

UNIVERSITY OF CAPE COAST

TEACHERS' KNOWLEDGE ON VALIDITY AND RELIABILITY OF  
CLASSROOM ASSESSMENT: EVIDENCE FROM SENIOR HIGH  
SCHOOL TEACHERS IN SEKONDI-TAKORADI METROPOLIS

ABIGAIL ABA TAWIAH MENSAH

2021

UNIVERSITY OF CAPE COAST

TEACHERS' KNOWLEDGE ON VALIDITY AND RELIABILITY OF  
CLASSROOM ASSESSMENT: EVIDENCE FROM SENIOR HIGH  
SCHOOL TEACHERS IN SEKONDI-TAKORADI METROPOLIS

BY

ABIGAIL ABA TAWIAH MENSAH

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

SEPTEMBER 2021

DECLARATION

**Candidate's Declaration**

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

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

Name: .....

**Supervisors' Declaration**

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

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

Name: .....

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

Name: .....

## ABSTRACT

The study focused on investigating knowledge on validity and reliability of classroom assessment among SHS teachers in Sekondi-Takoradi Metropolitan Assembly. This study adopted the descriptive survey design. A sample of 278 was selected for the study through stratified random sampling technique. Questionnaire and observation checklist were designed for collecting data for the present investigation. Data were analysed using frequencies and percentages, means and standard deviations, one-way MANOVA, content analysis, and simultaneous multiple linear regression analysis. It was found that majority of the respondents possessed high knowledge on validity and reliability. It was also revealed that the respondents did not engage in several practices that enhance validity and reliability of assessment, however, they mostly engaged in only a few. The results further showed no statistically significant gender difference in knowledge on validity and knowledge on reliability. With respect to years of teaching, there was a statistically significant difference in knowledge on validity, however, no statistically significant difference in teachers' knowledge on reliability. Finally, it was revealed that both knowledge in validity and reliability were positive predictors of adherence to practices that enhance validity in assessment. It was concluded that the soundness of the interpretations and uses of the end of semester results of the schools in this study for certain purposes are questionable. It was recommended that the Metropolitan Directorate of Education, STMA, continues to intensify their workshops for teachers, particularly in the area of assessment.

## ACKNOWLEDGEMENTS

I am very grateful to my main supervisor, Dr. Kenneth Asamoah-Gyimah, for his contributions. I am also appreciative of efforts made by Prof. Eric Anane, my co-supervisor.

To the following persons: Dr. Regina Mawusi Nugba, Mr. Francis Ankomah, and my dearest husband, Mr. Sampson Okoh Okai, Jacob Fraikue and Mr. Moses Atta Mensah, I thank them for their support and encouragement.

I also thank Rev. Kingsley Bortey, my lovely mum and kids for their prayers. Finally, I acknowledge the Board of the School of Graduate Studies, UCC for granting me the scholarship, which helped me in carrying out this study.

## DEDICATION

To my late dad and sisters.

## TABLE OF CONTENTS

	Page
DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
DEDICATION	v
LIST OF TABLES	x
LIST OF FIGURES	xii
CHAPTER ONE: INTRODUCTION	1
Background to the Study	1
Statement of the Problem	5
Research Questions	8
Hypotheses	8
Significance of the Study	9
Limitations	10
Delimitation	10
Operational Definition	10
CHAPTER TWO: LITERATURE REVIEW	11
Introduction	11
Theoretical Review	12
Validity Theory	12
Classical True Score Theory	20
Conceptual Review	21
Concept of Assessment	21

