwoven. Quaigrain (1992) has stated that tests provide needed information for evaluation. The basic principles for the construction of teacher-made tests have been developed over the years by several educational measurement experts (Amedahe, 2000). While some of the test construction principles are general and apply to any type of test, others are specific and apply solely to the particular type of test under construction. From the available literature, the test construction principles that the researcher judged as most simple and practicable in the achievement testing situation were those suggested by Tamakloe and Amedahe (1996) and Etsey (2004). These are in eight steps. The steps are: define the objective of the test, review the item format to use, find out what is to be tested, write the individual items, review the items, prepare the scoring key, write instructions, and evaluate the test. It agreed that a good test must follow these steps to improve its validity and reliability. According to Mehrens and Lehmann, noise and distraction in the testing environment should be kept at the barest minimum if not eradicated. The distractions from outside can change the attention of test-takers which could contribute to the low performance of students. Amedahe & Asamoah-Gyimah (2003) & Etsey (2004) found that tests must not be given immediately before or just after a vacation, holidays or other important events where students are involved in either physically or psychologically activities. Amedahe & Asamoah-Gyimah (2003) opined that tests must also not be given when students would normally be doing something pleasant such as having co-curricular activities as this will inveigle students' concentration. On test scoring, Amedahe & Asamoah-Gyimah (2003);

and Etsey (2004), who agreeably suggested that responses of the item should be scored item by item rather than script by script? This principle is to reduce the carryover effect on the scores and thereby ensuring consistency. A study by Amedahe (1989), recounted that teacher in the schools used mainly the analytic method in scoring their essay-type tests. He further asserted that teachers in the schools scored their essay-type tests either item by item or script by script. On the part of Quaigrain (1992), he found that most of teachers in the schools used the analytic method in scoring their essay-type tests. Amedahe & Asamoah-Gyimah (2003) and Etsey (2004) indicated that the skills of using methods such as correct teaching and learning materials usage, the teachers' ability to control the class, guard students to solve mathematical problems and frequent use of test strategies, quality of construction and administration, be judged separately from subject matter correctness. This was because when teachers are influenced by factors other than the subject matter, the marks awarded would represent attribute irrelevant or construct representativeness. This simply meant that higher scores on tests might not reflect the ability of students on the subject matter but rather discriminate students in proficiencies they have over other students. Etsey (2004) also indicated that scripts must be scored anonymously. He suggested scripts should be identified by code numbers or any other means instead of the names of students. This principle is to reduce the halo effect. Halo-effect happens when a scorer's general impression of a person influences how the paper is scored. The theories that underpinned this study were the Constructivist Learning Theory and Classical True Score Theory. Constructivist learning theory says that all knowledge is constructed from a base of prior knowledge (Davis, 2000). According to Vygotsky (as cited in Davis 2000),

children are not tabular raster and knowledge cannot be imparted without the child making sense of it according to their current conceptions; therefore, children learn best when they are allowed to construct a personal understanding based on experiencing things and reflecting on those experiences. In the study of Amedahe (2000), it was evident that most Ghanaian teachers had limited skills for constructing achievement tests, the objective and essay type tests, which are the most frequently, used instruments in schools in Ghana. According to Amedahe teachers' limited skills in test construction were the result of lack of training in assessment techniques, large class size and lack of a particular school's policy in assessment standards with implications on validity and reliability of the assessment results. Given the extent of prevalence of teachers' achievement tests in Ghanaian schools and the variety of uses to which the results from these tests are put, there is the need for research into the achievement testing practices of teachers. Again, careful observation of the secondary education improvement schools in the upper west region easily reveals that the achievement test practices of the teachers were with a lot of challenges which may include poor construction of test items, poor administration, poor strategies that one begins to wonder whether training contributes to competence at all. It is based on the above problem this study sought to evaluate the achievement testing practices of teachers in secondary education improvement schools in upper west of Ghana.

Statement of the Problem

Test practices at the senior schools shows that, most teachers have had training in achievement testing as part of the evaluation process at their various colleges of education and universities. Previous research has shown that most of the teachers in the second cycle institutions in Ghana lacked the basic test practices skills. This was justified by the findings that not all teachers in the Secondary Schools in Ghana have undergone professional training in testing techniques (Amedahe, 1989). The studies by Amedahe (1989) and Quagrain (1992) indicated that most Ghanaian teachers had insufficient skills for constructing the objective and essay type tests, which are the most frequently, used instruments in our schools. The study of Amedahe (1989) showed that, to a larger extent, secondary school teachers in the upper west Region did not follow the basic prescribed principles of achievement testing. Quagrain replicated the study of Amedahe in 1992 and confirmed the report of Amedahe. This is because most initial teacher training programmes do not make sufficient provision for a course in testing. Amedahe (2000) stated that “teacher –made tests may be made of a number of factors, notably among them are, training in assessment techniques, class size and a particular school’s policy in assessment with implications on validity and reliability of the assessment results” (p. 112-113). On contrary to those previous studies, Oduro (2000) concluded in his study that to a larger extent, teachers followed the basic principles in test construction, administration, strategies, and scoring. The findings of the study of Boakye (2016) also showed that teachers to some extent follow the basic principles of test practices. Could it be seen that because these two studies Oduro (2000) and Boakye (2016) were conducted at Volta Region. In contrast, Sasu (2017) repeated the same study in upper East and found out that teachers,

to some extent, have little knowledge in achievement testing which was in support of the studies by Amedahe (1989) and Quagrain (1992). Gleaning from the literature, it was evident that the numerous studies in teachers test practices have concentrated attention on teachers in the northern part of Ghana; however, in the case of those in the southern part of the country; it indicates much have not been documented. Interestingly, as I went around and interacted with some headmasters and Directors of Education in some districts in the northern part of the country, it shown that most teachers within some of the districts in the Northern part of Ghana are not professionally trained teachers. Suprisely, these teachers construct, administer, score, and interpret results of their students. The question that really comes to mind is the soundness and appropriateness of results from these assessment results. It is pertinent to examine the achievement testing practices among teachers in the upper west region, since this would bring to bear the extent to which mathematics teachers' practices, and this would help identify the mistakes in achievement test practices in order to provide a remedy to the situation, Hence, the need to conduct this study in a different region to help add more knowledge to teachers test practices in Ghana. The formative and summative assessments are not consistent with accepted theories of learning, as well as with socially valued learning outcomes. Gitomer & Duschl (2007) describe two types of coherence, internal and external. Assessment components can be considered internally coherent when they are mutually supportive; in other words, formative and summative assessments need to be conformed to one another. The effectiveness of assessment would be restricted by the nature of the larger system in which it is surrounded and, particularly, by the content, format, and design. Even though

students' achievement levels or performance is as a result of many factors, one cannot down play the influence of the teacher and classroom engagement. Both professional and non-professional teachers instruct and carry out achievement testing in schools. Blatchford and Cline (1992) suggest that the achievement tests should practice without bias with respect to gender, social class, ethnicity, language use and religion. However, many researchers have exhibited the existence of bias in achievement testing practices among mathematics teachers in the Upper West Region of Ghana. Many researchers have argued that most achievement test practices are culture biased and discriminate against certain ethnic groups. Anastasi (1972) argues that it is not productive to attempt to develop tests that are „culture-free“ (free from cultural influences) and, instead, there should be efforts to develop tests that are „culture-fair“ (common to different cultures).

Achievement testing serves many objectives but improving learning outcomes is the ultimate goal (Azis, 2012; Adeyemi, 2008; Guskey, 2003). There is therefore the need for all classroom teachers to know the principles of achievement testing but teachers' achievement test practices have been questioned by various researchers and other stakeholders who matter in this regard (Adom, Mensah & Dake, 2020; Quansah, Amoako & Ankomah, 2019; Adeyemi, 2008; Guskey, 2003). The implication is that a weak correlation exists between internal scores of students and scores from external examinations. When this happens the purpose of using achievement tests to improve students' learning outcomes is not achieved. Any other decisions that will be made based on the results of students may be misleading. This prompted the researcher to

examine the classroom achievement test practices of Mathematics teachers in SEIP schools in the Upper West region of Ghana.

Purpose of the Study

The purpose of this study was to assess mathematics teachers' practices of achievement testing in secondary education improvement project schools in the upper west region of Ghana.

Specifically, the study sought to:

1. find out views Mathematics teachers' hold about achievement testing in SEIP schools in Upper West region.
2. find out the kind of achievement test strategies Mathematics teachers in SEIP schools use to determine students learning outcomes.
3. investigate the challenge mathematics teachers in upper west region encounter in practicing achievement testing/
4. evaluate how teachers follow guidelines to test construction in secondary improvement project schools.
5. Investigate how years of teaching experience influence challenges teachers encounter in quest of practicing achievement testing in Secondary Education Improvement Project Schools in the Upper West Region of Ghana.

Research Questions

The study was guided by four research questions. They are:

1. What views do Mathematics teachers in Secondary education improvement project schools hold about achievement testing?
2. What kinds of achievement test strategies do Mathematics teachers in Secondary education improvement project schools use to assess students learning outcomes?
3. What challenges do mathematic teachers encounter in practicing of achievement testing?
4. How do teachers in senior high schools follow guidelines to test construction?
5. How do years of teaching experience influence the challenges teachers encounter in practicing achievement testing?

Research Hypotheses

1. H0₁: There is no statistically significant difference in the views Mathematics teachers in SEIP hold about achievement testing.
2. H0₂: There is no statistically significant difference in achievement test strategies use in SEIP schools by Mathematics teachers to assess their students learning outcomes.
3. H0₃: There is no statistically significant difference between in challenges encounter in quest of practicing of achievement testing.
4. H0₄: There is no statistically significant difference in the principles to test construction teachers follow in practicing of achievement testing.
5. H0₅: There are no statistically significant differences in years of teaching experiences and its influences on the challenges teachers encounter in quest of practicing of achievement testing.

Significance of the Study

The results that were gathered from the study would help stakeholders to determine the state of affairs with regard to achievement testing in the Ghanaian educational system. This, it is believed, it would aid teachers who received instruction in assessment in education to be well vest and put their acquired knowledge into practice since testing principles would be related to practice throughout the study. Positive suggestions would be offered as a means of addressing these problems. It is hoped that these suggestions would help all teachers to improve on their testing practices. The results of the study would help to enlighten the mathematics teachers in the upper west region on their knowledge of assessment in general and achievement test in particular. The findings of this study would help curriculum developers, educators and teachers to understand the impact of teacher's views of achievement tests on instructional practices, student's performance and the objective of education. Specifically, the findings of this study would inform teachers about the value and impact of achievement test tasks on their instruction. The results of this study would provide insight for curriculum developers, educators and teachers regarding the challenges impeding the effective use of achievement test for appropriate intervention. The research would benefit government, teachers, students, district education offices, curriculum developers, test experts, departments of Mathematics in the various universities and other researchers. Also government would get data on the state of affairs of Secondary education improvement project schools in Ghana in terms of students' performance especially in Mathematics. This would help government to assess whether the resources being channeled into these schools are yielding the desired results. There may also be the need for government to sponsor some Mathematics

teachers to study courses on assessment and evaluation in Mathematics education at the masters and PhD levels in order to improve their testing skills. The district education offices within the study area would get to know how students' achievement levels are assessed and the challenges teachers encounter. This can be achieved through the organization of in-service training for teachers on the importance of regular achievement testing, construction of good achievement test items, how to best administer and use achievement test results to improve the teaching and learning of Mathematics. Researchers interested in this area can use the findings of the study as the baseline for other researches. The ultimate beneficiary will be students since they will be motivated and interested in studying Mathematics. , the study added to literature, has provided valuable information for researchers who might want to undertake a study in a related.

Delimitations

The study was conducted to only the senior high schools in the Upper West Region. It concentrated on only the Secondary Education Improvement project schools in the Upper West Region of Ghana. Also, it was delimited to only teachers teaching Mathematics. More so, the study concentrated on only teachers' achievement tests. Finally the study concentrated on only three aspects of teachers' achievement test practices; construction, administration, strategies, challenges, years of teaching experience and views held by mathematics teachers. Out of a total of 30 Senior High Schools in the Upper West region, there are seven Secondary Education Improvement Project schools. The study involved only six SEIP schools because the seventh one is considered fledgling.

All the Mathematics teachers of these schools were involved in the study. Students and school administrators were excluded from the study.

Limitations

A questionnaire was used for the data collection. Therefore, the choice of respondents providing responses to some of the questions, maybe, without correct understanding of the questions was high. Hence, the possibility of introducing errors into the findings of the study would be possible. Another limitation of the study was the likelihood of respondents giving socially desirable responses to the questions on the questionnaire, and that therefore, could affect the results of the study as well as the interpretations and its uses. The findings of this study pertained to only Mathematics teachers in SEIP schools in the Upper West region, thus generalizing the findings to all teachers was seen as the weakness of the current study. This is because teachers in the various regions in Ghana were different based on characteristics such as economic, cultural, social and educational background of parents. Again, the study relied on self-reported data by respondents which was likely to be inaccurate. Six SEIP schools were involved in the study and some teachers in these schools may decide not to take part causing the number of respondents to fall below. Also, using only six SEIP schools out of the 30 SEIP schools in Ghana was statistically insignificant to make generalizations.

Definition of Terms

SEIP schools: They are some Senior High Schools in Ghana with persistent poor performance in both Mathematics and Science over the years.

Teachers test practices: as reflected in such activities as stating desired learner outcomes, grouping pupils, instigating study activities, and providing feedback for monitoring teaching and learning, are an integral component of models of instruction (Brophy & Good, 1986; Rosenshine, 1985).

Achievement testing: refers to any procedure or instrument that is used to measure an examinee's attainment of knowledge or skills. Achievement testing can be done informally, as in when a teacher asks a student to perform a skill such as reading aloud or demonstrating correct laboratory technique. More formal, and perhaps more common, achievement tests are routinely administered in educational and occupational settings.

Secondary education: refers to the stage of formal education that follows primary education and precedes higher education. It is typically offered to students between the ages of 14 and 18, although the specific age range may vary depending on the educational system and country.

Test: is a device or technique used to measure the performance, skill level, or knowledge of learner on a specific subject matter.

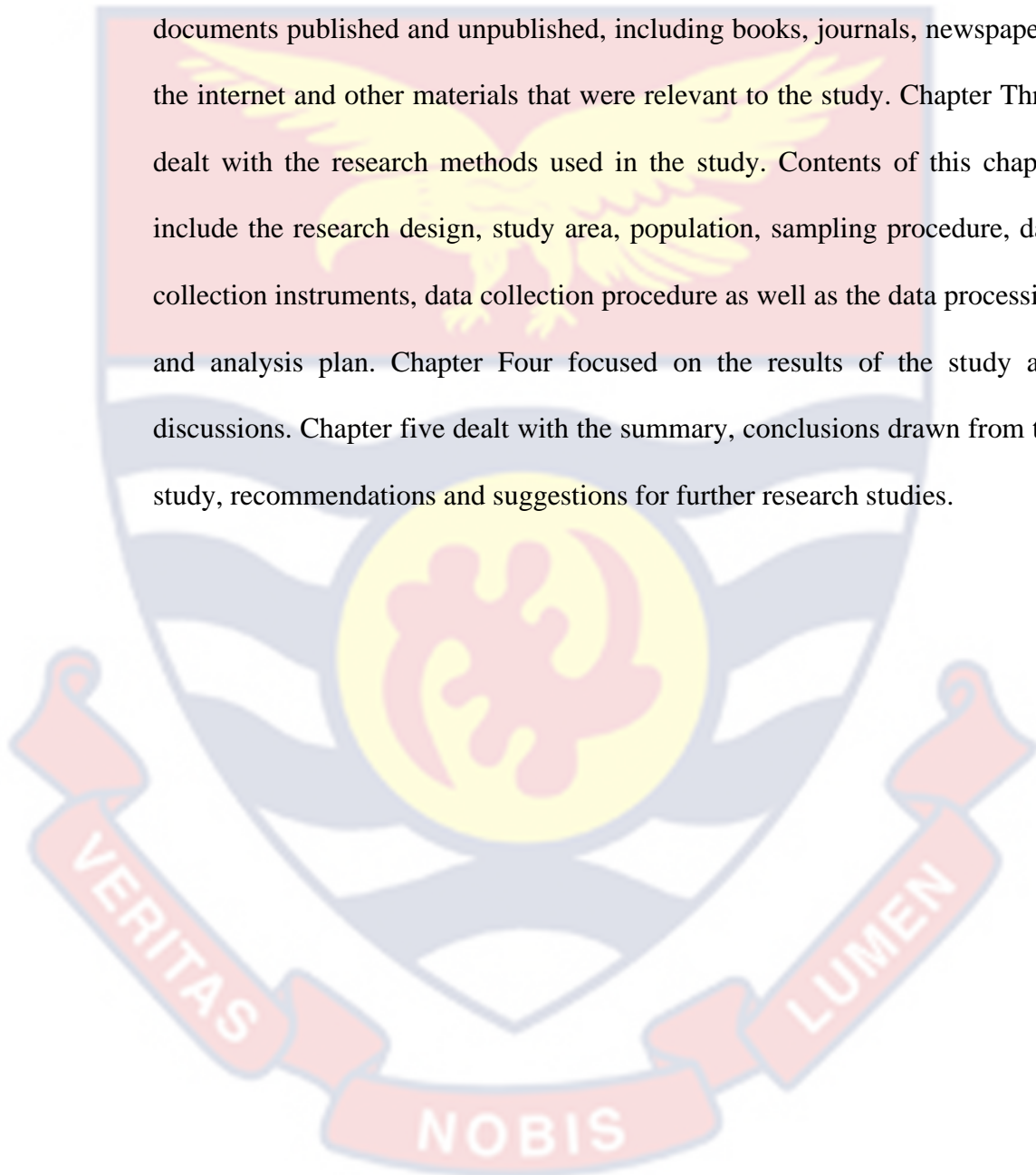
Assessment: is an inferential process because others cannot know with certainty what understanding exists inside a student's head.

Assessment Practice: is a manner of conducting assessment. These are teachers' actual application or use of idea, belief or method of assessment as opposed to what is meant or believed to happen

Organization of the Study

The study was organized into five chapters. Chapter one consists of an introduction to the study; the background of the study, statement of the problem,

the purpose of the study and objectives of the study. In addition, the research questions, significance of the study, delimitation, limitations, definition of terms as pertains to the study as well as organization of the study, are described. Chapter two dealt with the review of related literature to the study from documents published and unpublished, including books, journals, newspapers, the internet and other materials that were relevant to the study. Chapter Three dealt with the research methods used in the study. Contents of this chapter include the research design, study area, population, sampling procedure, data collection instruments, data collection procedure as well as the data processing and analysis plan. Chapter Four focused on the results of the study and discussions. Chapter five dealt with the summary, conclusions drawn from the study, recommendations and suggestions for further research studies.



CHAPTER TWO

LITERATURE REVIEW

Introduction

The major aim of the study was to investigate mathematics teachers' practices of achievement testing in secondary education improvement project schools in the upper west. This section reviewed the literature related to the topic. The study comprises of the conceptual review, theoretical review and empirical review

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

James and Pedder (2006) also state that the focus of constructivists is on how people construct meaning and make meaning out of the world through organizing structures, concepts and principles in schema (mental models).

According to James and Pedder (2006), prior knowledge is regarded as a powerful determinant of a pupil's capacity to learn new material. He then indicates that cognitive constructivists emphasize 'understanding,' thus problem solving is seen as the context for knowledge construction. Davis (1991) again argues that processing strategies, such as deductive reasoning from principles and inductive reasoning from evidence, are important and as a result, differences between experts and novices are marked by the way in which experts organize knowledge structures and their competence in processing strategies. Torrance and Pryor (2001), point out that the interaction between teacher-pupil goes further than just finding out whether the pupil has reached the target behavior, as in behaviorism. Teacher-pupil interaction in a test situation goes beyond the discussing of test results, the judgments of progress and the provision of additional instruction, to include a role for the teacher in assisting the pupil to understand and engage with new ideas and problems (Torrance & Pryor, 2001). To them, the process of evaluation itself is seen as having an impact on the pupil, as well as the product or the result. Harlen (2006) stated that the constructivists' view of learning concentrate attention on the processes of learning and the duty of learners. Teachers engage pupils in self-evaluation and use their own evaluation to try to identify their current understanding and levels of skills.

Constructivists' assessment

Traditionally, assessment in the classrooms is based on testing thus it is important for the student to produce the correct answers (Davis, 1991). However, he further posits that in constructivist teaching, the process of gaining knowledge is viewed as being just as important as the product. Thus, assessment

is based not only on tests, but also on observation of the student, the student's work, and the student's points of view (Davis, 1991). According to Davis (1991), some constructivists' assessment strategies include:

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

Examples of Constructivist Activities

The constructivist classroom, students work primarily in groups and learning and knowledge are interactive and dynamic (Harlen, 2006). Davis (1991) states that with the constructivist classroom, there is a great focus and emphasis on social and communication skills, as well as collaboration and exchange of ideas which is contrary to the traditional classroom in which

Students work primarily alone, learning is achieved through repetition. He further argues that the subjects are strictly adhered to and are guided by a textbook. According to Gielen, Dochy and Dierick (2003), some activities encouraged in constructivist classrooms are:

1. Experimentation: Students individually perform an experiment and then come together as a class to discuss the results.
2. Research projects: Students research a topic and can present their findings to the class.

3. Field trips. This allows students to put the concepts and ideas discussed in class in a real-world context. Field trips would often be followed by class discussions.

4. Films. These provide visual context and thus bring another sense into the learning experience.

5. Class discussions. This technique is used in all of the methods described above. It is one of the most important distinctions of constructivist teaching methods.