Teacher-made Assessment	24
Assessment Practices	25
Validity and Reliability of Assessment	31
Sources of Validity Evidences	33
Empirical Review	41
Knowledge in Validity and Reliability in Assessment	41
Adherence to Practices that Enhance Validity and Reliability of Assessment Results	43
Demographic Characteristic and Knowledge in Validity and Reliability	50
Conceptual Framework	51
Chapter Summary	53
<b>CHAPTER THREE: RESEARCH METHODS</b>	<b>55</b>
Sampling Procedures	58
Data Collection Instruments	59
Data Processing and Analysis	62
Chapter Summary	64
<b>CHAPTER FOUR: RESULTS AND DISCUSSION</b>	<b>65</b>
Demographic Characteristics of Respondents	65
Research Question 1	66
Research Question 2	69
Research Question 3	71
Research Question 4	74
Evaluation of Samples Tests	75
Observation of Examination Procedures and Conditions	81



Descriptive Information on Knowledge on Validity, Knowledge on Reliability, and Adherence to Validity and Reliability Practices	83
Hypothesis 1	84
Hypothesis 2	86
Hypothesis 3	89
Discussion	91
Teachers' Knowledge on Validity and Reliability	92
Teachers' Engagement in Practices that Enhance Validity and Reliability of Assessment Results	94
Sources of Test Score Invalidity	96
Demographic Characteristics and Teachers' Knowledge on Validity and Reliability	97
Chapter Summary	99
CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATION	
Summary	100
Conclusion	103
Recommendations	103
Suggestions for Further Studies	104
REFERENCES	106
APPENDICES	116
APPENDIX A: QUESTIONNAIRE	117
APPENDIX B: OBSERVATION CHECKLIST	124
APPENDIX C: ENGLISH LANGUAGE TEST SAMPLES	126

APPENDIX D: MATHEMATICS TEST SAMPLES	144
APPENDIX E: INTEGRATED SCIENCE TEST SAMPLES	159
APPENDIX F: NORMALITY TESTS	176
APPENDIX G: LINEARITY, HOMOSCEDASTICITY, NORMALITY OF RESISUALS	179
APPENDIX H: INTRODUCTORY LETTER	180
APPENDIX I: ETHICAL CLEARANCE	181

## LIST OF TABLES

Table	Page
1 The Distribution of Public SHS Teachers in STMA based on Gender	57
2 Sample Distribution	59
3 Distribution of Respondents based on Demographic Characteristics (N = 255)	66
4 Items on Validity	67
5 Level of Knowledge on Validity	69
6 Items on Reliability	69
7 Level of Knowledge on Reliability	71
8 Practices that Enhance Validity and Reliability	72
9 Room Conditions (n = 14)	81
10 Invigilation (n = 14)	82
11 Testing Conditions (n = 14)	82
12 Descriptive Statistics on Knowledge on Validity, Knowledge on Reliability, and Practices	83
13 Wilks' Lambda Multivariate Test for Gender Differences in Knowledge in Validity and Reliability	85
14 Univariate Test for Gender Differences in Knowledge in Validity and Reliability	85
15 Wilks' Lambda Multivariate Test for Differences in Knowledge in Validity and Reliability with respect to Years of Teaching	87
16 Univariate Test for Differences in Knowledge in Validity and	

	Reliability with respect to Years of Teaching	87
17	Tukey HSD Multiple Comparisons	88
18	Model Summary for Knowledge in Validity and Reliability	90
19	Regression Coefficients for Knowledge in Validity and Reliability	91

## LIST OF FIGURES

Figure		Page
1	Categories of evidence used to argue for the validity of test score interpretations and uses (AERA, APA, & NCME, 2014)	40
2	Flow chart showing the link between teachers' knowledge in validity and reliability, and valid and reliable assessment results	52

## **CHAPTER ONE**

### **INTRODUCTION**

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

The overall purpose of conducting assessment to learners is to ascertain that learners have undergone rigorous and holistic learning processes that are based on evidence through proper documentation, such that learners exhibit absolute competence of knowledge as well as in the ability to reason, perform certain skills and behavioural patterns in their quest to pursue an academic degree (Sadler, 2009). Essentially, assessment plays a critical role in the learner's educational journey since it greatly impacts the learner's study patterns as well as the quality of learning outcomes. According to Sadler, assessment additionally provides learners with a clear idea of their own progress, capabilities and challenges that arise at any point in time as they go through academic exercises. This emphasises the need for validity and reliability in assessment. Sadler further submitted that the validity of an assessment is when learners' grades are in tandem with their accomplishments in academics.