Classical True Score Theory

A test theory or test model is a symbolic representation of the factors influencing observed test scores and is described by its assumption. Classical true score theory is a simple, model that describes how errors of measurement can influence observed score. Classical true score theory states that an observed score (X) is equal to the sum of a true score, or true underlying ability (T), and the measurement error (E) associated with estimating observed scores, or $X = T + E$. It is believed that when students take a particular test measuring a construct twice in a succession, it is unlikely that their scores will affect their performance. This is due to the effect of some factors such as, fatigue, guessing, careless marking, or miss scoring. A different form of test would also result in a change in scores because of variation in content. These inconsistencies in individual scores due to the sampling of tasks or occasions must be regarded as measurement error, (Crocker & Algina, 2008). According to Crocker & Algina, the “True Score” can be interpreted as the average of the observed scores obtained over an infinite number of repeated testing of the same test. In the classroom setting, the “true score” is the score a teacher would obtain if he is to

take the average score from an infinite number of test administrations. Of course, in practice, one cannot administer a test an infinite number of times, and as noted previously, the vast majority of the time we get only one chance. Therefore, we use reliability coefficients to estimate both true and error variance associated with our observed test scores (Crocker & Algina, 2008). Several assumptions are made about the relationship among these three components (True Score, Observed Score and Error Score). Most of the standard procedures for creating and evaluating teacher made test are based on a set of assumptions on the Classical true-score theory. The model assumes certain conditions to be true; if these assumptions are reasonable, then the conclusions derived from the model are reasonable. However, if the conditions are not reasonable, then the use of the model leads to faulty conclusions.

Assumptions of the Classical True Score Theory

1. $X=T+E$ states that, the observed score “X” is the sum of the True score “T” and the error of measurement “E” 163. $\epsilon(X)=T$. This states that the expected value (population mean “ ϵ ”) of “X” is “T”. This assumption is the definition of T: T is the mean of the theoretical distribution of X scores that will be found in repeated independent testing of the same person with the same test $\rho_{E1T} = 0$. This assumption implies that examinees with high true score do not have systematically more positive or negative error of measurement than examinee with low true score. This assumption will be violated if for example, one administration of entrance exams, students with low true scores copied answers from those with high true scores; this situation will create a negative correlation between true score and error score 4. $\rho_{E1E2} = 0$, where E1 is the error score for Test 1 and E2 is the error score for Test2. This assumption states

that, the error scores of two different tests are uncorrelated. That is if a person has a positive error score in Test 1, he or she is not more likely to have a positive or negative error score in test 2. This assumption is not reasonable if the test scores are greatly affected by factors such as fatigue, practice effect, the examinee's mood, or effects of the environment. $\rho_{E1T2} = 0$; this assumption states that, the error scores on one test (E_1) are uncorrelated with the true scores on another test (T_2). This assumption would be violated if Test two measures personality trait or ability dimension that influences error on Test one. The assumption would also be violated if students with low true scores copied answers from those with high true scores. If two tests have observed score X and another X' that satisfies assumption one and if, for every population of examinees $T = T'$ and varies of $\sigma^2 E = \sigma^2 E'$ then the test are called parallel test. For, $\sigma^2 E$ equal to $\sigma^2 E'$ the condition leading to error of measurement, such as mood, and environmental effect, must vary in the same way for the two tests. If two tests have observed scores X_1 and X_2 that satisfies assumption one, and if, for every population of examinees, $T_1 = T_2 + C_2$, where C is a constant, then the test are called τ -equivalent test (Allen, & Yen, 2012, pp. 56-59). The implication of this theory therefore, means that in order to achieve the reliability and validity of teacher made test, the principles of the theory needs to apply.

Conceptual Review

Concept of Teachers Test Practices

Teachers' testing practices, as reflected in such activities as stating desired learner outcomes, grouping pupils, instigating study activities, and providing feedback for monitoring teaching and learning, are an integral component of models of instruction (Brophy & Good, 1986; Rosenshine, 1985).

The testing and evaluation process within learning models is variously described as providing practice, review, consolidation of learning, knowledge of results, feedback for redirecting efforts, feelings of accomplishment, a focus for efforts, etc. Relatedly, Crooks (1988) asserts that testing is one of the most potent forces influencing education. Elton and Laurillard (1979), also in describing the impact of achievement testing upon pupils, stated that the surest way to change pupil learning behavior is to change pupil assessment. Contrary to the common views that testing plays an important role in the teaching and learning process, actual elements of the evaluation schemas that teachers institute have received less research attention than most other aspects of education (Crooks, 1988). Further, the research of testing has been focused primarily upon standardized testing rather than upon the much more prevalent teacher devised testing, and those studies that have addressed teacher-made tests and teachers' testing practices have predominantly used teacher self-report data-gathering procedures. As a result, these limited and narrow research efforts have brought in testing professionals knowing little about the nature and quality of teacher-made tests, about how these tests are used within the classroom teaching-learning process, and about the adequacy of teachers' testing knowledge and skills (Stiggins, Conklin, & Bridgeford, 1986). The aim of this chapter is to provide a review of the professional literature devoted to testing in the SEIP schools in order to investigate what testing knowledge and skills mathematics teachers sought to have; what testing practices ought to be used to facilitate improvement of learning; what is known about teachers' actual testing knowledge, skills, and practices; and what impacts for the measurement

profession are suggested by any inconsistency identified between teachers' desired and actual testing knowledge, skills, and practices.

The researcher provides more complete discussions of several of the topics presented in this chapter and that a few rather extensive literature reviews of these or closely related topics also exist (e.g., the reviews provided by Balch, 1964; Bangert-Downs, Kulik, & Kulik, 1988; Crooks, 1988; and Kulik & Kulik, 1988). Tests guide and instigate effort: It is rather clear from the research on the impact of teachers' test practices; often involving interviews of students that students study is instigated by an announced test and is concentrated particularly upon content that they anticipate will appear in the test. In regard to this impact of teachers test practices, Rogers (1969) stated that teachers testing practices inform academia of the real aims of improvement of achievement testing in SEIP schools. Guiding of students study efforts toward content that is tested may have desirable or bad effects upon learning, depending upon how effective the test guides students to correct outcomes. In order for teachers to adequately practice achievement test in SEIP schools, the testing community advises teachers to use test specification tables to better link test questions to desired learner outcomes. This matching of test items with desired outcomes frequently is not done, and the resulting absence of match between content of standardized exams guidelines and more significant course content is often recognized by both teachers and students. A number of researchers have reported that teacher's test practices differ in a form of study when informed of the types of test questions to appear in a scheduled achievement testing. Balch (1964), after a review of teacher-instigated testing studies, revealed that students' awareness of the nature of the achievement test to be administered and the provision of

feedback regarding students' performance following a test are the two most potent testing variables influencing achievement testing. The researcher described achievement test strategies as concentrating on details when preparing for objective tests, and as searching for relationships and main points when preparing for essay tests. In response to this research, testing specialists commonly advise teachers to use a variety of question types on their achievement testing, when appropriate for the content to be examined, to encourage students to use more varied study patterns. Bangert-Downs, Kulik, & Kulik (1988), after reviewing a number of achievement test practices, testing frequency, concluded that students in classes with no tests scheduled were clearly disadvantaged, that moderately frequent tests appear to best facilitate teachers achievement test practices, and that as test frequency in a course increases teachers' skills in practicing achievement testing. Test frequency also facilitating effect of frequent testing upon teacher achievement appears to be consistent across subject content fields, to be more beneficial for less able teachers than for more able teachers, and to be more beneficial under certain testing conditions, such as the provision of feedback related to students' performance on tests following the examination period.

Research suggests that carefully administered and monitored tests, for which content and format are described to students prior to administration typically produce higher students' performance, less students cheating, and reduced students test anxiety (Bushway & Nash, 1977; Carrier & Titus, 1981; Hill & Wigfield, 1984; Saigh, 1984; Szafran, 1981; Trentham, 1975). Conversely, unannounced tests, carelessly administrated tests, poorly monitored tests, and

tests perceived by students to be unfair not only adversely impact upon student performance but tend to heighten test anxiety and encourage cheating.

Test feedback: The prompt return of achievement tests with the provision of knowledge of results or other forms of achievement test feedback, such as discussion of questions missed tends to increase achievement test practices (Kulik & Kulik, 1988; Marso, 1970a; Wexley & Thornton, 1972).

The impact of test question difficulty and of test question arrangement upon achievement test has been less conclusive than the findings from the research of many other aspects of testing. Similarly, teachers are commonly advised when constructing formal teacher-made tests that test difficulty should be approximately 50%, after corrections for probability of guessing relative to question types used, in order to assure an acceptable level of test reliability (Gronlund & Linn, 1990; Mehrens & Lehmann, 1984). However, those students' study efforts would be more efficiently rewarded by a moderately high level of students' success on teacher-made tests. Students having experienced one or more very hard tests in a course are less likely to be encouraged to persevere in their course study efforts if they assume that all subsequent tests in the course will be as difficult or more difficult. The research of test question difficulty arrangements, such as random placement or easy-to-hard placement within tests, indicates that arrangement patterns generally have little impact upon achievement test practices on teacher-constructed tests (Klimko, 1984; Marso, 1970b; Monk & Stallings, 1970; Newman, Kundert, Lane, & Bull, 1988). A common criticism of teacher-made tests, however, is that they tend to function almost exclusively at the recall or knowledge cognitive level (Fleming & Chambers, 1983; , and studies of teachers' testing practices show that

teachers generally do not use test specification tables to better match test questions with content objectives (Gullickson & Ellwein, 1985; Marso & Pigge, 1988a). During the late 1980s, the measurement profession, through the efforts of the National Council of Measurement in Education, the American Association of Colleges for Teacher Education, the American Federation of Teachers, and the National Education Association, developed standards teachers' competence in achievement testing that were published in 1990. The standards represent the measurement profession's views of what teachers supposed to know about testing. The measurement profession's standards for teacher confidence in the achievement testing shows that teachers supposed to be knowledgeable about and proficient in: the selection of appropriate assessment methods for making various instructional decisions, development of assessment guidelines appropriate for making various instructional decisions, appropriate administration and scoring of assessment measures and the appropriate interpretation of the results of classroom assessments, appropriate use of achievement test results in making instructional and related decisions about students and school curricula, appropriate communication of achievement test results to students and related audiences and identification and appropriate response to ethical and legal issues and concerns related to achievement test. It bonded together all the parts of the research work covering broad areas such as teachers' practices, achievement testing strategies Mathematics teachers' views, kinds of achievement strategies, test construction, test administration and scoring that are available and are used, and significant of achievement test results and challenges teachers face when practicing achievement testing. Validity, reliability would be discussed.

The researcher believes that teacher factors such work experience, qualification, programme offered at the degree level and the number of years for in-service. Training courses on assessment, measurement and evaluation programme attended and his or her views have a role to play in the kinds of assessment chosen. The factors also influence the skills of Mathematics teachers in testing the students and the kinds of achievement test strategies.

Teachers Years of Teaching Experience

The study contributes to the literature by focusing on senior high school teachers and by extending the analysis to student outcomes on test practices. The study contributes to policy debates by documenting that teachers can and do continue to learn on the test construction, administration and strategies use to assess students learning outcomes. In the late 1980s, most of the nation's teachers had considerable experience only 17 percent had taught for five or fewer years. By 2008, however, about 28 percent or more than one in four of America's teachers had five or fewer years of experience (Ingersoll & Merrill 2010).

Some observers applaud this "greening" of the teaching force because they believe that experience is not crucial to teacher effectiveness. Others decry the trend because they believe that good teaching is a complex process that can best be mastered through years of classroom experience. Most researchers and policy makers agree that regardless of how effective they may eventually become, novice teachers are typically less effective than their counterparts who have more experience. In recent years, researchers have used administrative data from individual states and districts to examine how teaching experience contributes to student test scores. Research by Kane et al(2008) concludes, for

example, that while teachers have a steep learning curve for the first few years the test score returns to experience level off after four or five years. Using data from New York City, Boyd et al. (2008) report a similar pattern for both basic and senior high school math teachers. Although studies by Clotfelter, Ladd & Vigdor (henceforth CLV), based on North Carolina data (2006, 2007a, and 2010) confirm that the more experience teachers are largest in the first five years, they report some limited gains beyond that for both basic school math and reading teachers, but none for high school teachers. Using Florida data, Harris & Sass (2011) find rising returns to experience through more than 20 years of experience for basic and senior high school teachers, but negative returns for high school teachers.

Teacher's experience contributes to four non-test score outcome measures: absences, reported disruptive classroom offenses, time spent completing homework, and time spent reading for pleasure. Although these behaviors depend in part on a student's home or community environment, they also reflect important facets of learned motivation, perseverance, and self-control, and may be particularly important for the future success of school students.

Evidence has shown that more experienced teachers generate more positive outcomes along these dimensions as well for test scores would greatly strengthen the policy argument for pursuing policies designed to develop and retain experienced teachers. Consistent with prior research, the returns to experience are largest during the first few years, especially for math teachers. Contrary to the received wisdom, however, the returns continue to rise well beyond the first five year for teachers of both subjects. The study has revealed

that, experienced teachers effectively increase performance as it will discourage any bad attitude avoid wrong method and others. In addition, experience appears to increase the ability of teachers to encourage free reading and possibly time spent on homework (as reported by students). Also, weak evidence supports the view that math teachers become more effective in reducing disciplinary offenses, especially in the first several years of teaching. We conclude that as individual teachers gain experience they become more effective not only in raising test scores but also in contributing to other valued behaviors, such as attending school or reading outside of school. In addition they are flexible in their practice, and have a broad repertoire of skills that they can easily access and implement to achieve their goals (Berliner 2001, 2004).

Assessment Standards

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

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

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

Newman and Wehlage (1993) noted that achievement tests tasks need to be organized and structured well so that they are contextualized, integrative, related to the curriculum taught, flexible (requires multiple applications of knowledge and skill), open to self-evaluation and peer-evaluation, contain specified standards and criteria. They again emphasize that correct assessment task must consider the following standards: Organization of information: The task asks students to organize, synthesize, interpret, explain, or evaluate complete information in addressing a concept, problem, or issue. Consideration of alternatives: The task asks students to consider alternative solutions, strategies, perspectives, or points of view in addressing a concept, problem, or issue. Disciplinary content: The task asks students to show understanding and

or use of ideas, theories, perspectives considered central to an academic or professional discipline. Disciplinary process: The task asks students to use methods of inquiry, research, or communication characteristic of an academic or professional.

Achievement Testing

Achievement tests are generally teacher-made tests (McDaniel, 1994). These tests are constructed by teachers to test the volume of learning done by students or their achievement at the end of a course unit, term or at the end of an academic year (Amedahe, 1989). According to Mehrens and Lehmann (1991), teacher-made tests usually measure achievement in a single subject in a specific class or grade. The predominance of teacher-made tests in every educational set up is given credence by the conclusions of studies by Herman & Dorr-Bremme & Stiggins & Bridgeford (cited in Mehrens & Lehmann, 1991) that, in the face of the ever-increasing use of portfolios and performance tests to assess student progress, teacher-made tests are mostly the significant starting point for evaluating student progress in school. The major objective of teacher-made tests has been delimited by measurement experts (Ebel & Frisbie, 1991; Etsey, 2004; Gronlund, 1988; Kubiszyn & Borich, 1984; Mehrens & Lehmann, 1991). All these scholars have accepted with the fact that the major objectives of a teacher-made test is to get valid, reliable, and useful information relating to students' achievements and thus add to the evaluation of educational progress and achievements for the total betterment of classroom teaching and learning. Teacher-made tests can be categorized in a variety of ways. According to Mehrens and Lehmann (1991), one type of classification is based on the type of item format used essay-type versus objective-type. Another classification is

based on the stimulus material used to present the tests to students verbal versus non-verbal, while other classifications may be based on the purposes of the tests and the use of the test results criterion-referenced versus norm-referenced, achievement versus performance, and formative versus summative. The teacher-made test classification that is most popular with testing experts is the classification based on the type of item format used, which classifies tests into objective-type tests and the essay-type tests (Cunningham, 2001; Etsey, 2004; Gronlund, 2012; Nunnally, 1964; Tamakloe et al, 1986). The aforementioned testing specialists have believed that essay-type tests can either be the opened or the restricted response types while objective-type tests can take the form of the short-answer, true-false, matching or multiple-choice. Testing in educational institutions is created to evaluate either curriculum based (classroom instructional) attainment or a variety of student attributes other than curriculum-based achievement. Tests such as career interest, attitudes, and personality tests evaluate a variety of students' attributes other than curriculum-based achievement (Nitko, 2001). Stainback & Stainback (1996) argued that depending on how it is interpreted, evaluating almost any student achievement deriving or related to the classroom curriculum, including achievement testing could be an example of curriculum-based assessment. It must be emphasized that achievement testing is concerned with evaluating students based on the domain of content areas they have studied, which are drawn from the school curriculum. Etsey (2012) stated that achievement test "measures the extent of present knowledge and skills. In achievement testing, test examinees are given the opportunity to exhibit their acquired knowledge and skills in specific learning situations" (p. 41). An extensive review of the literature posits two

main types of achievement tests. These are teacher-made tests and external tests (Nitko, 2001). Assessment made by teachers of students' attainment, knowledge and understanding is called variously as teacher-made tests. Teachers construct these tests to evaluate the amount of learning done by students (Amedahe, 1989). External tests or "extra-classroom assessments" (Nitko, 2001, p. 43), on the other hand, include assessment tools that are developed and/or graded by people who are not associated with the schools providing the students' learning (Lissitz & Schafer, 2002). Commercial test publishers, departments of education, and local school jurisdictions, usually develop external test (Reeves, 2003). According to the National Association of School Psychologists (NASP, 2002), external tests are usually allowed by core components of standard based rectification, which includes: content and performance standards set for all students, development of tools to measure the progress of all students toward the standards, and accountability systems that require continuous improvement of student achievement. External test can take the form of textbook accompaniments, survey tests and mandated tests (Munson & Parton, 2013; Nitko, 2001; Zucker, 2004).

Test

It is an essential tool that helps to quantify such constructs which helps one to make a value judgment about the degree to which such constructs might probably exist in an individual. A large number of assessment techniques are being used to collect information about students. These include formal and informal observation of students, paper-and-pencil tests, students' performance on homework, laboratory work and projects during oral questioning and analysis of students' records (NCME, 2014). Teachers in the educational setting

would want to assess their students over a given period. All the domain of such construct might not be known using a single test. Nevertheless, a well-constructed test could sample to a large extent a reasonable amount of the construct on which value judgment could be made. Educators and teachers must also know that a test is subject to errors that could adversely affect decision making.

Tom and Gary (2003) opined that test misuse and abuse can occur when users of test results are unaware of the factors that influence the usefulness of the test scores. Among the major factors are the technical adequacy of a test and its validity and reliability. Technical inadequacies might emerge from factors such as test appropriateness, content validity evidence, appropriateness of reading level, language proficiency and cultural characteristics of students and teachers may affect administration procedure and scoring of test. It must also be noted that even when a test is technically adequate, misuse and abuse can occur because technical adequacy does not ensure that test scores are accurate or meaningful. When students' achievement levels are not properly measured and interpreted, teachers and school administrators are not able to provide the right educational opportunities and support for student. Testing provides feedback that facilitates educational decisions. These decisions may be the ones that require information about the success of learning programmes or about students who have reached particular levels of skill and knowledge (Izard, 2005). It is a crucial tool for quantifying these constructions and for aiding in determining the likelihood that a person possesses them to a given degree. Information on students is gathered using a variety of assessment methods. These include formal and informal student observation, paper-and-pencil testing, homework

completion, laboratory work, projects, oral questioning, and record analysis (NCME, 2014). Teachers would wish to evaluate their students over a specific time period in an educational context. Nevertheless, a well-designed test could sufficiently sample a considerable portion of the construct from which value judgment could be made. According to Tom and Gary (2003), test misuse and abuse can happen when recipients of test findings are ignorant of the variables affecting the utility of the test results. The test's technical suitability, validity, and reliability are a few of the key elements. Technical errors could result from aspects including test appropriateness, topic validity evidence, appropriateness of reading level, language competency, and cultural traits of teachers that could influence test administration and scoring. Additionally, it should be summarized that even when a test is technically sound, misuse and abuse can still occur because technical soundness does not guarantee that test results are correct or significant. Teachers and school administration are unable to give pupils the right educational opportunities and help when students' accomplishment levels are not accurately monitored and interpreted. Testing offers feedback that helps with educational decision-making. These judgments can call for knowledge regarding the effectiveness of educational initiatives or about students who have attained specific skill and knowledge levels (Izard, 2005). One should not undervalue the value of tests. To start with, tests are used to grade students. The letters or numbers (A, B, C) that classroom teachers use to designate student accomplishments are their official evaluations or judgments on the caliber or value of those accomplishments. Once more, it is employed in selecting processes. Sometimes an institution will approve or reject applicants for particular programs depending on the results of testing. The institution no

longer cares about those who are rejected. Tests are essential components of all educational systems. Teaching and testing are intertwined. According to Quaignrain (1992), tests offer the necessary data for evaluation. There cannot be results knowledge without evaluation. The researcher determined the test construction concepts of Tamakloe & Amedahe (1996) & Etsey (2004) to be extremely complete and useful in the testing environment of a classroom based on the existing literature. This principle consists of eight steps. The phases are: defining the test's goal; choosing the structure of the test items; deciding what will be tested; writing the individual test items; and reviewing the test items. It held that an effective test had to come next. Mehrens and Lehmann contend that in order to reduce interruptions inside and outside the testing room that have the possibility to reduce students' performance, noise and distraction should be kept to a minimum. Tests shouldn't be given right before or immediately after a long vacation, , or any other relevant event where students are engaged in either physically or psychologically demanding activities, according to research by Amedahe & Asamoah-Gyimah (2003) and Etsey (2004). According to Amedahe & Asamoah-Gyimah (2003), examinations shouldn't be administered when students are participating in pleasure activities like co-curricular activities or other recreational activities because this would reduce their attention. According to Amedahe & Gyimah (2003), Etsey (2004), and others, the test responses should be graded item by item rather than script by script. This rule aims to reduce the influence of carryover on the scores in order to maintain uniformity. According to Amedahe & Gyimah (2003) & Etsey (2004), the mechanics of expression, such as proper formulas usage, application of

formulas, quality of drawing, orderly presentation of material, and identification of shapes, should be appraised independently of subject matter correctness.

Teacher-Made Tests

Evaluations are made by teachers on student's achievement, knowledge and understanding is called variously as teacher-made or classroom made test and school-based assessment (Amedahe, 1989). The motive of teacher-made tests is joined with the constructivist model of learning. In this model, it is vital to comprehend what the student knows and how he/she improves it in order to develop his/her knowledge of understanding. In this model, it is learning with understanding which counts and to this end, information about existing ideas and skills is relevant. Work in psychology and learning shows similarly that for efficient learning, the task must be matched to the student's current level of understanding Gipps (1992), and either pitched at the level to provide practice or slightly higher in order to expand and develop the student's skills. For content of a course to be sufficient and ensure that it is significant as well, the content should match the understanding level of a particular student. Salvia and Yesseldyke (2001) stated that, teacher made tests are better when used to evaluate students because they are curriculum matched. If the new task is much too easy, the students can become bored, and if much too difficult, the student can become de-motivated (Gipps, 1999). Importantly, there are two major forms of teacher or classroom-made test; formal and informal tests. Instructors may pose questions, observe activities, and assess students' work in an organized and gradually way (Gipps, McCallum, McAlister & Brown, 1995). Classroom tests are specifically teacher-made tests. Instructors have the duty to provide their students with the best guidelines as possible. This means that they must

have some important content procedures whereby they can reliably and validly evaluate how effectively their students have learnt what has been taught them (Mehrens & Lehmann, 2009). The pencil and paper or teacher-made test is one such tool. Classroom teacher-made tests mostly prevail in subjects –matter like Science and Social Studies. Classroom tests can also, be tailored to fit a teachers' particular instructional objectives, essentially, when one wishes to provide for optimal learning on the part of the pupil and optimal teaching on the part of the teacher (Bejar, 1984). Here, without achievement tests, the objectives that are unmatched to a particular school or teacher might not be evaluated. The emphasis on the desirability and importance of the teachers being able to construct their own personal, unique and relevant tests is based on the principles of assessment in education. A study conducted by Stiggins & Bridgeford (1985) on the uses of various types of tests reported that the tests are: to find out the achievement level and efficiency of methodology adopted, for diagnosis, to encourage students to learn to improve in their work and for certification. Upon all these relevance of teacher-made tests, a study conducted in the United States of America shown some challenges in teacher-made tests, in the way that, teachers were only nurture to teach but not to evaluate their students (Gullickson, 2001). To start with, ambiguous questions is when a statement have two or more meanings, one has ambiguity. Again frequent wording adds to difficulty in teacher-made test. Too often teachers think that the more wording there is in a question, the clearer it would be to the student. This does not always happen. The more concise and clear-cut the wording, the greater the probability that the student will not be confused. Mostly, teacher-made tests do not cover the objectives stressed and taught by the teacher and do not reflect

proportionally the teacher's judgments as to the importance of those objectives. Teacher-made achievement tests are mostly over loaded with items that only test the students' capacity to remember specific facts and information (Fleming & Chambers, 1983). Use of inappropriate item formats also adds to challenges in teacher-made tests. Some teacher uses different item formats like true-false or essay solely because they feel that diversity is desirable. But the need for change should not govern the type of item to be used therefore; teachers should be selective and choose the format that is most effective for measuring a particular objective. According to Nitko (2001), assessment content is significant when teacher-made or classroom test consists choice formats such as (multiple choice, true or false, matching exercise and other formats like greater - less same items), short answers and completion format and essay format (restricted responses and extended responses). Some educators argue that essay tests are more susceptible in scoring than the objective tests. However, classroom teachers exclusively use both since one cannot be used exclusively to measure all learning outcomes. According to Bartels (2003), with regard to the objective type tests, the multiple choice, short-answer/fill-in-the-blanks, matching and true or false types are the major ones used by tutors in the teacher colleges of education in Ghana.

Objective -Type Tests

The objective-type item was developed in response to the criticism made against the essay type tests. Some of the criticisms were, poor content sampling, unreliable scoring, time-consuming to grade, and encouragement of bluffing. The objective test-items normally consist of a large number of items and the responses are scored objectively, to the extent that competent observers can

agree on how responses should be scored (Amedahe & Etsey, 2003). Objective-type item formats are put into two groups; the supply type and the selection type. The supply type format consists of completion type, fill-in-the blanks and short answer. The selection type consists of true-false, matching, and multiple-choice item type. According to Amedahe & Etsey (2003) objective type test items are most useful when class sizes are very large and when there is limited time to submit the results of the test. The short-answer and completion format consist of one or more blanks in which the student writes his answers to the question with a word or, phrase. This type of objective test is also known as constructed – response type. It consists of a statement or question and the respondent is required to complete it with a short answer usually not more than one line (Etsey, 2012). It is used for testing knowledge of facts or recall of specific facts (example, “knowledge objective” in Bloom’s taxonomy of educational objectives). Short-answer and completion format can be used to assess higher-level abilities like, to make simple interpretations of data and applications of rules, to solve numerical problems in science and mathematics, and to manipulate mathematical symbols and balance mathematical and chemical equations. A true or false test consists of a statement to be marked true or false. Here their utilities are placed primarily in assessing knowledge of factual information. True or false items are difficult to prepare (Salvia & Ysseldyke, 2001). True or false test items are made up of four types; simple true or false, (here only two choices; true or false), complex true or false (comprises three choices; true or false and opinion), compound true or false (consists of two choices, true or false plus a conditional completion response) and finally

multiple true or false (consist of a stem with three, four or five options and the respondent indicates if the options are true or false (Etsey, 2012).

One of the limitations in constructing the true or false test items is that, the probability of getting right answer by guessing is high. It can be used to assess only a few numbers of educational objectives, and can be used to evaluate definitions, facts, meaning of the true or false, recognition, and interpretation of charts/graphs. An advantage of true or false test item is that, they can cover a wide range of content within a relatively short period of time. Matching test format is another choice format item which presents respondents with three things; directions for matching, a list of premises and a list of responses. The simple matching exercise requires simple matching based on association that a student must remember. This is basically done to assess respondents' comprehension of concepts and principles. One of advantages of matching test format is that, matching test format use pictorial materials to assess student's abilities to match words and phrases with pictures of objects or with locations on maps and diagrams. A multiple-choice item consists of a stem followed by a list of two or more proposed alternatives; here the respondents are expected to select the correct option from the alternatives. Normally, only one of the options is the correct or best answer to the question one poses. This is called the keyed alternative, keyed answer or basically the key whiles the remaining incorrect options are called foils or distracters. The purpose is to allow students to demonstrate their knowledge and understanding of the learning targets. There are three types of multiple-choice tests. These are the single correct type and the "multiple responses" type. The "single correct" type consists of a stem followed by three or more responses and the respondent is to select only one option to

complete the stem. The “multiple responses” type consists of a stem followed by several true or false statements or words. The respondent is to select which statement could complete the stem. Multiple-choice tests format does not require students to write out and elaborate their answers and minimize the opportunity for less knowledgeable students to “bluff” or “dress-up” their answers (Wood, 2007). According to Amedahe & Etsey (2003), essay test items consist of relatively few items, but each require an extended response. Essay test items provide respondents with the freedom to organize their own ideas and respond with limited restriction. Here respondents are asked to speak to a particular issue and for that reason they could not just write a single word as an answer than to express themselves in terms of what they know about the items. The ability of the respondents to express themselves clearly and fluently and with content required tells the instructor that they have actually mastered the content of the subject. Essay questions are most useful in assessing instructional objectives prepared at a comprehension level or higher order thinking (Salvia & Yesseldyke, 2001). Nitco (2001) noted that “what is perhaps unique about the essay format is that it offers students opportunity to display their abilities to write about, to organize, to express and to explain interrelationship among ideas” (p.187).

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

knowledge, evaluate factual knowledge, and organize ideas and present ideas in a logical, coherent written fashion.

The extended response makes the greatest contributions at the levels of synthesis and evaluation of writing skills (style, quality). Under the restricted-response essay questions, the student is more limited in the form and scope of the answer because it tells specifically the context that the answer is to take. This type of question is of greatest value for measuring learning outcomes at the comprehension, application, and analysis level, and its use is best reserved for these purposes.

Purposes of Achievement Test Results in Schools

There is a purpose for conducting classroom achievement testing (Ghaicha, 2016). Before developing an achievement test, there is the need to think of what the results will be used for. The main purpose of practicing achievement testing is to obtain valid and reliable results or data that are indicative of how much students have gained from classroom instruction and to improve students' learning (Amadehe & Asamoah Gyimah, 2014).

According to Barton and Coley (1994), there are several reasons for carrying out achievement testing in schools. These include; a means for accountability to tax payers, improve instruction, identify strengths and weaknesses of students, promotion of students into various grades and graduation, for accreditation purposes, rewarding students and schools and to know whether preschoolers are ready for formal education. Again, Etsey (1997) summarized from various researchers, ten (10) uses or purposes for engaging in classroom achievement testing. These include; identifying students' achievement differences in schools, putting students into appropriate programmes,

monitoring student progression in terms of achievement and identifying the difference in potential and actual achievement of students. Also, it is to put students into appropriate groups for further instruction, plan teaching and learning activities in schools, to modify the curriculum, and provide relevant information concerning the success and appropriateness of various instructional programmes.

Empirical Review

Academic Perspectives on teachers practice on achievement testing are that behaviors are based on the underlying theory or ideas that teachers subscribe to, behaviors and other factors have a significant impact on assessments of students' performance levels (Munoz, Palacio, & Escobar, 2011). According to Delanshere and Jones (1999), the definition and purpose of a test as set forth by an external examining body, teachers' beliefs about the curriculum and the subject matter, and teachers' overall views of how students learn, all have an impact on teachers' beliefs about student testing in the classroom depending on how long a teacher has been teaching, their attitudes about achievement testing may change over time. The assumption that teachers should test students' academic progress in the classroom has increased, according to Unal and Unal (2019), is based on experience. They claim that in contrast to their counterparts who are more seasoned teachers, rookie teachers think that student testing is unimportant and unchangeable.

According to Thomas (2012), alternative testing to the conventional paper and pencil assessments has a significant impact on students' achievement. This belief is shared by both trained and untrained teachers. The prevalent views of teachers regarding achievement test success. According to Jones & Egley

(2007), testing of students can be summarized as follows: testing of students impedes learning; tests in general have little or no impact on teaching and learning; students will learn better without tests; students prefer to focus on honing their test-taking skills; students' learning is not accurately measured with testing; testing forces teachers to teach in a way that is consistent with their beliefs; students' creativity is hampered by testing and test; More specifically, teachers think that student testing interferes with instruction and learning and makes it harder to acquire accurate data about pupils because of the possibility of inaccuracy. In addition to the discussion above, Thomas (2012) evaluated teachers' views of achievement testing and provided the following findings: formal testing has little impact on students' learning, informal testing can be used to assess students' learning, achievement testing has the power to pressure teachers to complete the curriculum, direct observation of students' learning can lower performance, and testing is necessary in aiding in student learning. It might be said that since test results don't always accurately reflect what kids are capable of, teachers' thoughts and perspectives on student testing have minimal bearing on teaching and learning. The fundamental cause is that teacher-made tests are susceptible to some measurement mistakes.

Teachers' views

Views can be an individual's subjective knowledge and emotions concerning Mathematics (cf. Lester, Garofalo & Kroll, 1989); these relationships are based on their personal experiences (Pehkonen & Kaasila, 2011). How students are taught in Mathematics will therefore have an impact on their views (Pehkonen, 2001). Pehkonen (1995) divides students' views on Mathematics into four subgroups: views on Mathematics as subject,

views on one's ability to do Mathematics, views on Mathematics teaching and views on how to learn Mathematics. The views of teachers as well as students become an important factor which has a substantial impact on the quality of the teaching and learning process (Spangler, 1992; Pehkonen, 1995).

Ernest (1991) describes five educational strategies in Mathematics while claiming that teachers' views on the Mathematics affected their teaching and thereby student learning problems. The first strategy is the 'industrial trainer' where teachers see mathematics as a set of truths and rules. Teaching involves authoritarian processes, such as transmission and drill. Students' learning is regarded as hard work, effort, practice and rote learning. The second strategy is called the 'technological pragmatist' where Mathematics is seen as an unquestioned body of useful knowledge. The teacher works as a skill instructor and motivates the learning of Mathematics through its relevance to work. Learning is characterized by accumulating skills and practical experience. Teachers explain, motivate and try to pass on the subject's structure so that the student can understand and apply Mathematical concepts. The fourth strategy is the 'progressive educator' which is characterized by a teaching process in which teachers facilitate personal exploration and try to prevent failures. The learning process is marked by activity, play and exploration. The view of Mathematics is a process view – personalized Mathematics. The final strategy is the 'public educator'. Teachers using this strategy see Mathematics from the perspective of social constructivism where teaching involves discussions, cognitive conflicts, questioning of content and pedagogy. Students learn by questioning, decision-making and negotiation.

Researchers have attempted to investigate teachers' perceptions of assessment in many different ways (Chester & Quilter, 1998). Chester and Quilter believed that studying teachers' views of authentic assessment is important in the sense that it provides an indication of how different forms of authentic assessment are being used or misused and what could be done to improve the situation. More critical also is the fact that perceptions affect behavior (Atweh, Bleicker & Cooper, 1998; Calderhead, 1996; Cillessen & Lafontana, 2002). Creswell (2012) engaged 25 teacher-volunteers to participate in a study representing six secondary rural schools from New South Wales, Australia. The researchers used the Structure of Observed Learning Outcome; a cognitive structural model which provided "a basis for both assessing students' understanding and identifying ways of enhancing students' learning" (Creswell, 2012, p. 420). Workshops were conducted at the college for these teachers, focusing around the model assessment tasks and teaching strategies of the 25 teacher-volunteers by Creswell (2012). The researchers primarily used two sources of data: "students' scripts coded using the structure of observed learning outcomes model" and interviews with teachers. They inquired from the teachers their experiences with the new approach to teaching i.e. (SOLO) and assessment practices to enhance students' learning. The researchers found out that all teachers who participated in this project represented a change in their perception enabling them use collaborative effort to engage students' understanding in their classrooms. According to Creswell (2007), the project helped teachers "recognize that restricting the type and style of questions in their teaching and assessment provide limited scope for students to demonstrate their conceptual understanding" (p. 431). Overall, the researchers found out that teachers

reported a shift in their perceptions of learning demonstrated in their teaching and assessment practices which was noticed by students and other teachers as well (Creswell, 2012). Chester and Quilter (1998) in their study on in-service teachers' views of teachers test practices; standardized testing, and alternative evaluation methods in Debre Markos University in Ethiopia concluded that teachers' views of achievement testing affected their classroom assessment practices. They found out that teachers that attached less value to classroom assessment used standardized tests most of the times in their classrooms. Chester and Quilter went further to say that teachers with negative experiences in alternative assessment and standardized testing are least likely to see the value in various forms of assessment for their classroom. They recommended, therefore, that in-service training should focus on helping teachers see the value of other assessment methods rather than "how to" do assessment. An interview with a fifth-grade teacher at Deerfield Elementary school in Lexington, USA by Kentucky Department of Education (1991) confirms that teachers are aware of the limitations of standardized tests. They further revealed that the teacher indicated that curriculum must emphasize subjects for which the state accountability test measures skills: math, reading, and science. The teacher argued further that test scores do not truly reflect her students' abilities and are too vague to help her pinpoint individual needs (Kentucky Department of Education, 1991). The teacher asserted that she longs for an assessment that relies on more than just written problems that could capture the more diverse skills visible in her classroom and valued in the workplace, such as artistic talent, computer survey, and the know-how to diagnose and fix problems with mechanical devices (Kentucky Department of Education, 1991).

Teachers Test Strategies

The best task or strategy to use in assessing students learning outcomes or achievement levels has not reached any consensus, (Denman & AL-Mahrooqi, 2018). However, the assessment task or strategy most relevant to measuring the performance or characteristics of the individual or student is chosen (Amedahe & Asamoah Gyimah, 2014). Amedahe & Asamoah Gyimah (2014) uses multiple strategies to assessing students' achievement levels.

Several strategies and procedures are available to the Mathematics teacher and can be used to assess achievement levels of students. According to Barton & Coley (1994), strategies used in achievement testing in America include; projects, portfolios, presentations, assessing work samples, multiple choice items, and problem sets. Others include interviews, observations, exhibitions, demonstrations, individual performance, written responses, and group performances. Students can also be given assignments in order to test their achievement level (Amedahe & AsamoahGyimah, 2014). Fox & Soller (2001), in their study on authentic assessment strategies and tools employed by teachers in Malawi found out that students prefer working collaboratively using projects, group work, class exercises, co-curricular activities such as playing ball, running and others. It was also revealed in the study that education systems that emphasize tests and examinations put some student at a disadvantage (Mbanjo, 2003; Nampota & Wella, 1999). Fook & Sidhu (2010) conducted a study in Malaysia to investigate the different types of authentic assessment used in higher education in Malaysia. The study employed a qualitative research method and involved the use of instruments such as interview, document analysis and classroom observations to collect relevant data in the classroom.

The researchers identified that different types of authentic assessment were used. The study revealed that teachers employed the following assessment tools; portfolio (15%), article review (14%) performance product (25%), project (20%) and test (50%). The findings indicated that alternative and authentic assessment have more acceptances from students and should, therefore, be viewed as an alternative to traditional standardized assessment. The study again revealed that assessment practices in some subject areas like Mathematics and Science indicated favourable emphasis being given to formative assessment because 80% of the total marks have been allocated to on-going assessment and 30% was for the test. Moreover, students interviewed also agreed that project and portfolio assignment given were to a large extent real and authentic tasks that they could relate to their future workplace. Beckmann, Senk & Thompson (1997) studied the assessment and grading practices of 19 high school mathematics teachers in the United States. Their study revealed that the most frequently used assessment tools were tests and quizzes and these determined about 65% of students' grades. Twelve of the nineteen teachers used other forms of assessment such as written projects, experiments, demonstrations or interviews with students. The study also revealed that teachers recorded a very high level of student participation in the written projects, experiments.

Challenges in practicing Achievement Test

A study was conducted by Kankam, Bordoh, Eshun, Bassaw & Korang (2014) to investigate the influence of teacher-made test on their classroom practices and the challenges teachers of Mathematics encounter in Ghana. A descriptive case study design was employed using ten (10) Senior High Schools and twenty (20) teachers were randomly selected from fifty-seven (57) public

schools in the Upper West Region of Ghana. Semi-structured interview guide was used as the main instrument to collect the data. From the study, it was concluded that multiple forms of teacher-made test were not used by classroom teachers due to limited time, examination policies, resources and methods of achievement testing employed by their schools. It was also revealed that schools were not having individual assessment policies except for the West African Examination Council's assessment policy. Again, students were not involved in the assessment process because of the achievement test methods they chose. Furthermore, some teachers confirmed that the issue of completing the syllabus restricted proper practicing of achievement testing. Again, large class size, inadequate time, inadequate technological and physical facilities, teachers inflation of students' marks, teachers inadequate knowledge on different forms of achievement testing and parents who do homework for their wards are some challenges teachers face when practicing achievement testing, Metin (2013). According to Hussain, Ahmad & Shaheen (2019), majority of teachers are without training in the field of classroom achievement testing. Teachers and policy makers heavily stick to traditional pencil and paper test practices coupled with large class size and the inability of teachers to statistically analyze students' achievements are major challenges affecting classroom achievement test practices in Pakistan. Again, Chand (2017) conducted a study on challenges teachers face in carrying out classroom achievement testing and concluded that overloaded content, inadequate in-service training and inadequate time to complete various syllabus are challenges teachers face in practicing proper achievement testing. It can therefore be concluded that achievement testing is bedeviled with several challenges not only peculiar to Ghana but the world at

large. Challenges faced when practice of achievement testing frequently cited in literature include inadequate resources and facilities, large class size, inadequate professional training and knowledge on the part of teachers, examination policies favouring traditional testing methods, academic dishonesty and overloaded content. Eshun et al. (2014) conducted a study to investigate the influence of achievement test on achievement test practices of teachers and the challenges they encounter in practicing of achievement testing in Ghana. The study used a descriptive case study design and it involved 10 senior high schools and twenty teachers randomly sampled from thirty (30) senior high schools in the Upper West Region of Ghana. Semi-structured interview guide was the main instrument used for data collection. The research found out that the forms of achievement test some teachers used in their classrooms were limited due to examination policies, time, resources and assessment methods employed by their schools. Furthermore, they revealed that most teachers they observed were not using evaluation techniques that involved students in the teaching and learning process. Again, they indicated that some teachers revealed that using the achievement test would delay them in completing topics in their syllabuses given to them. Beckmann, Senk and Thompson (1997) in their study conducted in USA identified three reasons why teachers do not use multiple assessment methods. First, some teachers had limited knowledge of different forms of assessment. Second, teachers felt they had no time to create/develop authentic assessment.

Construction of Achievement Tests

The basic principles for the construction of teacher-made tests have been developed over the years by a number of educational measurement experts

(Amedahe, 1989). While some of the test construction principles are general and apply to any type of test, others are specific and apply solely to the particular type of test under construction. From available literature, the test construction principles that the researcher judged as most simple and practicable in the classroom testing situation were those postulated by Tamakloe, Atta & Amedahe (1996) & Etsey (2004). These are in eight steps. The steps are: define the purpose of the test, determine the item format to use, determine what is to be tested, write the individual items, review the items, prepare the scoring key, write directions, and 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 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 & 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 behaviors and linked to what has been stressed in class. A test plan made up of a table of specifications should be made. The table of specifications matches the course content with the instructional objectives (Etsey, 2004). With the total number of items on the test in mind, the specification table helps to avoid overlapping in the construction of the test items, helps to determine the weighting of learning outcomes with respect to content areas, and makes sure that justice is done to all aspects of the course, thereby helping to ensure the content validity of the test. After the planning stage, actual writing of the individual test items follows. Tamakloe et al. (1996) & Etsey (2004) have pointed out that whichever test item types that are being constructed must follow the basic principles laid down for them. There are, however, general guidelines that according to Mehrens Lehmann (1991) & Etsey (2004), apply to all types of tests. These include:

1. The table of specifications must be kept before the teacher and continually referred to as the items are written.
2. The test items must be related to and match the instructional objectives.
3. Well-defined items that are not vague and ambiguous must be formulated. Grammar and spelling errors must be checked. Textbook or stereotyped language must be avoided.
4. Excessive verbiage and complex sentences must be avoided.