Teachers inevitably have significant roles to play in any assessment as they use diverse teaching strategies including observational techniques during instructional hours, or setting of special tasks and or projects to monitor learners' academic progress, or use self-constructed tests or class work (course work) that are formative in nature (Assessment and Learning Research Synthesis Group [ALRSG], 2004). Apparently, classroom teachers evaluate

learners' academic performance using their test scores. Learners' test scores are usually derived from the summation of all the formative scores of the academic contents they study within a specific period of time. Given that assessment directs the career paths or options of learners, it is imperative for teachers to ensure the validity and relevance of test or assessment instruments (Griswold, 1990).

Concerns regarding safeguarding that assessments are valid continuously remain critical in the further development of education worldwide. As Sadler (2009) concurred, cognisance accorded to valid judgements of individual work pieces should precede a reliable grading system in any academic setting which greatly emphasises diagnosis of learners' difficulties and the improvement of their capabilities. Fives and DiDonato-Barnes (2013) indicated that when clear discrepancies exist between the contents of the learning materials and the assessment or test materials at the end of the assessment, the credibility of the assessment tool is usually disputed since it does not provide a strong basis for the teacher to make absolute and independent judgement regarding the students' academic progress.

The issue of validity in classroom assessment manifests in several ways. First the test has to be planned after which a specification table is used. Developing a specification table is, thus, one of the strategies that educators can use to deal with the issue of validity. One of the most essential factors that the assessor needs to consider in developing a test is ensuring that they have developed an assessment instrument that makes it possible to infer students' performances in a content area, and also to feel confident about the results as a true reflection of the students' ability. This requires the assessor beginning with

a clearly structured, unambiguous and concise statement of the meaning and the purpose of the assessment (Kane, 2006), after the assessor has strictly adhered to the assessment protocols regarding the content that was taught, then the assessor can boldly conclude that the results from such an assessment is valid.

Messick (1989) submitted that validity is the extent to which concept and evidence corroborate the explanations of test scores detailing the projected uses of the test. More so, Messick posited that the procedures of validating an assessment instrument entail the accumulation of evidences that provide scientifically grounded basis for the projected explanations of test scores. Essentially, the interpretation of the scores of a test is the most important requirement for conducting an evaluation but not the actual test. According to Crocker and Algina (2008), the appropriate validation type hinges on the conclusions that would be made from the test scores. Generally, validity evidences are typically broken further into three basic types which are construct validity, criterion validity, and content validity.

In addition to getting evidence on the meanings and uses of test scores, students' results are expected to be reliable, thus consistent overtime. Reliability is the degree to which test scores are reliable and reproducible (Crocker & Algina, 2008). In assessment, teachers are interested in knowing whether or not, students' test results would be similar or be about the same results should the test be taken again and again. In the field of testing, reliability also implies giving an assurance that a student's test result is an absolute representation or reflection of their content knowledge and or real scores obtained provided the measurement procedure had no or a minimum margin of error. However, an absolutely perfect reliability cannot be guaranteed since measurement errors are



inevitable and that testing only provides an estimation of students' actual score as creating an assessment that in our estimation yields a near accurate assessment as possible is the desired goal (Shillingburg, 2016).

Amedahe and Asamoah-Gyimah (2015) posit that achievement tests in the classroom are usually designed by teachers (teacher-made). These tests are designed by teachers with the aim of determining students' learning attainments or progress following the delivery of a course, term or contact hours of an academic period. Teacher-made tests are most of the times used in measuring students' academic attainments in a specific or independent subject area in class setting. This implies that teachers ought to be well vested in knowledge related to practices in assessing their learners. The purpose of teacher testing has been described by measurement experts (Etsey, 2004). All of these authorities agree that the primary purpose of a test administered by teachers is to provide valid, reliable and useful information about student progress and also to help assess educational progress and achievement in order to improve overall classroom teaching and learning to improve .

Amedahe (2014) found that while teachers can achieve some degree of success in their classrooms without following prescribed principles in testing their students, more could be achieved by following scientific principles and practices inherent in testing measurement are considered useful. This situation is very critical in the Ghanaian education system, where a number of studies have shown that teachers have poor assessment practices and do not adhere to recommended assessment principles (Anhwere, 2009; Oduro-Kyireh, 2008; Quansah, Amoako, & Ankomah, 2019) . This, therefore, triggers the need for a

study of this kind to examine whether teachers really understand the issues of validity and reliability which are the bedrock of assessment.

### **Statement of the Problem**

Assessment is an indispensable tool in every educational system. Assessment and teaching are interwoven. Assessment also serves as a rich source of information that is required to evaluate learners. Essentially, feedback and having knowledge of one's own results would be impossible, and not having knowledge of one's own results hinders the smooth progress of learners especially regarding their academic achievement. Amedahe (2014) found that Basic One teachers, even from kindergarten to university in the Ghanaian education system, use some kind of assessment practice to determine whether or not learning has taken place, or sometimes for the selection of the next leader of education. In addition, teachers construct tests to identify student problem areas in specific areas of the subjects covered.

In the Ghanaian education system, standardised achievement tests are not predominantly used to assess students in the classroom; All the information needed to make important teaching decisions is provided by non-standard face-to-face tests. As a result, teachers must always ensure that they follow accepted standards when designing, administering and scoring learner tests in order to increase the validity of the test results obtained. If so, they would endeavour to make their test results more reliable so that the use of the test results would be as reasonable and appropriate as possible. This, in a way, minimizes the negative consequences that students and test users may experience if their test scores are used as intended.

However, my personal observation of constructed test items by a number of teachers, testing conditions in schools, and scripts marking in some schools in Sekondi-Takoradi Metropolis suggests to me that most teachers appear to lack the fundamental principles pertaining to validity and reliability of classroom testing. Amedahe (2014) indicated that teachers are expected to deliver in constructing, administering, scoring and interpreting the results of classroom achievement tests. Amedahe further stated that, while some teachers have received in their college courses, pre-service instruction concerning the construction, administration, scoring of tests and the interpretation of test results, others have not. However, these teachers are required to conduct timely assessments for their students, in order to have information on how much learning has taken place, and the need to vary or modify instruction, whenever necessary. This, therefore, implies that teachers should possess some competencies regarding assessment, in order to carry out this mandate successfully.

Empirical evidence from Postareff, Virtanen, Katajavuori and Lindblom-Ylänne (2012) indicated that, classroom teachers lack the necessary abilities for conducting valid and reliable assessments for their learners. Some investigations also demonstrate that teaching and assessment actions move together (Deneen & Boud, 2014; Postareff, Lindblom-Ylänne, & Nevgi, 2007). Problems associated with assessment practices may possibly reflect in the reliability and validity of the results of the assessment. In such instances, the quality of learning outcomes is not reflected in learners' grades. In the Ghanaian context, a number of studies have equally found poor assessment practices among teachers, and lack of assessment skills, in general (Anhwere, 2009;

Oduro-Kyireh, 2008; Quansah, Amoako, & Ankomah, 2019). The implication of the results of the aforementioned studies is that uses and interpretations to those scores have put to, were limited in terms of their soundness. On the other hand, few other studies have found high level of knowledge in test construction among teachers (Ankomah et al., 2020; Ankomah & Nugba, 2020). It is worthy to mention that the aforementioned primarily focused attention on test construction. It be emphasised that teachers' knowledge in validity goes beyond knowledge in test construction, but other areas of assessment practices.