5. The test items must be based on information that students should know.
6. More items than are actually needed in the test must be prepared in the initial draft. Mehrens & Lehmann (1991) suggested that the number of items should be 25% more while Hanna (as cited in Amedahe, 1989) has suggested 10% more items than are actually needed in the test.
7. Items of varying levels of difficulty must be used. This, however, depends on the purpose of the test.
8. The items and the scoring keys must be written as early as possible after the material has been taught.
9. The test items must be written in advance (at least two weeks) of the testing date to permit reviews and editing.

After the items have been written, Tamakloe et al. (1996) call the next stage the item preparation stage. At this stage the test items must be reviewed and edited. Etsey (2004) has suggested that the items must be critically examined at least a week after writing them. He has emphasized that where possible, fellow teachers or colleagues in the same subject area should review the test items. Reviewing and editing the items are for the purpose of removing or rewording poorly constructed items, checking difficulty level of items, checking the length of the test, and the discrimination level of the items (items must discriminate between low- and high-achievers). All test items should be checked for technical errors and irrelevant clues. After reviews and editing, the test items can now be assembled. In assembling test items, the following points must be considered (Etsey, 2004; Kubiszyn & Borich, 1984; Mehrens & Lehmann, 1991; Tamakloe et al., 1996:

1. The items should be arranged in sections by item formats. The sections must progress from easier formats (true-false) to more difficult formats (interpretive exercises and essay).

2. Within each section or format, the items must be arranged in order of increasing difficulty. One way of achieving this is to group items in each format according to the instructional objectives being measured and make sure that they progress from simple to complex.

According to Mehrens & Lehmann (1991), such a grouping has the advantage of helping the teacher to ascertain which learning activities appear to be most readily understood by students, those that are least understood and those that are in-between. According to Hambleton & Traub (cited in Mehrens & Lehmann, 1991), ordering items in ascending order of difficulty leads to better performance than either a random or hard-to-easy ordering. Lafitte (cited in Mehrens & Lehmann, 1991) on the other hand, has reported inconclusive data. Although, empirical evidence is also inconclusive about the effectiveness of using statistical item difficulty as a means of ordering items, Sax & Cromack (cited in Mehrens & Lehmann, 1991), Mehrens & Lehmann (1991) and other testing experts have recommended that for lengthy or timed tests, items should progress from the easy to the difficult-if for no other reason than to instill confidence in the examinee, especially at the beginning. It should be noted however, that, the use of statistical item difficulty or item difficulty indexes by the classroom teacher seems impracticable to a large extent (Kubiszyn & Borich, 1984; Tamakloe et al., 1996). This is because statistical item difficulty data are always gathered after test administration test try-outs and teacher-made test items are usually not pre-tested. Mehrens & Lehmann (1991) however,

recommended that subjective judgment must be relied on to determine difficulty level of items. They have stated that -teachers could only categorize their items as difficult, average or easy.

3. The items must be spaced and numbered consecutively so that they are not crowded and can easily be read.

4. All stems and options must be together on the same page and if possible, diagrams and questions must be kept together.

5. If a diagram is used for a multiple-choice test, the diagram must be placed above the stem.

6. A definite response pattern to the correct answer must be avoided.

In addition to the above, Gronlund (2012) & Etsey (2004) have recommended that for objective-type tests, the options must be written vertically below the stem rather than across the page. Further, Etsey (2004) has suggested that test items can also be arranged according to the order in which they were taught in class or the order in which the content appeared in the textbook. After the test items have been assembled, the next task is the preparation of the scoring key, the marking scheme or the scoring rubric (Etsey, 2004). The marking scheme according to Etsey (2004) & Amedahe & Asamoah Gyimah (2003), must be prepared when the items are still fresh in the teacher's mind and always before the administration of the test. This way, defective items that do not match their expected responses would be recognized and reviewed. For objective-type tests, correct responses to items should be listed. For essay-type tests, points or marks should be assigned to

Various expected qualities of responses. Mehrens & Lehmann (1991) have pointed out that if the teacher considers it prudent to have differential weighting

for different essay questions, then factors such as the time needed to respond, the complexity of the question, and emphasis placed on that content area during the instructional phase must be considered. Immediately following the preparation of the marking scheme is the writing of clear and concise directions for the entire test and sections of the test. Here, the time limit for the test must be clearly stated. As argued by Nunnally (1964), and Ebel and Frisbie (1991), a good working rule is to try to set a time limit such that about 90 percent of the students will feel that they have enough time to complete the test. Directions according to Etsey (2004), must include penalties for undesirable writings, number of items to respond to, where and how the answer should be written, credits for orderly presentation of material (where necessary), and mode of identification of examinees. The last stage of the test construction process is the evaluation of the test on the criteria of clarity, validity, practicality, efficiency and fairness. Clarity refers to how simply and clearly the items are written vis-à-vis the ability level of the examinees and the material the test is measuring. It also refers to the kinds of knowledge the test is measuring and how adequately the test items relate to the content and course objectives (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996). Validity bothers on how closely the test represents the material presented in the course unit or chapter and how faithfully the test reflects the difficulty level of the material taught in class. The issue of validity establishes the content validity evidence of the test (Amedahe & Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996). On practicality, consideration is given to whether students will have enough time to complete the test. It also bothers on whether there are enough materials (chairs, tables, answer booklets) to present the test and complete it effectively (Amedahe &

Asamoah Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996). Efficiency bothers on finding out whether the test is the best way to measure the desired knowledge, skill or attitude. Consideration must also be given to the problems that might arise due to material difficulty or shortage and these expected problems well catered for (Amedahe & Asamoah Gyimah, 2003; Etsey, 2004; Tamakloe et al., 1996).

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

Achievement Test Administration

Next important level in practice of achievement test after constructing achievement test is administering the test. Administering achievement test must be based on the basic guiding principle that all students are habituated same opportunity to exhibit their skills on the intended mastering purpose (Rukundo & Magambo, 2010). To support debris argument, Roshan (2016) argues that allowing all students equal opportunity to express themselves during examinations is one important tenet of administering classroom test. According to Rukundo & Magambo (2011), both physical and psychological factors guiding the conduct of the examination must be conducive for the examinee to showcase what they have achieved in the instructional process. With this,

maintaining uniformity pertaining to test administration conditions should be the guiding principle to achieving reliable, valid and consistent results devoid of errors of chance (Oduro-Okyireh, 2008). This stage is so important because any test administered under unhealthy conditions and without regards for best principles of administering test will not produce the desired results and will affect the evaluation of the results. The argument is backed by Roshan (2016) when he said that test which is not properly administered yields results that will not be valid and reliable. She then added that qualified persons should be called upon to administer achievement test to improve validity and reliability of test results. Therefore, best principles of test administration according to Roshan (2016), & Rukundo & Magambo (2016) should be carried out in three phases; before, during and after the test. Firstly, some important activities are supposed to be carried out by test administrators before the actual day of administering test to students. These activities are nonnegotiable to test administrators the test is going to fulfill its purpose. These activities include; securing the test instrument from unnecessary access, giving accurate date and time of the test to both parents and students, organizing materials for the test and if possible informing students of the materials they will bring, catering for special needs students, planning for adequate invigilation, providing test administrators with knowledge on test administration regulations and procedures, providing proper lighting and ventilation, properly arranging seats to avoid cheating during the test and preventing intruders from accessing places of examination. Also, during the test, certain steps factor in place to ensure proper conduct. Roshan (2016) and Rukundo & Magambo (2011) these measures are; avoid impersonation, announce time at regular intervals as the test progresses,

carefully monitor testing proceedings, discourage cheating of all forms, avoid unnecessary interruptions such as noise and assisting students who you think are failing, avoid giving students instructions and new areas to read and avoid eating in the exam room. Lastly, whatever happens after the test can still mar the test administration process. Test administrators are tasked to collect and count test materials whether used or not and check candidates scripts to be sure all scripts are collected. Candidates' scripts should be kept safely away to prevent protected people inserting foreign materials before scoring test.

Test Scoring

lettering of objective test has limited issues as compared to scoring essay type test as the latter is prone to scorer's subjective (Khairuddin, 2018). Using essay type test to measure students' ability is dependent on the rater or scorer as well as the method the scorer decides to choose in the process (Amedahe & Asamoah Gyimah, 2014). According to Amedahe & Asamoah Gyimah (2014), how well an essay type test is graded determines its effectiveness. Two most used methods in grading or scoring essay type test are the analytic method and the holistic method. In using analytic method to score essay type test, the skill or attribute to measure is broken into its component parts and graded. The various grade points are collated to determine the overall mark. According to Khairuddin (2018), this method is most suitable for classroom purposes. On the other hand, the holistic method of grading essay type test looks at the overall quality of a student's response and a single score is assigned (Amedahe & Asamoah Gyimah, 2014). They believe that the holistic method is more suitable for extended response type of essay test and that of the analytic method is good for restricted response type of essay test.



CHAPTER THREE

RESEARCH METHODS

Introduction

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

Research Design

The research design chosen for the study was the descriptive sample survey. According to Amedahe (2004), “descriptive research is which specifies the nature of a given phenomenon” (p. 50) alpha. Gay (cited in Amedahe, 2004), explains that descriptive research involves the collection of data in order to test hypotheses or answer research questions concerning the current status of the subjects of the study. Dawson (2002) posits that a research design is the conceptual structure within which research would be conducted. In this regard, this study would adopt the descriptive research design. Descriptive research was used because; the data collected was used to investigate achievement test practices of teachers in secondary education improvement project schools in

upper west region. Surveys was assisted in gaining a better understanding of achievement test practices of mathematics teachers in secondary education improvement project schools in upper west region. Again, the researcher chose this approach because he was interested in learning about the practice of achievement test from SHS teachers' views in the upper west region. According to Murphy (2009), the major advantage that goes with this type of design is that, the data collection techniques present several advantages as they provide a multifaceted approach for data collection. For example, a survey can provide statistics about an event while also illustrating how people experience that event. Again, he states that the descriptive research design also offers a unique means of data collection thus it provides more accurate picture of events and seeks to explain people's perceptions and behavior on the basis of data gathered at a point in time (Murphy, 2009). However, the design has some weaknesses. Confidentiality is the primary weakness of descriptive research (Murphy, 2009). According to Murphy (2009), respondents are often not truthful as they feel the need to tell the researcher what they think the researcher wants to hear and also participants may refuse to provide responses they view to be too personal. Another weakness of this design, according to Murphy (2009) is that it presents the possibility for error and subjectivity. However, the design was used despite its weaknesses because it seeks to explain people's perceptions and behavior on the basis of data gathered at a point in time and can provide statistics about an event while also illustrating how people experience that event thus providing a multifaceted approach for data collection.

Study Area

The study area chosen for the research work is Upper West region of Ghana. It shares border to the north with Northern Region, south with Cote D'Ivoire, west with Burkina Faso and east with Upper East. The 2021 National Population and Housing Census (NPHC) results revealed a total population of 901,502 in the Upper West region. This is about 4.22% of Ghana's total population of 33,512,935. The population comprises 440,317 males and 461,18 females representing 48.8% and 51.2% respectively with a sex ratio of 94 males to 100 females (Source: 2021 NPHC). Using the growth rate of 1.7%, the projected population for 2020 is 80,502, comprises 39,043 males and 41,459 females representing 48.2% and 51.2% respectively with a sex ratio of 94 males to 100 females [Source: Upper West Region, Ghana]. Out of the thirty (30) Senior High Schools in the region, there are seven SEIP schools which is approximately a quarter of the Senior High Schools. SEIP schools are considered non-performing especially in Mathematics and inadequate human and material resources. Therefore, poor performance in the SEIP schools can statistically affect the overall performance in Mathematics in the region. This is why the researcher chose the Upper West region.

Population

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

of the sample. The population for this research consisted of all Mathematics teachers in the thirty (30) Senior High Schools in the Upper West region. The accessible population was all the Mathematics teachers of the seven SEIP schools. Only mathematics teachers in the SEIP schools were involved because the researcher believes the mathematics teachers were familiar with the practices of achievement testing in SEIP schools.

Sampling Procedure

Sarantakos (2005) postulated that a sample consists of a carefully selected unit that comprises all the categories of the population. Sarantakos (2005) indicates that estimation of the sample size varies significantly, with some researchers showing interest in pure quantity, others in quality and yet others combining in what is called triangulation of sources, data and methods. However, a wise rule is that the sample size must be as large as necessary, and as small as possible. An estimated sample size of 234 mathematics teachers selected for the study using Krejcie & Morgan (1970) sampling table. According to the table, a population of 600 gives a sample size of 234. Fraenkel & Wallen (2009) have also indicated that for descriptive studies, larger sample size produces desirable results to generalize over the population. Therefore, a sample size of 234 for this study was considered large enough to produce the desired results and allowed for generalization of the findings over the population. The study employed the multistage sampling techniques (purposive, stratified and simple random sampling procedure). Per the nature of the study population, purposive, stratified and simple random sampling procedure were used to select cases in the secondary education improvement project schools in the upper west region. Purposive sampling was used because the researcher selectively chose to study

only Senior High School teachers teaching mathematics. According to Crossman (2017), purposive sampling is a non-probability sample that is selected based on characteristics of a population and the objective of the study. Purposive sampling is also known as judgmental, selective, or subjective sampling. This type of sampling can be very useful in situations when you need to reach a targeted sample quickly, and where sampling for proportionality is not the main concern (Crossman, 2017). Stratified random was used because the population comprised of different districts in the region. Stratified sampling was used to select equivalent number of Senior High Schools from the seven districts. According to Van Dalen (2012), stratified sampling is a procedure for selecting a sample that includes identified subgroups from the population in the proportion that they exist in the population. Van Dalen posited that stratified sampling can be used to select equal numbers from each of the identified subgroups if comparisons between subgroups are important. According to Cohen, Manion & Morrison (2011), the quality of any research not only stands or falls by the appropriateness of methodology and instrumentation but also by the suitability of the sampling strategy that is adopted. Therefore, the researcher used stratified sampling to guarantee the desired distribution among the selected subgroups of the population and to aid equivalent selections of senior High Schools from all the various districts where the schools are located and municipals. At the last stage, the researcher used simple random to select the Senior High Schools teachers in the seven districts and municipalities in the Upper West Region. The simple random technique was used in order to give Senior High Schools equal chance of being selected and it helped to avoid biases in selecting the Senior High Schools. This is to

help improve the representativeness of the sample by reducing sampling error (Saunders et al., 2007).

Table 1: The enrolment of the sampled Mathematics teachers

District	Name of school	Mathematics teachers		
		SHS 1,	SHS 2,	SHS 3
Lambussie	Piina S H S	7	9	12
Lambussie	Holy Family S H S	5	11	8
Nandowli	Kaleo S H S	6	13	9
Bussey	Daffiamah S H S	11	5	10
Jirapa	Jirapa S H S	14	8	13
Jirapa	Ullo S H S	9	11	7
Wa- East	Fusion S H S	5	9	13
Total		57	66	72

Data Collection Instrument

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

were prepared based on the objectives of the study to elicit the needed information. Saunders (2007) reiterates that a questionnaire is an ideal tool when collecting a lot of information over a short period of time. Again, the researcher deemed it ideal to use questionnaire because his respondents were literate. The questionnaire was closed ended type. The questionnaire was developed using four- point Likert-type scale ranging from “Strongly Disagree to Strongly Agree”. The research instrument was organized into six sections (A, B, C, D and E). Section ‘A’ comprises the background information of the mathematics teachers’ views. The Section, ‘B’, constitutes the strategies mathematics teachers use to assess students outcomes by teachers in senior High Schools in the Upper West Region. “Section C” was based on challenges teachers face in quest of practicing achievement testing. “Section D” was based on how mathematics teachers follow the principles of test construction in secondary improvement project schools in the Upper West Region of Ghana and finally, Section E” was based on how senior high school teachers follow the guidelines to test administration. The questionnaire was a four-point Likert -type scale which requires participants to indicate their level of agreement or disagreement to the items using strongly agree, agree, disagree or strongly disagree. The responses were scored as follows: Strongly Agree = 4; Agree = 3; Disagree = 2; Strongly Disagree =1.

Instrument validity

According to Copper & Schinder (2003), validity is defined as the soundness of the interpretations and use of student’s assessment result. Validity emphasizes the interpretation and use of results and not the test instrument. Validity is unitary concept that is based on various kinds’ evidence. Validity of

a research is the appropriateness of the interpretation of the results of an assessment procedure for a group of individuals (Mugenda & Mugenda, 1999). Validity was carried out in relation to the study in order to validate and modify the questionnaire's content. Also, in order to improve content validity, the research instrument was evaluated by the thesis supervisors in the Department of Education and Psychology. The updated draft was then given to a three-member panel team consisting of highly inclined academic researchers with relevant knowledge and expertise in the assessment. The researcher ensured that all the necessary corrections were made based on the comments from the panel. The questionnaire was finally given to the researcher's supervisor for further assessment. Careful attention was given to key areas notably research objectives, communication method, potential respondents, cost and time constraint. Also, attention was given to the layout structure (i.e., strategies, views of teachers, test construction, test administration, challenges teachers encounter, etc), items design while stressing on good and relevant questions. Simply put, validity was achieved by first developing the questionnaire instrument and assigned to research experts and the researcher's supervisor for extensive assessment. All the criticisms raised were duly addressed and the drafted questionnaire underwent pre-testing.

Reliability of Instrument

A reliability test (Cronbach's alpha test) which is a scale used to measure the internal consistency of multiple Likert questions in a survey to determine the reliability was further carried out to check for reliability of the research instrument. According to Beins & McCarthy (2019) the degree to which a research instrument generates consistent results when used multiple

times in different locations and at different times is referred to as reliability. Internal consistency of the questionnaire items was determined using Cronbach's alpha (α) in connection to the study. According to previous studies, the closer the value of Cronbach alpha (α) is to 1, the more dependable the research instrument is (Beins & McCarthy, 2017; Bolarinwa, 2015). A threshold of 0.7 or higher, on the other hand, is generally considered acceptable. The statistical product for service solution (SPSS versions 21.0) software was employed for the computation. Prior to the actual data collection exercise; pretesting was conducted using 234 responses obtained from mathematics teachers selected secondary education improvement project schools in Wa. The data obtained was then processed using the IBM SPSS Statistics and analyzed using the “reliability analysis” tool. The α score was then reported and tested against the proposed threshold. (See appendix B).

Data Collection Procedures

A letter of introduction was collected from the Department of Education and Psychology, University of Cape Coast, to seek for permission from the Head teachers in the various secondary improvement project schools in the Region. The questionnaires were administered by the researcher to two-hundred and thirty four teachers (234) in the Upper West Region. The researcher had also established the necessary contacts with the head teachers of the selected schools to seek permission to administer the questionnaire. A brief self-introduction was made by the researcher to explain the purpose of the study to the respondents before the questionnaires were distributed to them. The researcher stayed with them and had interactions with them. The researcher has appealed to all the respondents to take their time to read the questionnaire and respond to it

appropriately. The researcher also explained to the respondents the purpose of the study and assured them of their confidentiality. The researcher also explains areas that need clarifications and some teachers were given the chance to send home and answered.

Data processing and analysis

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

Chapter Summary

The research was quantitatively motivated. The design employed for the study was the descriptive survey design. Data were analyzed using Inferential (ANOVA and sample-t- test), and descriptive statistics (means and standard deviations).

The demographic data from the questionnaires were statistically analyzed and computed using an alpha(α) level of 0.05 to achieve statistical significance. Research question 1, 2, 3 and 6 were analyzed with one way analysis of variance to compare the score mean of the independent variable performance of public junior high school students). (Pallant,2004).Also, data was collected and coded by the use of quantitative methodology which aid in the generation of the descriptive data and inferential statistics. Hypotheses 1 tested using One-way Analysis of Variance (ANOVA) in order to determine whether there is statistical evidence that, the associated population means are significantly different. Data cleaning and screening are carried out to ensure that

there are no missing values and that the data to be coded is consistent. (Treiman, 2014). According to Treiman (2014) these procedures improve data analysis accuracy while guaranteeing that data analysis methodologies' assumptions are not violated. They added that, checking for data accuracy is vital for verifying irregular responses, means, standard deviations and values for credibility purposes.