From the foregoing discussions, one must note that validity and reliability remain the benchmarks of assessment. All the principles and theories in assessment are geared towards making assessment results valid and reliable. When teachers are not knowledgeable in validity and reliability, and when they do not know the validity and reliability implications of the assessments they conduct, they may not see the need to follow certain principles to construct, administer, score, and interpret assessment results. Subsequently, they are more likely to do things anyhow, without recourse to the meaningfulness and uses of assessment results. The quest of the current study is to investigate teachers' knowledge on validity and reliability in classroom assessment, since findings of this study would provide some evidence to the root cause of poor assessment practices among teachers.

## **Purpose of the Study**

The present study focused on investigating the knowledge of Senior High School (SHS) teachers in Sekondi-Takoradi Metropolitan Assembly on validity and reliability of classroom tests. Specifically, the study sought to;

1. Assess teachers' knowledge on validity;
2. Assess teachers' knowledge on reliability;
3. Examine the extent to which teachers engage in practices that enhance (a) validity and (b) reliability of assessment results;
4. Determine gender difference in teachers' knowledge on (a) validity and (b) reliability in assessment;
5. Determine the difference in teachers' knowledge on (a) validity and (b) reliability with respect to years of experience in teaching;
6. Examine the influence of teachers' knowledge on (a) validity and (b) reliability on the extent to which teachers engage in practices that enhance validity of assessment results; and
7. Determine the sources of test scores invalidity.

## **Research Questions**

The study was guided by the following research questions:

1. What is the level of knowledge of teachers on validity?
2. What is the level of knowledge of teachers on reliability?
3. What is the extent to which teachers engage in practices that enhance (a) validity and (b) reliability of assessment results?
4. What are the sources of invalidity in test scores?

## **Hypotheses**

The study tested the following hypotheses:

1. H<sub>0</sub>: There will be no statistically significant gender difference in teachers' knowledge on (a) validity and (b) reliability of assessment.

H<sub>1</sub>: There will be a statistically significant gender difference in teachers' knowledge on (a) validity and (b) reliability of assessment.

2. H<sub>0</sub>: There will be no statistically significant difference in teachers' knowledge on (a) validity and (b) reliability of assessment with respect to years of teaching experience.

H<sub>1</sub>: There will be a statistically significant difference in teachers' knowledge on (a) validity and (b) reliability of assessment with respect to years of teaching experience.

3. H<sub>0</sub>: Teachers' knowledge on (a) validity and (b) reliability of assessment will not significantly predict the extent to which they engage in practices that enhance validity of assessment results.

H<sub>1</sub>: Teachers' knowledge on (a) validity and (b) reliability of assessment will significantly predict the extent to which they engage in practices that enhance validity of assessment results.

### **Significance of the Study**

Findings of this present investigation enormously impact the assessment practices in SHSs in Sekondi-Takoradi Metropolitan Assembly, and the entire country. Information on teachers' knowledge in validity and reliability, which are the bedrock in assessment would give foresight and the way forward in providing remedies for enhancing the assessment practices among teachers. The findings of this study, in a way, provide much information to the Ministry of Education (MoE), Ghana Education Service (GES), headmasters, as well as the

teachers in the selected SHSs used in the study, to help them in addressing the lapses in the education system in terms of classroom assessment.

### **Limitations**

Much caution should be exercised in attempts to interpret the findings of the study, since teachers were not tested in the quest to measure their knowledge on validity and reliability. In addition, the validity and reliability of findings of this study, to some extent, depends largely on accuracy and honesty of information provided by the respondents.

### **Delimitation**

The study was delimited to only public SHSs in STMA. Moreover, this study focused on validity and reliability issues in assessment. In addition, the study examined difference in knowledge on validity and reliability in terms of gender, and teaching experience.

### **Operational Definition**

For the purpose of this study, the following terms were conceptualised as follows:

**Knowledge on validity:** This refers to teachers' understanding on evidence which support the interpretations and use of assessment results.