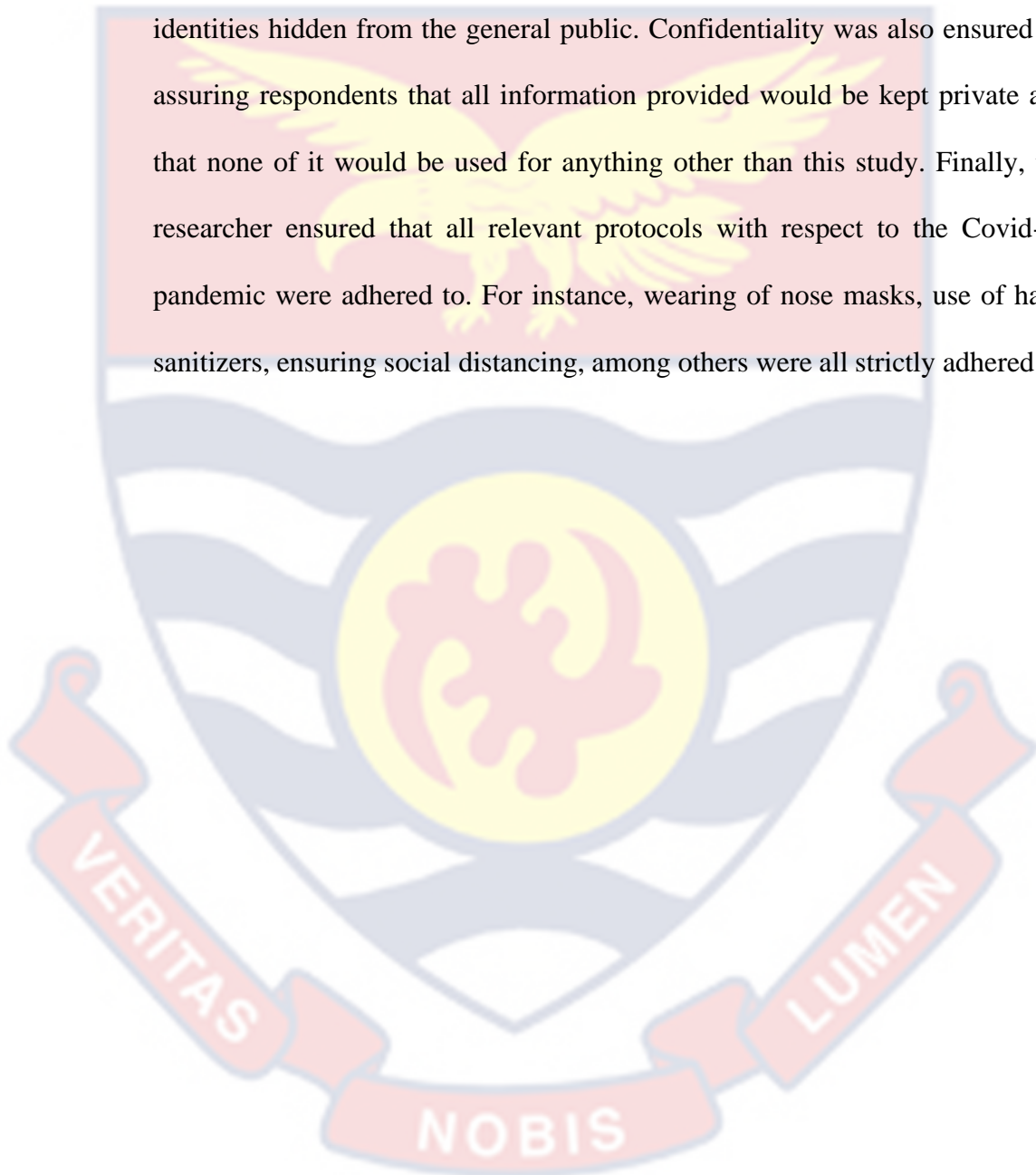
Data Processing and Analysis

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

Ethical considerations

Informed consent, voluntary participation, right to privacy, plagiarism, anonymity, and confidentiality are among the fundamental rules of ethics in data gathering proposed by Neuman & Robson (2014). In terms of informed consent, the respondents were informed of their participation in the data gathering process. This was accomplished in practice by first securing consent from the various junior high schools under investigation. Each questionnaire was also accompanied by a copy of the authority notice. In terms of voluntary participation, none of the respondents were coerced into taking part in the survey against their will. In addition, the respondents' right to privacy was protected by allowing them to choose whether or not to participate in the study. All relevant information obtained from various sources was paraphrased and properly referenced to avoid plagiarism (in-text and end-text). The study was then put

through a plagiarism test to see whether there is any evidence of plagiarism. Anonymity was ensured by removing all personally identifiable information, such as names and other sensitive personal data that could identify the respondents. These safeguards were implemented to keep the respondents' identities hidden from the general public. Confidentiality was also ensured by assuring respondents that all information provided would be kept private and that none of it would be used for anything other than this study. Finally, the researcher ensured that all relevant protocols with respect to the Covid-19 pandemic were adhered to. For instance, wearing of nose masks, use of hand sanitizers, ensuring social distancing, among others were all strictly adhered to.



CHAPTER FOUR

RESULTS AND DISCUSSIONS

Introduction

This chapter presents the results and discussion of the analysis of data for the study. The analysis and interpretation of data were based on the results of five (5) research questions and two (1) hypothesis formulated for the study.

The first part of the analysis described the demographic characteristics of the respondents using frequencies, percentages and graphs (bar and pie). In the second part, the research findings are presented based on the research questions and the hypothesis formulated for the study. This chapter presented an analysis of the data gathered from the field in relation to achievement test practices of teachers in the upper west region. The study aimed at finding out whether senior high teachers in the Upper West Region practice the basic principles of construction, administration and scoring of achievement tests. The data was analyzed using descriptive statistics (frequencies, percentages, means and standard deviations) and inferential statistics (ANOVA). The analysis and interpretation of the data were conducted through the lens of five research questions and two hypotheses that were meticulously formulated for this study. The initial segment of the analysis presents a comprehensive overview of the participants' demography. To address the research questions, descriptive statistics, encompassing means, standard deviations, frequencies, and percentages, were employed as analytical tools. Additionally, to scrutinize the hypotheses formulated, a one-way analysis of variance was utilized as a method of examination.

Presentation of Results**Table 2: demographic distribution of participants**

Variable	Freq.	Percent. %
Gender		
Male	21	70.0
Female	09	30.0
Qualification		
Diploma	01	3.33
Bachelor's degree	24	80.0
Postgraduate Certificate	01	3.33
Masters	02	6.67

N=234

The table presents data on gender and qualification distribution within a certain context. It offers information on the number of individuals belonging to each gender category and qualification level, as well as the percentage representation of each category relative to the total number of individuals.

At a glance, the table portrays a distinct gender disparity within the sampled population. Males constitute a significant majority, accounting for 70% of the total, while females comprise a smaller share at 30%. This observation underscores the potential presence of gender-related imbalances in the context under study. This gender skew could be reflective of prevailing societal or occupational trends that either encourage or inhibit certain genders' participation. Turning to the qualification distribution, the data reveals a compelling narrative about the educational attainment of individuals within the sample. The overwhelming majority - 80%- of the individuals hold a

"Bachelor's degree," indicating a substantial prevalence of individuals with at least an undergraduate education. This prevalence might signal the significance of higher education within the context, potentially suggesting that individuals with advanced educational backgrounds are more likely to be included in the sampled population. Beyond the dominance of "Bachelor's degree" holders, the table also showcases a minority representation of individuals with other qualification levels. The presence of "Diploma" holders, "Postgraduate Certificates," and "Masters" degree holders, though relatively smaller in number, suggests a varied educational landscape within the sample. These individuals could potentially contribute specialized expertise or advanced knowledge to the context, providing a diverse array of skills and perspectives.

Description of Respondents

This section discussed the background information of selected Mathematics teachers. The results are presented below.

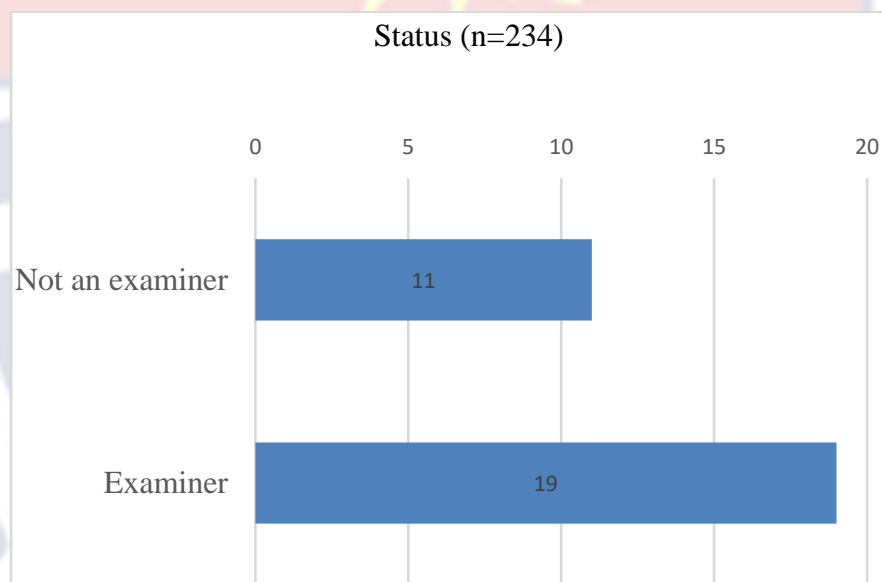
Table 3: Characteristics of Respondents

Variable	Sub-scale	Freq.	Percent. %
Gender	Male	21	70.0
	Female	09	30.0
Qualification	Diploma	01	3.33
	Bachelor	24	80.0
	Postgraduate Certificate	01	3.33
	Masters	02	6.67

Others 00 0.00

Source: Field data, 2022. n=234

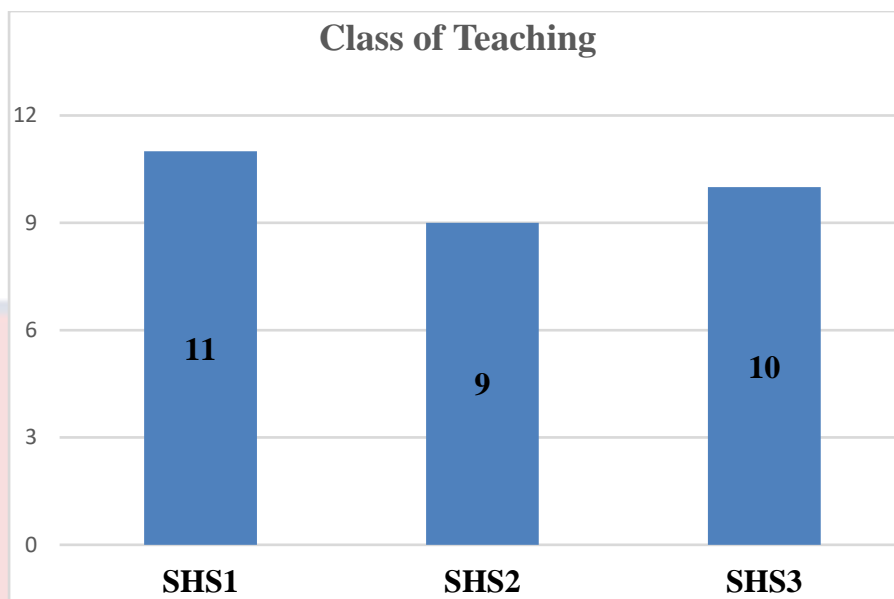
Table 3 contains the characteristics of respondents who took part in the study. From Table 3, there were more male teachers (n=21, 70.0%) as compared to females (n=09, 30.0%). The table further revealed that most of the teachers were Bachelors degree holders (n=24, 80.0%). Only one had Diploma and postgraduate certificate (n=01, 3.33%). Two of the teachers had Masters (=02, 6.66%).



Source: Field data, 2022.

Figure 1: The Bar chart showing the status (examiner or not examiner) of the teachers.

Figure 1 showed that most of the teachers were examiners (n=19, 63.33%). Only eleven of the teachers were not examiners (n=11, 36.77)



Source: Field data, 2022.

Figure 2: The Bar chart showing the classes the teachers teach

Figure 2 showed that most of the teachers were teaching at SHS 1 (n=11, 36.67%). Nine of them were teaching at SHS 2 (n=09, 30.0%) and 10 of them were teaching at SHS 3 (33.33%).

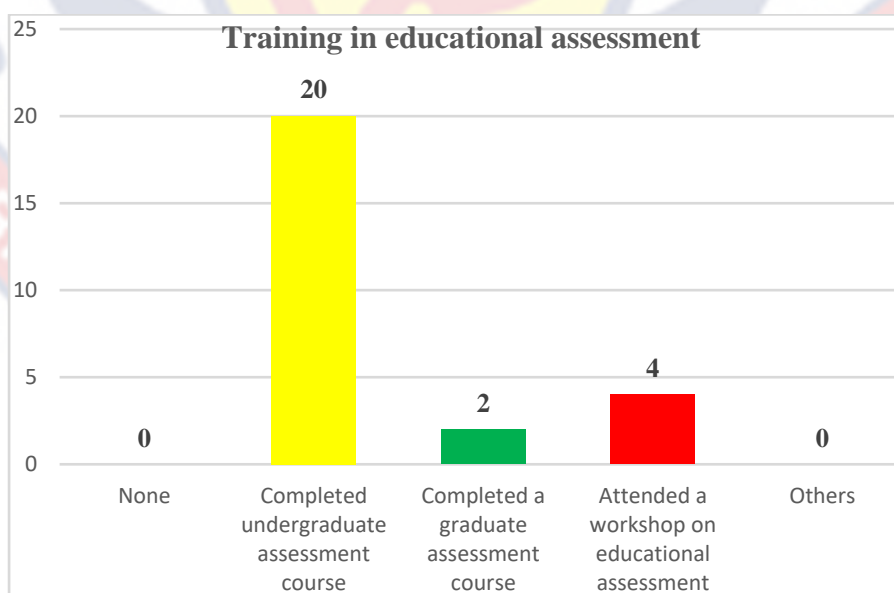


Figure 3: The Bar chart showing teachers training in educational assessment

The results from Figure 3 showed that most of the teachers had their training in educational assessment at the undergraduate level (n=20, 66.67%).

Analysis of the Research Questions

According to Gujarati (2013), descriptive statistics utilize statistical, numerical and graphical methods to look for patterns in a data set. It usually provides the information in a data set by revealing the average

Research Question One: This research question sought to find out views mathematics hold about achievement testing tests. In addressing this research questions (Q1-Q4), means and standard deviations were used for the analysis.

The teachers were given a four-point Likert scale items on teachers views holds about achievement testing for teachers to respond to. The scoring of items was based on the four-point Likert scale of measurement ranging from “Strongly Agree” (scored 4) to “Strongly Disagree” (scored 1). In the analysis, means provides the summary of the responses from teachers and the standard deviation indicates whether teachers’ responses were clustered to the mean score or dispersed. The criterion value (CV) of 2.50 was established for the scale. To obtain the criterion value (CV=2.50), the scores were added together and divided by the number of the scale ($4+3+2+1= 10/4=2.50$) (Green & Neil, 2014). To understand and interpret the mean scores, any items/statements that scored a mean of 2.50 and above indicate respondents’ positive believe of the variables under study while a mean of 2.49 and below indicates a negative believe towards variables under study. The findings are presented as below:

Research Question 1: What views do mathematics teachers in secondary education improvement project hold about achievement testing?

In the quest of achieving the purpose of the study, I evaluated what views teachers hold about achievement testing in the Upper West Region. In achieving this, the responses from the teachers were analyzed using Means and Standard Deviations. The results are presented

Table: 4 Results on what views do teachers hold about achievement testing in SEIP schools in Upper West Region

Table 4: Mathematics Teachers View about Achievement Testing

Statements	M	SD	N
Achievement testing establishes what students have learned	3.9	.34	234
Achievement test measures students higher order thinking	3.8	.55	234
Achievement test results are trustworthy	3.7	.91	234
Number of times Achievement tests are conducted play a major role in students' performance	3.7	.63	234
Achievement test results can be depended on	3.7	.14	234
Achievement tests provide information on how well schools are performing	3.6	.71	234
Achievement testing is an integral part of teaching	3.6	.40	234
Quality of achievement test questions influence students' performance	3.5	.13	234
Achievement tests provide feedback to students about their performance	3.5	.16	234
Achievement testing of students' learning should be embedded in the teaching process	3.4	.17	234

Achievement testing places students into categories	3.4	.10	234
Achievement tests results should be treated cautiously because of measurement error	3.4	.53	234
Achievement testing should be on going	3.2	.42	234
Teachers should take into account the error factor in conducting all achievement tests	3.2	.19	234
Achievement test results show how much students are benefiting from classroom instruction	3.1	.14	234
Achievement testing forces teachers to teach in a way different to their beliefs	2.2	.83	234
Achievement testing brings about unnecessary competition among students	2.1	.71	234
Students will learn better in the absence of test	2.1	.52	234
Achievement testing interferes with teaching	2.1	.51	234
Achievement testing is unfair to students	2.1	.44	234
Achievement test has little impact on teaching and learning	2.0	.54	234
Students' failure in mathematics is indicative of teachers' failure in teaching properly the subject	1.8	.04	234
I carry out achievement testing because it is mandatory to do so	1.3	.32	234
Achievement testing of students is a waste of time	1.0	.57	234

Source: Field data, 2022

Cut-off Mean value=2.50

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

The table contains a series of statements related to achievement testing, along with corresponding mean (M) scores and standard deviation (SD) values. These values have been interpreted based on a criterion: a mean greater than or equal to 2.5 indicates agreement among teachers, while a mean lesser than 2.5 indicates disagreement. The analysis of the table highlights the prevailing trends and patterns in teachers' responses: Central to the findings is the widespread belief in the utility of achievement tests as tools for gauging student learning. With a mean score of 3.95 and a standard deviation of 0.349, teachers overwhelmingly concur that these tests adeptly determine what students have assimilated. Furthermore, they're not just superficial measures. A consensus, supported by a mean score of 3.82 and a standard deviation of 0.553, posits that these tests measure beyond basic recall, assessing students' prowess in higher-order thinking skills. These tests, as indicated by a mean score of 3.51 with a standard deviation of 0.161, are pivotal for evaluating students' academic performance. Parallel to their utilitarian value is a palpable trust in the integrity of these tests. The majority of teachers, as evidenced by a mean of 3.74 and a standard deviation of 0.917, regard achievement test results as trustworthy indicators of student performance. Their dependability isn't taken lightly either; a mean of 3.71 (SD = 0.145) points to a strong confidence in the results' reliability. Moreover, a significant finding, with a mean of 3.68 and a standard deviation of 0.719, indicates that teachers perceive achievement tests as instrumental in providing insights into the broader school performance. Another salient point is the intrinsic link between achievement testing and the teaching process. Not only do teachers see it as an indispensable component (M = 3.64, SD = 0.403), but they also advocate for its continual integration into education,

a sentiment reflected by a mean of 3.28 and a standard deviation of 0.424. However, the influence of external factors on testing outcomes hasn't been overlooked. Participants acknowledged, with a mean score of 3.72 (SD = 0.634), that the frequency of tests can considerably sway students' performance.

The quality of questions, too, emerges as critical, having a mean of 3.59 and a standard deviation of 0.135. Yet, this prevailing positive sentiment doesn't eclipse the concerns that lurk in the background. While the majority lauded the reflective nature of achievement tests on effective classroom instruction (M = 3.14, SD = 0.146), concerns about unnecessary student competition (M = 2.19, SD = 0.719) and potential interference with teaching (M = 2.13, SD = 0.517) surfaced. Furthermore, the strong disagreement, evidenced by a mean of 1.02 and a standard deviation of 0.576, against the notion that achievement testing is a waste of time, emphasizes the nuanced views of educators.

The findings from the present study agree with the assertion of Tom and Gary (2003), who indicated that, teachers believe in the utility of achievement tests as tools for gauging student learning and therefore concluded the achievement testing is best way of assessing student performance. The possible cause of this finding may be teachers' intensive teaching, clear instructions in exams and enough time in taking tests. The study further revealed that teachers often use the correct methods of teaching. This attitude might have a great deal of implication on the understanding of methods of solving mathematics problems on various topics which is related to the standard. The study also revealed that teachers using the right teaching and learning materials during instructional periods. This enables students to know the correct mathematical tool to use in any situation in examination. This might appropriately create competence of the

students' content coverage and their thinking process on a particular topic. The result from the study also revealed that, teachers often guide students in working mathematics to identify shortfalls and weakness of them. This confirms the findings of Quaigrain (1992) who indicated that some teachers believe in the achievement aids in students learning.

The findings on the views of mathematics teachers hold about achievement testing supports the assertion of Wiliam (2008), who stated that, to increase the trustworthiness of a test, teachers consider how successfully the test results serves a purpose efficiently and effectively. He further suggested that test would be not be trustworthy when it is not free from malpractice, teacher not using scoring guide not giving clear instructions and limited time. The use of such test result for any purpose might be faulty because it might not reflect the true academic performance of the examinees. To obtain a comprehensive result, I evaluated the kinds of achievement test strategies that mathematics teachers in SEIP schools use to assess their students' learning outcomes. In accomplishing this, the responses from the teachers were compiled using Means and Standard Deviations. The results are presented in Table 5.

Research Question Two: What kinds of achievement test strategies do mathematics teachers use to assess their students' learning outcomes?

Table 5: Achievement test strategies

Statements	M	SD	N
Class tests/quizzes	3.95	.349	234 S
Objective tests (e.g., multiple choice, short answer, matching)	3.82	.553	234 S
Essay type questions	3.74	.917	234 S

Projects completed by individual students	3.72	.634	234N
Major examinations (e.g., midterm, end of term, end of year)	3.71	.145	234 S

Source: Field data, 2022

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

The exploration of mathematics achievement test tasks and strategies employed in Secondary Education Improvement Project (SEIP) schools is a pivotal endeavor to understand the diverse approaches educators take to assess students' learning outcomes. The detailed analysis and discussion of the survey responses provide a nuanced understanding of the prevailing practices and how they collectively respond to the research question: What kinds of mathematics achievement test tasks or strategies are used to assess students' learning outcomes in SEIP schools?

The substantial proportion (86.7%) of educators utilizing exams or tests in class underscores the significance of traditional assessment methods. These tests provide a snapshot of students' comprehension of specific topics, aiding teachers in gauging the effectiveness of their instructional strategies. This aligns with the research question, indicating that mathematics teachers rely on conventional assessments to measure immediate learning outcomes.

The high preference for unbiased evaluations, such as multiple-choice and short-answer questions, serves as a testament to educators' endeavor to ensure objective and standardized assessments. This aligns with the research

question's intent, revealing that SEIP schools prioritize evaluation methods that provide fair and uniform measurement of student knowledge.

The extensive use (80.0%) of essays-based inquiries suggests a commitment to fostering critical thinking and in-depth understanding among students. These assessments encourage students to articulate their knowledge in-depth, demonstrating their grasp of complex concepts. This approach resonates with the research question's exploration of varied strategies that delve into students' analytical abilities. While a minority (16.7%) of educators employ individual student projects, their inclusion highlights a commitment to practical application of knowledge. These projects require students to synthesize theoretical concepts into real-world scenarios, fostering creativity and problem-solving skills. This strategy responds to the research question by showcasing a diverse range of assessment practices aimed at holistic skill development. The extensive reliance (90.0%) on main examinations, including midterms and end-of-year tests, underscores the importance of comprehensive evaluations that span multiple topics. This aligns with the research question's exploration of assessment methods that encompass a broad array of learning outcomes over a specific period. The detailed analysis and discussion of the survey responses effectively address the research question by revealing a tapestry of assessment methods. Mathematics teachers in SEIP schools employ a combination of traditional assessments, unbiased evaluations, essays-based inquiries, projects, and main examinations to measure students' learning outcomes. This diverse approach showcases their commitment to holistic education, encompassing immediate understanding, critical thinking, practical application, and comprehensive knowledge. The findings of the study clearly indicate that SEIP

schools recognize the multifaceted nature of learning outcomes and tailor their assessment strategies to align with these varying dimensions. Through a balanced combination of assessment methods, educators respond to the research question by offering a comprehensive evaluation of students' mathematical abilities, ensuring a well-rounded educational experience within the SEIP school environment. From the results, it is evident that few of the achievement test strategies are used. Some of the strategies include: Class tests/quizzes (M=3.95, SD=.349, n=234); Objective tests (e.g., multiple choice, short answer, matching (M=3.82, SD=.553, n=234); Essay type questions (M=3.74, SD=.917, n=234).

Some of the kinds of achievement test strategies mathematics teachers used to assess their students learning outcomes' include Projects work (M=3.72, SD=.634, n=234); Major examinations (e.g., midterm, end of term, end of year)(M=3.71, SD=.145, n=234)

It was also revealed in the study that education systems that emphasize tests and examinations put some student at a disadvantage (Mbano, 2003; Nampota & Wella, 1999).The findings indicated that alternative and authentic evaluation have more acceptances from students and should, therefore, be viewed as an alternative to traditional standardized assessment. The study again revealed that evaluation practices in some subject areas like Mathematics indicated favourable emphasis being given to formative evaluation because 80% of the total marks have been allocated to on-going assessment and 20% was for the test. Moreover, students interviewed also agreed that project and portfolio assignment given were to a great extent real and authentic tasks that they could relate to their future workplace.

It was also revealed in the study that education systems that emphasize tests and examinations put some student at a disadvantage (Mbanjo, 2003; Nampota & Wella, 1999). Fook & Sidhu (2010) conducted a study in Malaysia to investigate the different types of authentic assessment used in higher education in Malaysia.

The researchers identified that different types of authentic assessment were used. The study revealed that teachers employed the following assessment tools; portfolio (15%), article review (17%) performance product (24%), project (50%) and test (28%). Moreover, students interviewed also agreed that project and portfolio assignment given were to a great extent real and authentic tasks that they could relate to their future workplace.

Research Question Three: What are challenges encounter in practicing achievement testing?

I evaluated challenges by teachers encounter in practicing of achievement testing in the upper west region.

Table 6: Results on challenges teachers encounter in practicing achievement testing

Statements	M	SD	N
Inadequate knowledge on how to construct achievement test hinders testing of students learning outcomes	3.84	.234	234
Inadequate time to gather information and materials to use hinders achievement testing	3.72	.112	234
Loaded timetable is a factor for not testing students	3.64	.303	234
Inadequate teacher motivation on the part of school authority hinders achievement testing	3.62	.142	234

Inadequate knowledge on how to administer achievement test is a factor for not testing students	3.58	.244	234
The school system does not encourage free and fair achievement testing	3.55	.549	234
Inadequate time to complete the mathematics syllabus hinders achievement testing	3.52	.134	234
Some topics are actually difficult to test students on	3.49	.232	234
Large class size makes it difficult to assess students through achievement test	3.43	.023	234
The school assessment system makes it difficult to use achievement test	3.41	.445	234
Inadequate knowledge on the usage of achievement test results hinders achievement testing	3.39	.713	234
Inadequate knowledge on how to score achievement test is a factor for not testing students	3.38	.132	234
Developing mathematics achievement test is a difficult task	3.35	.562	234
Inadequate knowledge on the usage of achievement test results hinders achievement testing	3.34	.159	234
Some forms of achievement testing are not carried out because of inadequate funds	3.23	.257	234
Poor supervision affects achievement testing	3.13	.512	234

Source: Field data, 2022

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

The table present results on challenges encounter in practicing of achievement testing by mathematics teachers. The results showed that there are

numerous challenges that confront the use of achievement test. Some of the challenges included: inadequate knowledge on how to construct achievement test hinders testing of students learning outcomes ($M=3.83$, $SD=.234$, $n=234$). Another challenge was that there was inadequate support from school authority in terms of logistics and facilities affected achievement testing ($M=3.72$, $SD=.112$, $n=234$). Loaded timetable was also a major factors ($M=3.64$, $SD=.303$, $n=234$). It was also found that inadequate teacher motivation on the part of school authority hindered achievement testing ($M=3.62$, $SD=.142$, $n=234$). It was again found that inadequate knowledge on how to administer achievement test is a factor for not testing students ($M=3.58$, $SD=.244$, $n=234$). The table further revealed that the school system does not encourage free and fair achievement testing ($M=3.55$, $SD=.549$, $n=234$). Inadequate time to complete the mathematics syllabus hindered achievement testing ($M=3.52$, $SD=.134$, $n=234$). Lastly, many teachers admitted that some topics are actually difficult to test students on ($M=3.49$, $SD=.232$, $n=234$).

Eshun et al. (2014) conducted a study to investigate the influence of achievement test on achievement test practices of teachers and the challenges they encounter in the Social Studies classroom in Ghana. The study used a descriptive case study design and it involved 10 senior high schools and twenty teachers randomly sampled from thirty (30) senior high schools in the Upper West Region of Ghana. Semi-structured interview guide was the main instrument used for data collection. The research found out that the forms of achievement test some teachers used in their classrooms were limited due to examination policies, time, resources and assessment methods employed by their schools. Furthermore, they revealed that most teachers they observed were

not using assessment techniques that involved students in the teaching and learning process. Again, they indicated that some teachers revealed that using the achievement test would delay them in completing topics in their syllabuses given to them. Beckmann, Senk & Thompson (1997) in their study conducted in USA identified three reasons why teachers do not use multiple assessment methods. First, some teachers had limited knowledge of different forms of evaluation. Second, teachers felt they had no time to create/develop authentic assessment.

Research Question four: How do mathematics teachers in the secondary education improvement project schools adhere to the construction of test items?

In the quest of achieving the purpose of the study, I assessed how mathematics teachers in secondary education improvement project schools adhere to the principles of construction of test items.

Table 7: Results on how mathematics teachers in secondary education improvement project schools follow the principles of construction of test items.

Statements	M	SD	N
Specify the construct to be measured	3.59	.174	234
Questions reflect the construct to be measured	3.55	.253	234
Write more test items than required	3.45	.567	234
Follow the principles of test construction for each format used	3.44	.443	234
Provide clear instructions on how test items should be answered	3.43	.139	234

Consider the time to be spent on the test items	3.38	.132	234
State the purpose of the test	3.36	.322	234
Construct test items when it is time to assess	3.34	.563	234
Consider students' language maturity	3.32	.335	234
Construct test items from the scheme of work	3.28	.146	234
Avoid leading questions	3.24	.153	234
Write items weeks before the actual date for the test	3.23	.123	234
Use a test specification table	3.20	.245	234
Ask colleagues in the subject area to help review the items	3.19	.123	234
Prepare marking scheme after administering the test items to students	3.15	.145	234
Consider the weightings of the profile dimensions in the construction of test items	3.13	.542	234
Ask colleagues in the subject area to help construct the items	3.12	.134	234
Achievement test items are designed primarily by yourself	3.09	.432	234
Sometimes ask students to construct their own questions	3.03	.165	234
Construct test items during teaching	3.01	.160	234
Set test items from past questions	2.38	.234	234
Teacher solves the questions to determining the time required	1.85	.166	234
Arrange test items in ascending order of difficulty	1.82	.246	234
Use test items from test books	1.62	.122	234
Double barrel questions are necessary in Achievement testing	1.41	.146	234

Source: Field data, 2022

N=234

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

Table presents results on how mathematics teachers in SEIP reported that they adhere to the principles of construction of test items. In the table are some of the test construction principles: (M=3.59, SD=.174, n=234). Again, teachers were skilled in constructing test that reflected construct to be measured (M=3.55, SD=.252, n=234). Furthermore, the teachers knew how to write required test items (M=3.45, SD=.567, n=234). Most of the teachers indicated that they followed the principles of test construction for each format used (M=3.44, SD=.443, n=234). Also, teachers were skilled in providing clear instructions on how test items are answered (M=3.44, SD=.139, n=234). The teachers again indicated that they considered time duration for test items (M=3.38, SD=.132, n=234) and stated the purpose of the test (M=3.36, SD=.322, n=234).

Similar to the above results, it was found that many teachers were skilled constructing test items at appropriate times (M=3.34, SD=.563, n=234) and considered students' language maturity (M=3.32, SD=.335, n=234). Again, many teachers were skilled in constructing test from the scheme of work (M=3.28, SD=.146, n=234) and avoided leading questions (M=3.24, SD=.153, n=234). Lastly, teachers were skilled in writing items weeks before the actual date for the test (M=3.23, SD=.123, n=234).

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

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

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

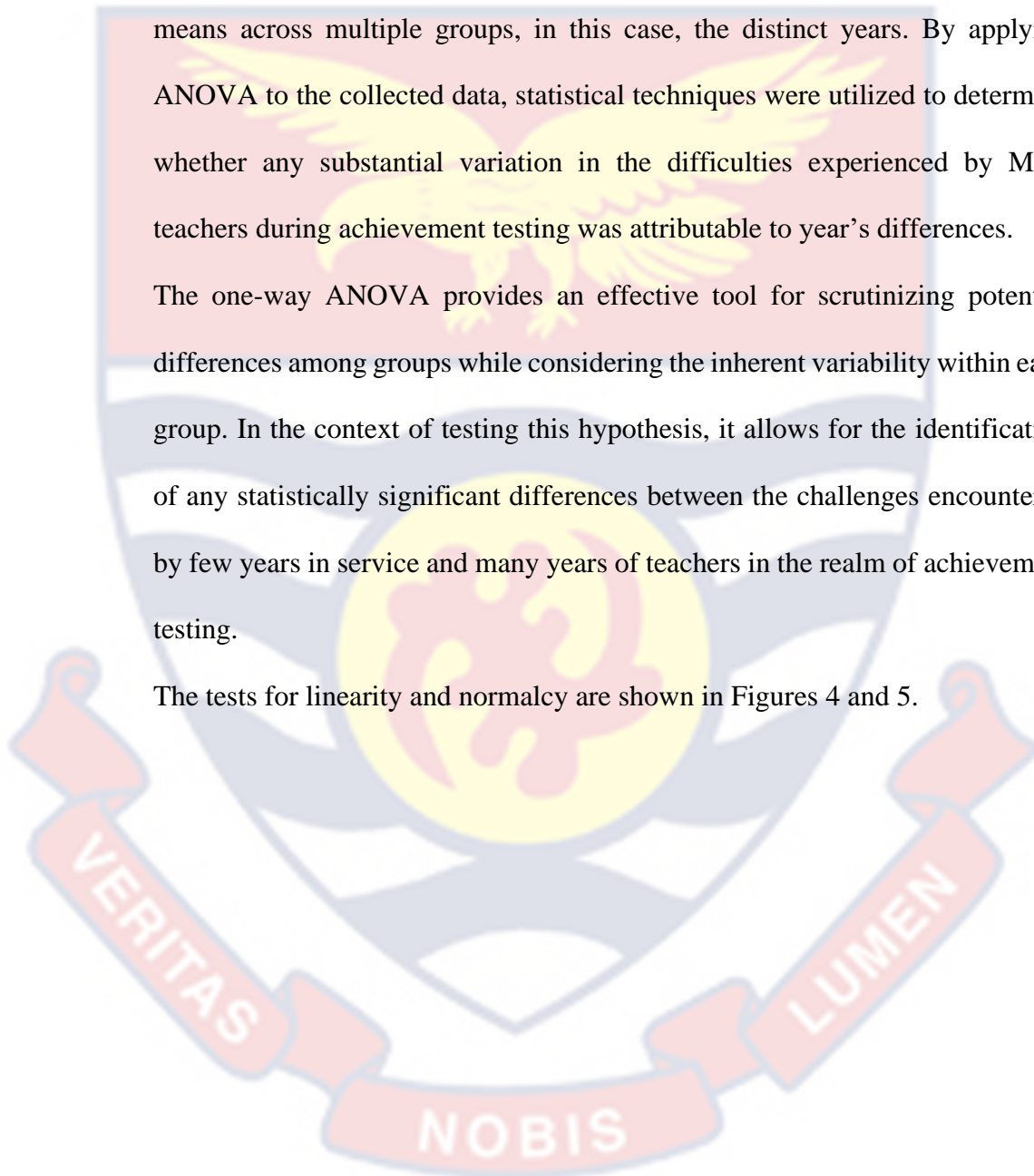
H01: There is no statistically significant difference in the challenges faced by SEIP Mathematics teachers in achievement testing based on teaching experience. The objective of testing the hypothesis is to ascertain whether there exists a significant difference in the difficulties encountered by Math teachers in achievement testing, contingent upon years of experience. This endeavor

seeks to explore whether gender has any discernible impact on the challenges faced by educators when engaging in the practice of achievement testing.

To assess the null hypothesis, a one-way analysis of variance (ANOVA) was employed as the analytical approach. ANOVA facilitates the comparison of means across multiple groups, in this case, the distinct years. By applying ANOVA to the collected data, statistical techniques were utilized to determine whether any substantial variation in the difficulties experienced by Math teachers during achievement testing was attributable to year's differences.

The one-way ANOVA provides an effective tool for scrutinizing potential differences among groups while considering the inherent variability within each group. In the context of testing this hypothesis, it allows for the identification of any statistically significant differences between the challenges encountered by few years in service and many years of teachers in the realm of achievement testing.

The tests for linearity and normalcy are shown in Figures 4 and 5.



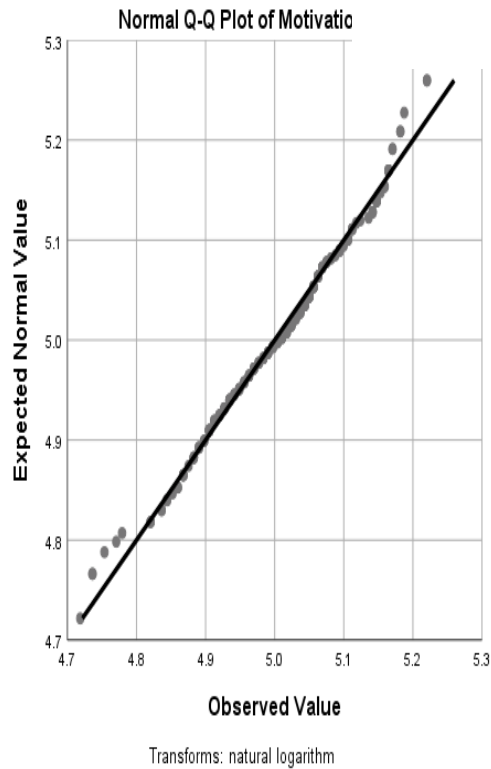


Figure 4 Diagnostic test for linearity and normality

Field study is its source (2022)

Indicators of normality and linearity, according to Pallant (2007), include a straight normal probability plot. It delivers a trustworthy result, he said, when several normalcy assumptions are satisfied.

In Figure 5, the plot of the data showing the normality and linearity of the data between the two variables could be observed as a good straight line.

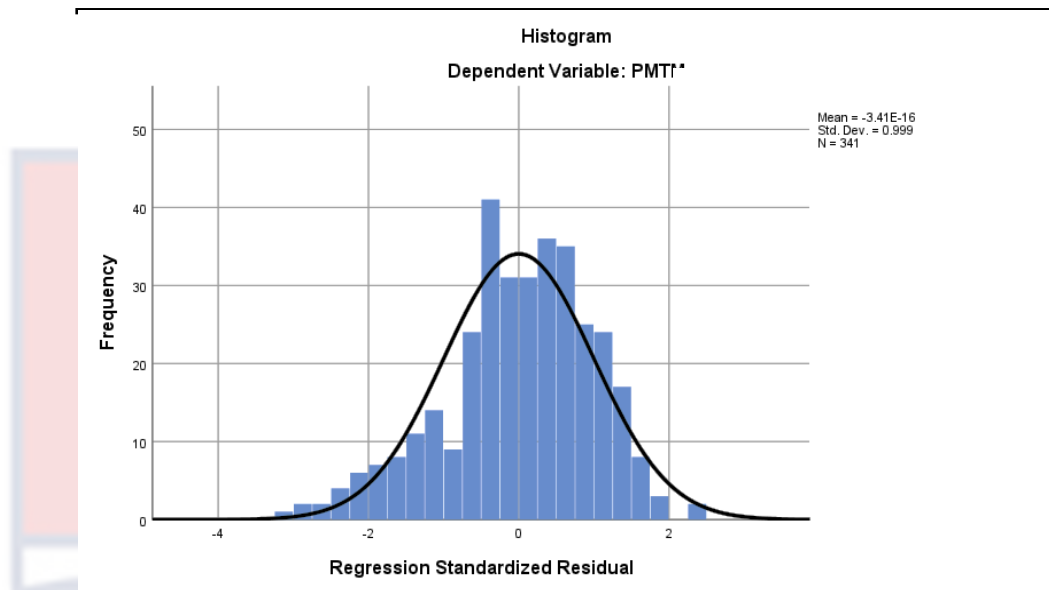


Figure 5: Histogram Test of Normality and Linearity

Origin: Field Study (2022)

The data met the requirements of normality of variance and linearity, and the residuals were roughly normally distributed, according to the Histogram plot of standardized predicted values versus standardized residuals.

Table 8: Normality Test Results of the Variables

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	Df	Sig. value	Statistic	Df	Sig. value
Years Range						
0-3 years	.067	06	.106(ns)*	.865	06	.245(ns)

*

4-6 years	.187	11	.095(ns)*	.295	11	.353(ns) *
7-10 years	.051	09	.081(ns)*	.143	09	.542(ns) *
11 years and above	.463	04	.070(ns)*	.191	04	.075(ns) *

*. This represents a lower limit of the real significance, a.

The source for the Lilliefors Significance Correction is a field survey conducted in 2022 (.187, df=11, p=.095**, p>0.05, n=30).

Teachers with tenures of 7 to 10 years also reported a Kolmogorov-Smirnova test result of .051, df=09, p=.081**, p>0.05, n=30, showing that the data was normally distributed.

Finally, teachers with at least 11 years of classroom experience stated that the data was regularly distributed with a Kolmogorov-Smirnova test value of .463, df=04, p=.070**, p>0.05, and n=30. This indicates that the distribution of the data was normal.

Prior to evaluating if the independent variables were homogeneous, the researcher first confirmed that the data were normal.

Table 8 presents the findings P0.05, n=30, indicates a significant difference. The Kolmogorov-Smirnova test findings in Table 8 are presented under the presumption that the sample size is smaller than 50 (n 50). If the significant value is more than 0.05, the Kolmogorov-Smirnova test findings are considered to be normal. According to the findings of the Kolmogorov-Smirnova test, the

dependent variable (challenges of achievement testing) was normally distributed across years of teaching experience. For instance, teachers who have been in the classroom for 0–3 years received a Kolmogorov–Smirnova result indicating that the data was normally distributed (Kolmogorov–Smirnova test value=.067, df=06, p=.106**, $p>0.05$, n=30); teachers who have been in the classroom for 4–6 years also received a result indicating that the data was normal (Kolmogorov–Smirnova indicating that the data was normally distributed (Kolmogorov-Smirnov^a test value= .067, df=06, p=.106**, $p>0.05$, n=30); teachers who have taught for 4 - 6yrs also recorded a Kolmogorov-Smirnov^a results indicating that it was normal (Kolmogorov-Smirnov^a test value= .187, df=11, p=.095**, $p>0.05$, n=30).

Also, teachers who have taught for 7-10 years reported a Kolmogorov-Smirnov^a test result signifying that the data was normally distributed (Kolmogorov-Smirnov^a test value = .051, df=09, p=.081**, $p>0.05$, n=30). Finally, teachers who have taught for 11 years and above reported a Kolmogorov-Smirnov^a test result suggesting that the data was normally distributed (Kolmogorov-Smirnov^a test value = .463, df=04, p=.070**, $p>0.05$, n=30). After checking the normality of the data, the researcher proceeded to check the homogeneity across the independent variables. The results are presented in Table 9

Table 9: Homogeneity of Variances Test of the independent variables

Parameters of Estimate	Levene			
	Statistic	df1	df2	Sig. value

Using the Mean	.395	4	26	.093(ns)*
Using the Median	.340	4	26	.068(ns)*
Using the mean and with adjusted df	.430	4	26.521	.078(ns)*
Based on trimmed mean	.395	4	26	.053(ns)*

Field Data as a Source (2022)

- $P < 0.05$, $n=30$, indicates a significant difference.

The test to determine whether the independent variable's variances are homogeneous is shown in Table 9. The results of the homogeneity of variances test indicated from Table 9 that the assumption of homogeneity had not been broken using means values. [t ($df_1=4$; $df_2=26$) = .395, Sig. value = .093; $p > 0.05$, 2-tailed] Thus, it was statistically justified to do a between-groups one-way analysis of variance (ANOVA) test.

Results of the descriptive statistics test for the ANOVA are shown in Table.

Table 10: Descriptive Statistics of the Study Variables (QoL*YiW)

Years Range	N	M	Std. D	Std. Error	Ranking
0-3 years	06	58.07	4.190	.199	1 st
4-6 years	11	43.03	4.712	.902	2 nd
7-10 years	09	32.37	3.922	.703	3 rd

11 years and above	04	16.45	6.130	.722	4 th
--------------------	----	-------	-------	------	-----------------

Field Data, n=30, 2022

The descriptive data the table highlight the disparities between the mean scores for years of teaching experience and the difficulties math teachers have while administering achievement tests. For instance, math instructors with more than three years of experience had the highest mean scores (M= 58.07, SD= 4.190, Std. Error=.199, n=06), showing that instructors with less years of experience (0-3) were facing the most difficulties with accomplishment testing. The data also showed that those between the ages of 4 and 6 faced some significant obstacles in achievement testing (M= 43.03, SD= 4.712, Std Error= 902, n= 11). Achievement assessment revealed some minor difficulties for those between the ages of 7 and 10 (M=32.37, SD=3.922, Std Error= 703, n=09). In accomplishment assessment, those who have been teaching for more time faced very few difficulties (M= 16.45, SD= 6.130, Std Error= 722, n=04).

To gather more statistical proof that the observed difference was not a result of chance, a one-way analysis of variance (ANOVA) was carried out.

Table 11: Summary of One-way Analysis of Variance (ANOVA)

Results

Sources	Sum of Squares	Df	Mean Square	F-value	Sig.	Rks
---------	----------------	----	-------------	---------	------	-----

Between Groups	1777.073	4	464.268	1.445	.000(s	Diff.
Within Groups	41819.690	26	266.052)*	exited
Total	43596.763	30				

Source: Field Information (2022)

At $p < 0.05$ and $n = 230$, there is a significant difference. To compare the mean scores of the independent variables, a one-way Analysis of Variance (ANOVA) was carried out. The results of the between-groups one-way analysis of variance (ANOVA) in Table 11 demonstrate that there is a statistically significant difference between the years of teaching experience and difficulties Mathematics teachers encounter in achievement testing [$F(df_1=4, df_2=26) = 1.445, p = .000^{**}, n=30, 2\text{-tailed}$]. This provides statistical proof that the mean scores for the various years of teaching experience varied significantly. It is clear from the ANOVA findings that the significant value of 0.000^* (2-tailed) did not indicate where the disparities between the years of teaching experience exist. Consequently, the Post Hoc test was performed to determine the statistically significant differences.

Table 12: Post-Hoc Test among the years of teaching Experiences

(I) Years of teaching experiences	(J) Years teaching experiences	Mean Difference (I-J)	Std. Error	Sig. value

0-3 years	4-6 years	-8.323*	.144	.000(s) *
	7-10 years	-4.859*	.351	.000(s) *
	11 years and above	-4.443*	.459	.000(s) *
4-6 years	0-3 years	8.323*	.133	.000(s) *
	7-10 years	3.464*	.186	.000(s) *
	11 years and above	3.880*	.299	.000(s) *
7-10 years	0-3 years	4.859*	.151	.000(s) *
	4-6 years	-3.464*	.186	.000(s) *
	11 years and above	.4157	.308	.003(s) *
11 years and above	0-3 years	4.448*	.279	.000(s) *
	4-6 years	-3.881*	.299	.000(s) *

7-10 years	-0.4157	.308	.003(s) *
------------	---------	------	--------------

*. A 0.05 level of significance is used to the mean difference.

Field Data as a Source (2022)

The post-Hoc test's findings are shown in the table. (Games-Howell).

The Post-Hoc test identifies the areas where the teachers' varying years of teaching experience differ. The results are statistically significant when comparing 0-3 years and 7-10 years, for instance, where the mean difference and standard error of (MD=-8.323*, SR=.133) with a significant value of 0.000* (2-tailed) show that there was a difference. Additionally, the results between ages 0 to 3 and 11 and older demonstrate a significant difference (p=0.000**), with a mean difference and standard error of (MD=-4.443*, SR=.279). The researcher rejects the prior assertion that there is no statistically significant difference between the years of teaching experience and difficulties mathematics teachers face in achievement testing to accept that there is statistically significant difference among the years of teaching experiences and difficulties mathematics teachers face in achievement testing in the Upper West Region based on the significant differences of the results found.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction

The study sought to investigate if mathematics teachers in secondary education improvement project School teachers in the in the Upper West Region follow the basic prescribed principles in the area of construction and administration of achievement tests. The study was guided by the following research objectives:

1. Evaluate how mathematics teachers in the Upper West Region follow the principles of test:
 - a. construction
 - b. administration
2. Find out the kinds of achievement test strategies mathematics teachers in the in the Upper West Region use to assess their students' learning outcomes.

Mathematics teachers in the Upper West Region encounter in the use of achievement test.

4. Assess difference among the years of teaching experience of mathematics teachers in Schools in the Upper West Region and it gender influence in practicing of achievement testing.

A descriptive sample survey was conducted in Upper West Region using questionnaire as the data collection instrument. Stratified proportionate sampling, random sampling and purposive sampling were used to select two

hundred and thirty four (234) Senior High School teachers from the Upper West Region for the study. The analysis focused on descriptive statistics that involved computing of frequencies, percentages, means and standard deviations. The hypothesis one was analyzed using One Way Analysis of Variance (ANOVA) and the hypothesis two was analyzed using Sample t test.

Key Findings

The results show that generally, mathematics teachers in the Senior High Schools teachers in the Upper West Region averagely follow most principles of test construction ($M=2.38$, $SD=.234$). This may be as a result of some teachers relying on already existing questions instead of constructing the items on themselves. Some teachers copy test items directly from text books. Yet these problems may happen due to inadequate knowledge of teachers in test constructions. Because if teachers have knowledge in the principles of test constructions they will know that it is not good to use already constructed items to assess their students.

The results show that generally, majority of the mathematics teachers in the upper West Region averagely follow test administration guidelines in their achievement testing ($M=3.76$ $SD=.432$). The researcher found out from the study that some teachers do invigilation during examinations, guard students in solving mathematics problems and more teaching and learning materials during teaching. The results gave evidence that most mathematics in teachers in SEIP schools in the Upper West Region do not face more challenges in practicing of achievement testing with regard to their years of teaching experience and this always improve the achievement test results. It was revealed that most teachers have been able to instill discipline in class. It was also showed that most teachers

face challenges with large class size and therefore most of the assignments are always marked and submitted to students. From the results, it was evident that most of the achievement test strategies were used among mathematics teachers in the Upper West Region ($M= 3.71$ $SD= .145$).

1. The study revealed that there various views was held by mathematics teachers among mathematics teachers in Secondary Education Improvement project Schools in the Upper West Region. A lot of them believe that achievement testing establishes what students learn, help in order of high thinking ($M=3.95$ $SD= .349$). From the hypothesis, the tested hypothesis suggested non-significant difference existed among the years of working experience and test construction among the teachers. Hence, null hypothesis which states that, “the results gave statistical evidence to the effect that there were significant differences in mean scores of the years of teaching experiences ($df_1=4$, $df_2=26$) $=.1.445$, $p = .000^{**}$, $n=30$, 2-tailed). The results practically means that the challenges the teachers face in achievement testing is influenced or by years in teaching experience.

Conclusions

It was evident from the findings of the study that teachers in the Secondary Education Improvement schools in Upper West Region were well equipped with test construction, skills. Teachers having such a sensitive responsibility of achievement test practices and making decision concerning achievement testing’ it proper practice are expected to be professional in the process of achievement testing strategies. However, teachers engaging in some negative test practices when constructing test items and administering as well the test

items maybe that they are comfortable with such practices without recognizing the impact of their practices on issues of validity and reliability.

There was also clear evident that numbers years of teaching does not influence the challenges in quest of practicing achievement testing in secondary Education Improvement Project Schools in the Upper West Region.

Recommendations

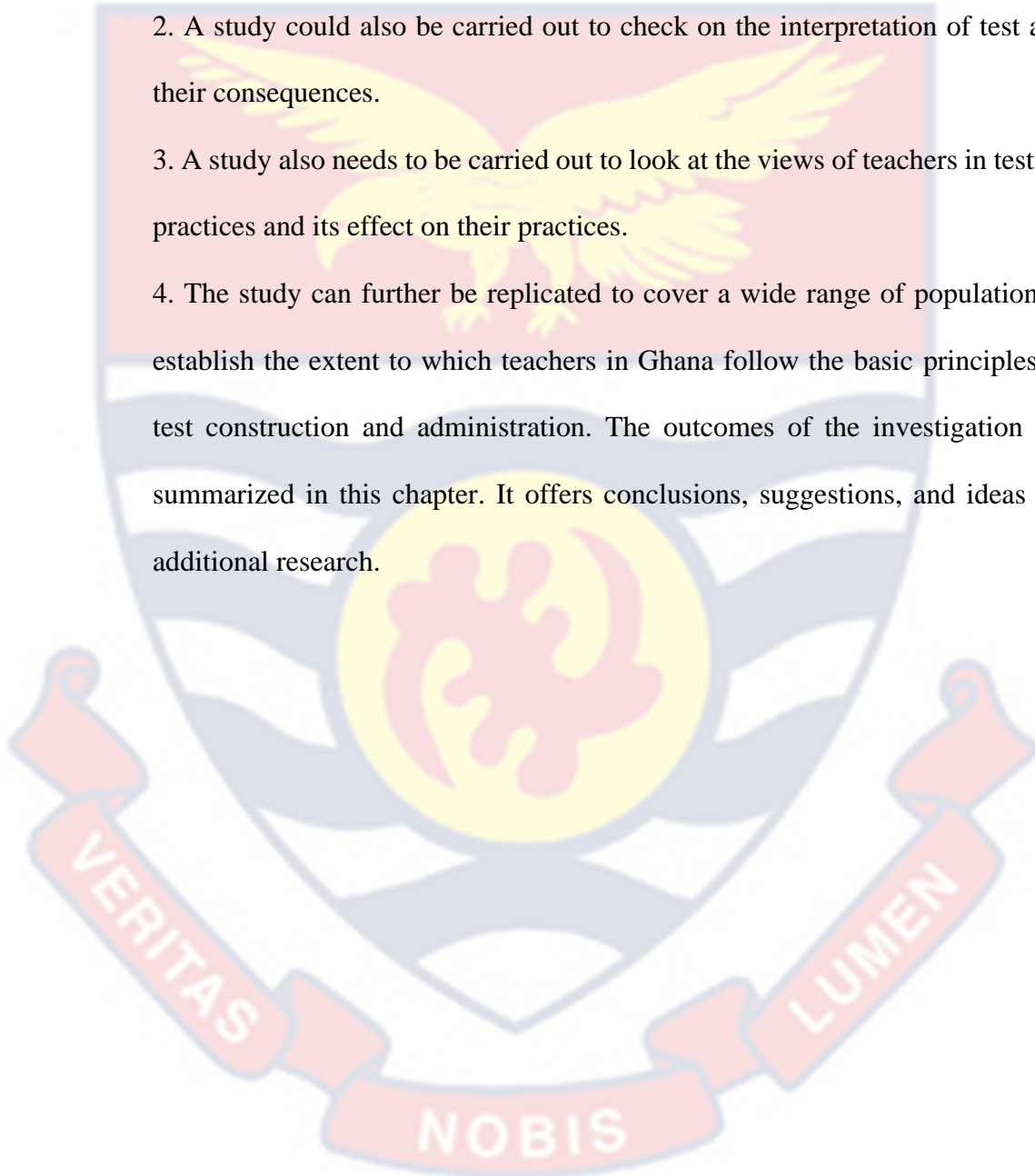
With respect to the findings resulting from the study, the following recommendations are made for the improvement of testing practices among mathematics teachers in the Upper West Region.

1. I suggest, regular workshops and in-service training should be organized by the Ghana Education Service for mathematics teachers in Senior High Schools on how to plan achievement test (especially test construction and administration) effectively. This could be achieved through the collaboration of the ministry of education, the institute of education and other stakeholders of education.
2. Since it was evidence from the findings that teachers use paper and pen as the only strategy to assess their students, teachers are encouraged to use of other equally important assessment strategies such as class test, major examinations. This will help to assess the students as a whole.
3. There should be an intensive monitoring by head teachers and other supervisors of education on how teachers practice achievement test. Head teachers should ensure that teachers provide test specification table for the test items they construct and make sure that other test practice principles are followed.

Suggestions for Future Research

The following are suggested for future research:

1. A study could be carried out to look into testing practices in terms of strategies used by teachers to assess students' outcomes.
2. A study could also be carried out to check on the interpretation of test and their consequences.
3. A study also needs to be carried out to look at the views of teachers in testing practices and its effect on their practices.
4. The study can further be replicated to cover a wide range of population to establish the extent to which teachers in Ghana follow the basic principles of test construction and administration. The outcomes of the investigation are summarized in this chapter. It offers conclusions, suggestions, and ideas for additional research.



REFERENCES

- Adom, D., Mensah, J. A., & Dake, D. A. (2020). Test, measurement, and evaluation: Understanding and use of the concepts in education. *International Journal of Evaluation and Research in Education*, 9(1), 109-119.
- Alexander, S. P., Christopoulos, A., Davenport, A. P., Kelly, E., Mathie, A., Peters, J. A. & Yao, C. (2019). The Concise Guide to PHARMACOLOGY 2019/20: G protein-coupled receptors. *British journal of pharmacology*, 176, S21-S141.
- Al-Ibadi, S. L. J. (2019). The Effect of Table of self (KWLH) Strategy on the Achievement of Second Class Intermediate School Girl Students at History. *Diyala Journal of Human Research*, 2(79).
- Al-Mahrooqi, R., & Denman, C. (2018). Alternative assessment. *The TESOL encyclopedia of English language teaching*, 10, 9781118784235.
- Alrfooh, A. E. (2012). Types of Achievement Tests Which Are Preferred By Outstanding Students at Al-Hussein Bin Talal University.
- Al-Saigh, Z. Y., & Munk, P. (1984). Study of polymer-polymer interaction coefficients in polymer blends using inverse gas chromatography. *Macromolecules*, 17(4), 803-809.
- Amedahe, F. K. (2000). Issues in combining continuous assessment scores with external examination scores for certification at the basic and secondary school levels: *Journal of Educational Management*, 3, 110-127.
- Amedahe, F. K., & Asamoah-Gyimah, A. (2003). Introduction to educational research for diploma in basic education.

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

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

Amrock, L. G., Neuman, M. D., Lin, H. M., & Deiner, S. (2014). Development and validation of a simple risk model incorporating a chart-derived frailty score. *Journal of the American College of Surgeons*, 219(4), 684-694.

Atweh, B., Bleicher, R. E., & Cooper, T. J. (1998). The construction of the social context of mathematics classrooms: A sociolinguistic analysis. *Journal for Research in Mathematics Education*, 29(1), 63-82.

Bagshaw, C. R., Trentham, D. R., Wolcott, R. G., & Boyer, P. D. (1975). Oxygen exchange in the gamma-phosphoryl group of protein-bound ATP during Mg^{2+} -dependent adenosine triphosphatase activity of myosin. *Proceedings of the National Academy of Sciences*, 72(7), 2592-2596.

Bushway, A., & Nash, W. R. (1977). School cheating behavior. *Review of Educational Research*, 47(4), 623-632.

Bashaija, A., Atibuni, D. Z., & Rukundo, A. (2021). Alcohol Consumption and its Socio-demographic Correlates among Secondary School Teachers in Uganda.

Balabanis, G., & Diamantopoulos, A. (2004). Domestic country bias, country-of-origin effects, and consumer ethnocentrism: a multidimensional unfolding approach. *Journal of the academy of marketing science*, 32(1), 80-95.

Barton, P. E., & Coley, R. J. (1994). Testing in America's Schools. Policy Information Report.

Bejar, I. I. (1984). Educational diagnostic assessment. *Journal of educational measurement*, 21(2), 175-189.

Bergold, J., & Thomas, S. (2012). Participatory research methods: A methodological approach in motion. *Historical Social Research/Historische Sozialforschung*, 191-222.

Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, 24(3), 200-212.

Blanton, R. L., & Nunnally, J. C. (1964). Evaluational language processes in the deaf. *Psychological Reports*, 15(3), 891-894.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008).

The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools (No. w14021). National Bureau of Economic Research.

Bois, P. R., Borgon, R. A., Vornrhein, C., & Izard, T. (2005). Structural dynamics of α -actinin-vinculin interactions. *Molecular and cellular biology*, 25(14), 6112-6122.

- Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, 40(1), 90-98.
- Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, 22(4), 195-201.
- Brophy, J. (1986). Teacher influences on student achievement. *American psychologist*, 41(10), 1069.
- Brempong, M. B., Norton, U., & Norton, J. B. (2019). Compost and soil moisture effects on seasonal carbon and nitrogen dynamics, greenhouse gas fluxes and global warming potential of semi-arid soils. *International Journal of Recycling of Organic Waste in Agriculture*, 8, 367-376.
- Calderhead, J. (1996). Anwar, M. N. (2018). Exploring Creative Thinking of Undergraduate Students through the Lens of Selected Demographic Variables. *NICE Research Journal*, 15-22.
- Carrier, C. A., & Titus, A. (1981). Effects of notetaking prôt raining and test mode expectations on learning from lectures. *American Educational Research Journal*, 18(4), 385-397.
- CDC, A. (1998). Measles, mumps, and rubella--vaccine use and strategies for elimination of measles, rubella, and congenital rubella syndrome and control of mumps: recommendations of the Advisory Committee on Immunization Practices (ACIP). *Mmwr*, 47(RR-8), 1-57.
- Chand, R. (2017). Doubling Farmers' Income Rationale, Srategy, Prospects and Action Plan.

Cizek, M., Horáček, J., & Domcke, W. (1998). Nuclear dynamics of the H collision complex beyond the local approximation: associative detachment and dissociative attachment to rotationally and vibrationally excited molecules. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 31(11), 2571.

Clark, A. (2000). *Mindware: An introduction to the philosophy of cognitive science*. Oxford University Press.

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

Coleman, J. S. (1991). *Parental involvement in education*. US Department of Education, Office of Educational Research and Improvement, Programs for the Improvement of Practice.

Cooper, D. R., Schindler, P. S., Cooper, D. R., & Schindler, P. S. (2003). *Business research methods*.

Cohen, S., Manion, L., & Morrison, K. (1972). 2011. *Folk Devils and Moral Panics*.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2012). *Algebra for 8th graders: Evidence on its effects from 10 North Carolina districts (No. w18649)*. National Bureau of Economic Research.

Creswell, J. W., & Tashakkori, A. (2007). Differing perspectives on mixed methods research. *Journal of mixed methods research*, 1(4), 303-308.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of educational research*, 58(4), 438-481.

Cronbach, L. J. (1960). *Essentials of psychological testing (2e ed. rev.)*.

Cowan, C. A., Klimanskaya, I., McMahon, J., Atienza, J., Witmyer, J., Zucker, J. P., ... & Melton, D. A. (2004). Derivation of embryonic stem-cell lines from human blastocysts. *New England Journal of Medicine*, 350(13), 1353-1356.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological science*, 12(2), 163-170.

Davis, G. F. (1991). Agents without principles? The spread of the poison pill through the interoperate network. *Administrative science quarterly*, 583-613.

Davidson, N. C. (2014). How much wetland has the world lost? Long-term and recent trends in global wetland area. *Marine and Freshwater Research*, 65(10), 934-941.

Dawson, C. (2002). *Practical research methods: A user-friendly guide to mastering research*. How to books.

Delandshere, G., & Jones, J. H. (1999). Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum and Supervision*, 14(3), 216.

DePascale, C. (2022). *Multiple Measures*. Routledge.

Ebel, R. L., & Frisbie, D. A. (1972). *Essentials of educational measurement*.

Ernest, P. (1994). The philosophy of mathematics and the didactics of mathematics. *Didactics of mathematics as a scientific discipline*, 335-350.

Elton, L. R. B., & Laurillard, D. M. (1979). Trends in research on student learning. *Studies in higher education*, 4(1), 87-102.

Etsey, K. (2005, November). Causes of low academic performance of primary school pupils in the Shama Sub-Metro of Shama Ahanta East Metropolitan Assembly (SAEMA) in Ghana. In Proceedings of the Regional Conference on Education in West Africa.

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

Ephraim, R. K. D., Osakunor, D. N. M., Denkyira, S. W., Eshun, H., Amoah, S., & Anto, E. O. (2014). Serum calcium and magnesium levels in women presenting with pre-eclampsia and pregnancy-induced hypertension: a case-control study in the Cape Coast metropolis, Ghana. *BMC pregnancy and childbirth*, 14(1), 1-8.

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

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

Geske, D. H., Ragle, J. L., Bambenek, M. A., & Balch, A. L. (1964). Study of Steric Effects by Electron Spin Resonance Spectroscopy and Polarography. Substituted Nitrobenzenes and Nitroanilines. *Journal of the American Chemical Society*, 86(6), 987-1002.

Ghaicha, A. (2016). Theoretical Framework for Educational Assessment: A Synoptic Review. *Journal of Education and Practice*, 7(24), 212-231.

Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects.

In Optimising new modes of assessment: In search of qualities and standards (pp. 37-54). Dordrecht: Springer Netherlands.

Gitomer, D. H., & Duschl, R. A. (2007). Indicator Systems: Establishing Multilevel Coherence in Assessment. *Teachers College Record*, 109(13), 288-320.

Gipps, C., & Stobart, G. (2009). Fairness in assessment. In *Educational assessment in the 21st century: Connecting theory and practice* (pp. 105-118). Dordrecht: Springer Netherlands.

Gobert, J. D., & Pallant, A. (2004). Fostering students' epistemologies of models via authentic model-based tasks. *Journal of Science education and Technology*, 13, 7-22.

Grotzinger, J. P., Sumner, D. Y., Kah, L. C., Stack, K., Gupta, S., Edgar, L., ... & Sirven, J. B. (2014). A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale Crater, Mars. *Science*, 343(6169), 1242777.

Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. *Educational Measurement: Issues and Practice*.

Gujarati, A., Cerqueira, F., & Brandenburg, B. B. (2013, July). Outstanding paper award: Scheduled ability analysis of the linux push and pull scheduler with arbitrary processor affinities. In *2013 25th Euromicro Conference on Real-Time Systems* (pp. 69-79). IEEE.

Hambleton, R. K., Roberts, D. M., & Traub, R. E. (1970). A comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple-choice test. *Journal of Educational Measurement*, 7(2), 75-82.

- Harlen, W. (2006). The role of assessment in developing motivation for learning. *Assessment and learning*, 61-80. at History. *Diyala Journal of Human Research*, 2(79).
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, 95(7-8), 798-812.
- Haynie III, W. J. (1992). Post hoc analysis of test items written by technology education teachers. Volume 4 Issue 1 (fall 1992).
- Henderson, P., & Karr-Kidwell, P. J. (1998). *Authentic Assessment: An Extensive Literary Review and Recommendations for Administrators*.
- Hill, K. T., & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *The elementary school journal*, 85(1), 105-126.
- Hira, T. K., & Mugenda, O. M. (1999). The relationships between self-worth and financial beliefs, behavior, and satisfaction. *Journal of family and consumer sciences*, 91(4), 76.
- Hudson, J. G., & Frisbie, P. R. (1991). Cloud condensation nuclei near marine stratus. *Journal of Geophysical Research: Atmospheres*, 96(D11), 20795-20808.
- Ingersoll, R. M. (2012). Beginning teacher induction what the data tell us. *Phi delta kappa*, 93(8), 47-51.
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, 17(2), 109-138.
- Jones, B. D., & Egley, R. J. (2007, September). Learning to take tests or learning for understanding? Teachers' beliefs about test-based accountability.

In The Educational Forum(Vol. 71, No. 3, pp. 232-248). Taylor & Francis Group.

Jönsson, S., & Grönlund, A. (1988). Life with a sub-contractor: new technology and management accounting. *Accounting, Organizations and Society*, 13(5), 513-532.

Kankam, B., Bordoh, A., Eshun, I., Kweku Bassaw, T., & Yaw Korang, F. (2014). Teachers' perception of authentic assessment techniques practice in social studies lessons in senior high schools in Ghana.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education review*, 27(6), 615-631.

Kara, F., & Çelikler, D. (2015). Development of Achievement Test: Validity and Reliability Study for Achievement Test on Matter Changing. *Journal of Education and Practice*, 6(24), 21-26.

Klassen, A. C., Creswell, J., Plano Clark, V. L., Smith, K. C., & Meissner, H. I. (2012). Best practices in mixed methods for quality of life research. *Quality of life Research*, 21, 377-380.

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American journal of health-system pharmacy*, 65(23), 2276-2284.

Klimko, I. P. (1984). Item arrangement, cognitive entry characteristics, sex, and test anxiety as predictors of achievement examination performance. *The Journal of experimental education*, 52(4), 214-219.

- Koch, N., Fuss, S., Grosjean, G., & Edenhofer, O. (2014). Causes of the EU ETS price drop: Recession, CDM, renewable policies or a bit of everything?—New evidence. *Energy Policy*, 73, 676-685.
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and psychological measurement*, 30(3), 607-610.
- Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of educational research*, 58(1), 79-97.
- Kubiszyn, T., & Borich, G. (1984). *Educational Testing & Measurement*. Glenview, Illinois: Scott Foresman and Company.
- LaFontana, K. M., & Cillessen, A. H. (2002). Children's perceptions of popular and unpopular peers: a multimethod assessment. *Developmental psychology*, 38(5), 635.
- La Notte, A., D'Amato, D., Mäkinen, H., Paracchini, M. L., Liqueste, C., Egoh, B., ... & Crossman, N. D. (2017). Ecosystem services classification: A systems ecology perspective of the cascade framework. *Ecological indicators*, 74, 392-402.
- Lester, F. K., Garofalo, J., & Kroll, D. L. (1989). Self-confidence, interest, beliefs, and metacognition: Key influences on problem-solving behavior. In *Affect and mathematical problem solving: A new perspective* (pp. 75-88). New York, NY: Springer New York.
- Liesbet, H. O. O. G. H. E., & Gary, M. (2003). Unraveling the central state, but how? Types of multi-level governance. *American political science review*, 97(2), 233-243.

- Lissitz, R. W., & Schafer, W. D. (2002). *Assessment in Educational Reform: Both Means and Ends*. College Division, Allyn & Bacon, Inc., 75 Arlington Street, Suite 300, Boston. MA 02116.
- Luo, Y. X., Wu, S. C., Gilat, J., Rasmussen, J. O., Hamilton, J. H., Ramayya, A. V., ... & Gelberg, A. (2004). Level structures of Rh 110, 111, 112, 113 from measurements on Cf 252. *Physical Review C*, 69(2), 024315.
- Marso, R. N., & Pigge, F. L. (1993). 6. Teachers' testing knowledge, skills, and practices.
- Mbano, N. (2003). The effects of a cognitive acceleration intervention programme on the performance of secondary school pupils in Malawi. *International Journal of Science Education*, 25(1), 71-87.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of applied psychology*, 79(4), 599.
- Meyer, F., Goesmann, A., McHardy, A. C., Bartels, D., Bekel, T., Clausen, J., & Pühler, A. (2003). GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic acids research*, 31(8), 2187-2195.
- Mehrens, W. A., & Lehmann, I. J. (1987). *Using standardized tests in education*. Longman/Addison Wesley Longman.
- Mehrens, W. A., & Lehmann, I. J. (1984). *Measurement and evaluation*. Education and Psychology, 1991-145.
- Moore, L. L., Grabsch, D. K., & Rotter, C. (2010). Using achievement motivation theory to explain student participation in a residential

leadership learning community. *Journal of leadership education*, 9(2), 22-34.

Monk, J. J., & Stallings, W. M. (1970). Effects of Item Order on Test Scores. *The Journal of Educational Research*, 63(10), 463-465.

Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile Issues in Teachers Professional Development*, 14(1), 143-158.

Munson, R., & Parton, C. (2013). Bias and fairness in state testing.

Mussawy, S. A. J. (2009). Assessment practices: Student's and teachers' perceptions of classroom assessment.

Murphy, T. H., & Corbett, D. (2009). Plasticity during stroke recovery: from synapse to behaviour. *Nature reviews neuroscience*, 10(12), 861-872.

Nampota, D., & Wella, K. (1999). Report on primary school teacher development workshop.

Newman, D. L., Kundert, D. K., Lane Jr, D. S., & Bull, K. S. (1988). Effect of varying item order on multiple-choice test scores: Importance of statistical and cognitive difficulty. *Applied Measurement in education*, 1(1), 89-97.

Newman, F. M., & Wehlage, G. G. (1993). Five standards for authentic learning. *Educ Lead*, 50(879), 15-19.

Nitko, A. J. (1996). Educational assessment of students. Prentice-Hall Order Processing Center, PO Box 11071, Des Moines, IA 50336-1071.

Nitko, A. J. (2001). Conceptual frameworks to accommodate the validation of rapidly changing requirements for assessments. *Curriculum and assessment*, 1, 143-163.

Nugba, R. M., Amedahe, F. K., & Asamoah-Gyimah, K. Role of Input, Process Factors and School Context on the Effectiveness of Junior High School Education in the Central Region of Ghana.

Oduro-Okyireh, G. (2008). Testing practices of senior secondary school teachers in the Ashanti region of Ghana (Doctoral dissertation, University of Cape Coast).

Oduro-Okyireh, G., & Partey, D. N. L. (2014). Effects of assessment on classroom learning in the Ashanti Mampong Municipality of the Ashanti Region of Ghana. *The International Journal of Humanities and Social Studies*, 2(7), 150-155.

Othman, B., Harun, A., Rashid, W., Nazeer, S., Kassim, A., & Kadhim, K. (2019). The influences of service marketing mix on customer loyalty towards Umrah travel agents: Evidence from Malaysia. *Management Science Letters*, 9(6), 865-876.

Parikh, P. P., Kanabar, M. G., & Sidhu, T. S. (2010, July). Opportunities and challenges of wireless communication technologies for smart grid applications. In *IEEE PES general meeting* (pp. 1-7). IEEE.

Pedersen, M. M., Marso, E., & Pickett, M. J. (1970). Nonfermentative bacilli associated with man: III. Pathogenicity and antibiotic susceptibility. *American Journal of Clinical Pathology*, 54(2), 178-192.

Pehkonen, E. (1995). *Pupils' View of Mathematics: Initial Report for an International Comparison Project. Research Report 152*. University of Helsinki, Department of Teacher Education, PO Box 38 (Ratakatu 6A), Helsinki 00014, Finland.

Pehkonen, E. (2001). A hidden regulating factor in mathematics classrooms: Mathematics-related beliefs. *Research on mathematics and science education: From beliefs to cognition, from problem solving to understanding*, 15-39.

Pehkonen, E., & Kaasila, R. (2011). Students' conceptions On Effective Mathematics Teaching. *Current State Of Research On Mathematical Beliefs XVI*, 218.

Pirttiniemi, P., Poikela, A., Huggare, J., & Löppönen, T. (2004). Dental maturation in children with shunt-treated hydrocephalus. *The Cleft palate-craniofacial journal*, 41(6), 651-654.

Plake, B. S., & Wise, L. L. (2014). What is the role and importance of the revised AERA, APA, NCME Standards for Educational and Psychological Testing?. *Educational Measurement: Issues and Practice*, 33(4), 4-12.

Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in Senior High Schools in Ghana: Document analysis. *International Journal of Assessment Tools in Education*, 6(1), 1-8.

Quagrain, A. K. (1992). Teacher competence in the use of essay tests: A study of secondary schools in the Western Region of Ghana. Unpublished Master's Thesis, University of Cape Coast, Ghana.

Quansah, Frank, Isaac Amoako, and Francis Ankomah. "Teachers' test construction skills in Senior High Schools in Ghana: Document analysis." *International Journal of Assessment Tools in Education* 6, no. 1 (2019): 1-8.

Reeves, G. D., McAdams, K. L., Friedel, R. H. W., & O'Brien, T. P. (2003). Acceleration and loss of relativistic electrons during geomagnetic storms. *Geophysical Research Letters*, 30(10).

Ritchie, D., & Thorkildsen, R. (1994). Effects of accountability on students' achievement in mastery learning. *The Journal of Educational Research*, 88(2), 86-90.

Rogers, E. M., & Svenning, L. (1969). Modernization among peasants: The impact of communication. *Modernization among peasants: the impact of communication*.

Rue, R. C., & Rosenshine, M. (1985). The application of semi-Markov decision processes to queueing of aircraft for landing at an airport. *Transportation science*, 19(2), 154-172.

Rukundo, A., & Magambo, J. (2010). Effective test administration in schools: Principles and good practices for test administrators. *African Journal of Teacher Education*, 1(1).

Rukundo, A., & Magambo, J. (2011). Quality assessment: implications and applications in schools and colleges in Uganda. *Journal of Educational Sciences & Psychology*, 1(1).

Sawilowsky, S. S. (2000). Psychometrics versus datametrics: Comment on Vacha-Haase's "reliability generalization" method and some EPM editorial policies. *Educational and Psychological Measurement*, 60(2), 157-173.

Saunders, M., Lewis, P. H. I. L. I. P., & Thornhill, A. D. R. I. A. N. (2007). *Research methods. Business Students 4th edition* Pearson Education Limited, England, 6(3), 1-268

Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for research in Mathematics Education*, 28(2), 187-215.

Smith, K. T., Tilset, M., Kuhlman, R., & Caulton, K. G. (1995). Reactions of $(\text{PiPr}_3)_2\text{OsH}_6$ Involving Addition of Protons and Removal of Electrons. Characterization of $(\text{PiPr}_3)_2\text{Os}(\text{NCMe})_x\text{Hyz}^+$ ($x=0, 2, 3$; $y=1, 2, 3, 4, 7$; $z=1, 2$), Including Dicationic. *Journal of the American Chemical Society*, 117(37), 9473-9480.

Smith, S. M., & Brady, J. M. (1997). SUSAN—a new approach to low level image processing. *International journal of computer vision*, 23(1), 45-78.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: issues and practice*.

Stude, D. E., & Gullickson, J. (2001). The effects of orthotic intervention and 9 holes of simulated golf on gait in experienced golfers. *Journal of Manipulative and Physiological Therapeutics*, 24(4), 279-287.

Stainback, S. B. E., & Stainback, W. C. (1996). *Inclusion: A guide for educators*. Paul H Brookes Publishing.

Songnalle, S. (2019). *Formative Assessment Practices of Senior High School Teachers in the Upper West Region of Ghana* (Doctoral dissertation, University of Cape Coast).

Sultana, U., Khairuddin, A. B., Sultana, B., Rasheed, N., Qazi, S. H., & Malik, N. R. (2018). Placement and sizing of multiple distributed generation

and battery swapping stations using grasshopper optimizer algorithm. *Energy*, 165, 408-421.

Tamakloe, E. K. (1997). *An Evaluation of the National Conference on Teacher Education, 1986*. African Books Collective, Ltd., The Jam Factory, 27 Park End Street, Oxford OX1 1HU, United Kingdom.

Thomas, A. E., & Grimes, J. E. (2002). *Best practices in school psychology IV, Vols. 1-2*. National Association of School Psychologists.

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

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53.

Thomas, G. L., & Thorne, R. E. (2003). Acoustical-optical assessment of Pacific herring and their predator assemblage in Prince William Sound, Alaska. *Aquatic Living Resources*, 16(3), 247-253.

Troeger, C., Khalil, I. A., Rao, P. C., Cao, S., Blacker, B. F., Ahmed, T., ... & Reiner, R. C. (2018). Rotavirus vaccination and the global burden of rotavirus diarrhea among children younger than 5 years. *JAMA pediatrics*, 172(10), 958-965.

Tsegay, G., Edris, M., & Meseret, S. (2013). Assessment of voluntary counseling and testing service utilization and associated factors among Debre Markos University Students, North West Ethiopia: a cross-sectional survey in 2011. *BMC public health*, 13, 1-7.

- Van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism practice*, 6(5-6), 648-658.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.
- Viberg, O., & Grönlund, Å. (2012). Mobile assisted language learning: A literature review. In 11th World Conference on Mobile and Contextual Learning.
- Wahyuni, D. (2012). The research design maze: Understanding paradigms, cases, methods and methodologies. *Journal of applied management accounting research*, 10(1), 69-80.
- Weiwei, C., Runlin, G., Lisheng, L., Manlu, Z., Wen, W., Yongjun, W., & Zhengpei, Z. (2016). Outline of the report on cardiovascular diseases in China, 2014. *European Heart Journal Supplements*, 18(suppl_F), F2-F11.
- Wheaton, K. J., & UNITED STATES EMBASSY THE HAGUE (NETHERLANDS). (2000). Success! Security Assistance And Its Impact On Croatia, 1995-2000. *DISAM Journal*.
- Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological review*, 114(4), 843.
- Yang, N., Shaheen, N., Xie, L., Yu, J., Ahmad, H., & Mao, H. (2019). Pesticide residues identification by optical spectrum in the time-sequence of enzyme inhibitors performed on microfluidic paper-based analytical devices (μ PADs). *Molecules*, 24(13), 2428.

Yan, X., Wang, W. D., Walters, A. S., Wang, Q., Liu, Y. J., & Chu, F. Y. (2012).

Traditional Chinese medicine herbal preparations in restless legs syndrome (RLS) treatment: a review and probable first description of RLS in 1529. *Sleep Medicine Reviews*, 16(6), 509-518.

Yeboah, A., Gyamfi, A., Wintson, D. K., & Prempeh, A. D. (2022). Item analysis, reliability estimates, percentile ranks and stanine scores among university students in Ghana. *Adv Soc Sci Res J*, 9(8), 42-60.

Young-Loveridge, J. (2010). A decade of reform in mathematics education: Results for 2009 and earlier years. Findings from the New Zealand Numeracy Development Projects 2009, 15-35.

Ysseldyke, J. (2001). Reflections on a research career: Generalizations from 25 years of research on assessment and instructional decision making. *Exceptional children*, 67(3), 295-309.

Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2003). Research methods. *Health economics research method*, 2.

Zhu, W., Michalsky, R., Metin, O., Lv, H., Guo, S., Wright, C. J., ... & Sun, S. (2013). Monodisperse Au nanoparticles for selective electrocatalytic reduction of CO₂ to CO. *Journal of the American Chemical Society*, 135(45), 16833-16836.

APPENDICES

APPENDIX A: QUESTIONNAIRE FOR RESPONDENTS

Dear Respondents

The study seeks to assess **Assessment of mathematics teachers' practice of achievement testing in secondary education improvement project schools in Upper West Region**. Your full input will help to make informed decisions about the **achievement testing**. It will be appreciated if you can **honestly** provide responses to **all the** items. You are assured of complete **confidentiality** and **anonymity** of all the information you will provide. Your participation in this study is **completely voluntary**. Again, questions on this questionnaire have gone through a thorough review and have been declared **ethical** for educational research.

Thank You.

Please, kindly tick [] the appropriate response

Section A: Demographic Characteristics.**1. Indicate your sex. (Tick [] one only)**

Male []

Female []

2. Educational Qualification:

a) Teachers' Certificate A []

b) Diploma with Education []

c) Bachelors with Education []

d) Bachelors without Education []

e) Masters with Education []

f) Masters without Education []

g) Others, specify.....

3. What was your programme of study at that level?

.....

4. Number of years in teaching service

a) Under 5 years []

b) 6 – 10 years []

c) Above 11 years []

5. Indicate your status. (Tick [√] one only).

Examiner

Not an examiner

6. What class do you teach? (Tick [√] one only)

SHS 1

SHS 2

SHS 3

1. Do you have any training in educational assessment? (Tick [√] where applicable)

None

Completed undergraduate assessment course

Completed a graduate assessment course

Attended a workshop on educational assessment

Others,

specify.....

Section B: Teachers Knowledge On Views Of Achievement Testing

This part of the survey looks at the views/ beliefs mathematics teachers hold on achievement testing. Express your opinion on the subject using the following rating scale: **Strongly Agree = SA; Agree = A; Neutral = N; Disagree = D; and Strongly Disagree = SD**; choose the response that comes closest to describing your opinion. (Tick [] only one).

Item No.	Views/Beliefs about achievement testing	SA	A	N	D	SD
1	Achievement tests provide information on how well schools are performing					
2	Achievement test results show how much students benefit from classroom instruction					
3	Achievement tests provide students with feedback on their performance					
4	Achievement testing is an integral part of teaching					
5	Achievement testing places students into categories					
6	Achievement test results are trustworthy					

7	Achievement testing forces teachers to teach in a way that is different from their beliefs					
8	Achievement tests results should be treated cautiously because of measurement error					
9	Achievement testing establishes what students have learned					
10	Students will learn better in the absence of test					
11	Achievement testing is unfair to students					
12	Teachers should take into account the error factor when conducting all achievement tests					
13	Achievement test measures students higher order thinking					
14	Achievement test results is reliable					
15	Achievement testing interferes with teaching					
16	Achievement test has little impact on teaching and learning					

17	The failure of students in mathematics is an indication of the failure of teachers to properly teach the subject					
18	I carry out achievement testing because it is mandatory to do so					
19	Achievement testing brings unnecessary competition among students					
20	Using achievement testing to test students is a waste of time					
21	Achievement testing should be on going					
22	Number of times Achievement tests are conducted play a major role in students' performance					
23	Quality of achievement test questions influence students' performance					
24	Testing students' learning should be embedded in the teaching process					

Section C: Teachers Knowledge Kind of Achievement Test Strategies

This part of the questionnaire basically looks at **achievement test tasks/strategies** mathematics teachers **use** in testing students learning outcomes. Express your opinion on the subject using **Yes** or **No**; and choose the response that comes close to describe your opinion. (Tick [\surd] only one).

Item No.	Achievement test tasks/strategies	Yes	No
25	Class tests/quizzes		
26	Objective tests (e.g., multiple choice, short answer, matching)		
27	Essay type questions		
28	Projects completed by individual students		
29	Major examinations (e.g., midterm, end of term, end of year)		
30	“What Do You Know”		
31	Group project		
32	Homework		
33	Class assignments		
34	Presentations		
35	Authentic tests (e.g., “real world” performance task)		
36	Performance tests (e.g., structured teacher observations or ratings of students’ performance on a task)		
37	Assessing students’ work samples		

Section D: Teachers Knowledge on Challenges Teachers Face in Quest of Practicing Achievement Testing

This part of the questionnaire basically looks at the **challenges** mathematics teachers face in carrying out achievement testing. Express your opinion on the subject using the following rating scale: **Strongly Agree = SA; Agree = A; Neutral = N; Disagree = D; and Strongly Disagree = SD;** choose the response that is closer to describing your opinion. (Tick [] only one)

Item No.	Challenges of achievement testing	SA	A	N	D	SD
38	Large class size makes it difficult to assess students through achievement test					
39	Inadequate knowledge on how to construct achievement test hinders testing of students learning outcomes					
40	Inadequate knowledge on how to administer achievement test is a factor for not testing students					
41	Some topics are actually difficult to test students on					
42	Inadequate knowledge on how to score achievement test is a factor for not testing students					

43	Inadequate knowledge on the usage of achievement test results hinders achievement testing					
44	Loaded timetable is a factor for not testing students					
45	Developing mathematics achievement test is a difficult task					
46	Inadequate support from school authority in terms of logistics and facilities affects achievement testing					
47	Inadequate time to complete the mathematics syllabus hinders achievement testing					
48	Inadequate time to gather information and materials to use hinders achievement testing					
49	Inadequate teacher motivation on the part of school authority hinders achievement testing					
50	Some forms of achievement testing are not carried out because of inadequate funds					

51	The school assessment system makes it difficult to use achievement test					
52	The school system does not encourage free and fair achievement testing					
53	Poor supervision affects achievement testing					

Section E: Teachers Knowledge On Test Construction

This part of the Questionnaire looks at how mathematics teachers **construct, administer, and score** achievement tests.

Section A basically looks at how mathematics teachers **construct achievement test**. Express your opinion on the subject using the following rating scale: **Strongly Agree = SA; Agree = A; Neutral = N; Disagree = D;** and **Strongly Disagree = SD;** choose the response that comes closest to describing your opinion. (Tick [√] only one).

Item No.	Construction of achievement test	SA	A	N	D	SD
54	State the purpose of the test					
55	Specify the construct to be measured					
56	Use a test specification table					
57	Construct test items when it is time to assess					
58	Use test items from test books					

59	Set test items from past questions					
60	Questions reflect the construct to be measured					
61	Ask colleagues in the subject area to help you to construct the items					
62	Ask colleagues in the subject area to help you to review the items					
63	Prepare marking scheme after administering the test items to students					
64	Double-barrelled questions are necessary in Achievement testing					
65	Consider the language maturity of the students					
66	Consider the time that will be spent on the test items					
67	Teacher solves the questions to determine the required time					
68	Provide clear instructions on how test items should be answered					
69	Write more test items than the required number					

70	Write items at an early date before the actual date for the test					
71	Follow the principles of test construction for each format used					
72	Sometimes ask students to construct their own questions					
73	Construct test items from the scheme of work					
74	Construct test items during teaching					
75	Consider the weightings of the profile dimensions when constructing test items					
76	Arrange test items in an ascending order of difficulty					
77	Avoid leading questions					
78	Achievement test items are primarily designed by the examiner					

87	Inform students early on the areas for the test					
----	---	--	--	--	--	--

88	Prepare the classroom before students sit for the test					
89	Tests are given after long vacations or after holidays					
90	Proof read all test items					
91	Make provision for emergencies during the time the test is taken					
92	Make sure test items are visible, clear and properly arranged (i.e., well photocopied)					