**Knowledge on reliability:** This refers to teachers' understanding on practices and issues that bother on how test results would be consistent over time.

**Practices that enhance validity and reliability of assessment results:** These are activities that when teachers engage in, would help improve the soundness of the interpretations and how assessment results are used.

## **CHAPTER TWO**

### **LITERATURE REVIEW**

#### **Introduction**

The present study focused on investigating the knowledge of SHS teachers in STMA on validity and reliability of classroom assessment. The focus of this current chapter was to review related literature in the area of validity and reliability of classroom assessment practices. The literature was organised into theoretical review, conceptual review, and empirical review, and conceptual framework.

1. Theoretical Review
  - a. Validity Theory
  - b. Classical True Score Theory
2. Conceptual Review
  - a. Concept of Assessment
  - b. Teacher-made Assessment
  - c. Assessment Practices
  - d. Validity and Reliability of Assessment
  - e. Sources of Validity Evidences
3. Empirical Review
  - a. Knowledge in Validity and Reliability in Assessment
  - b. Adherence to Practices that Enhance Validity and Reliability of Assessment Results



- c. Demographic Characteristic and Knowledge in Validity and Reliability

#### 4. Conceptual Framework

##### **Theoretical Review**

This theory was founded on two measurement theories, namely, the validity theory and the classical true score theory. The theory of validity provides a framework that serves as a benchmark in the practice of assessment and usefulness and meaningfulness of assessment results. In line with the classical true score theory, it provides a formidable foundation of issues relating to reliability of assessment results. In addition, it touches on the practices that can introduce errors to students' assessment results. The two theories underpinned this study.

##### **Validity Theory**

Generally, validity is seen from two theoretical perspectives. The early validity theory and the contemporary validity theory (Stringer, 2008; Kane, 2001). The early validity theory is founded on the measurement instruments and what they purport to measure, with three separate types: content, criterion and construct, with criterion-related validity usually subdivided into concurrent and predictive depending on the timing of the collection of the criterion data. On the other hand, contemporary validity theory is the unitary theory of validity.

In the past, validity was best described as a concept whereby an assessment tool or test truly fulfills the assessment purpose for which it is intended. Validity theory has a much broader scope, however, and views on validity and the framework of validity changed several times during the 1900s and early 2000s. Validity was introduced in the 1920s and initially focused

solely on what later became known as criterion validity. One of the earliest formal descriptions of validity was presented by Cureton (1951), who proposed that validity is associated with a specific purpose for which a test result is used, rather than being a static attribute of a test. Cureton also discussed an aspect of validity he called relevance, and this concept was similar to what later came to be known as construct validity.

The modernised or recently conceptualised meaning of validity states that validity is a single outcome that derives its facts or evidence from diverse sources (AERA, APA, NCME [Standards], 1999; Messick, 1989). The evidence-producing sources are usually proposed by rational anticipations of related meanings or interpretations of measurement outcomes. Construct validity is a recent concept which has received extensive description by Messick (1989), and enshrined in the most recent Standards of Educational and Psychological Measurement (SEPM). The erstwhile definition of construct validity as a form of validity was first published in an academic newsletter by Loevinger in the year 1957 (AERA, APA, NCME, 1999). In the previous years, validity was viewed as encompassing three separate forms which are content, construct and criterion types of validity, where criterion-related validity was further categorised as predictive and concurrent, which is contingent on the purpose and period during which the criterion data is gathered.

According to SEPM (1999), validity refers to “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations” (p. 9). Recent standards are in total support of validity’s unitary conceptualisation after

a critical consideration of Messick's project that embraces the entire validity as construct validity and viewed as a process of investigating clearly determined constructs and effectively gathering data and facts that are used in supporting or opposing the meaning of assessment results in case of any argument or agitation (Kane, 2001). Various methods of validity as well as the forms of evidence that are related to construct validity are based on a number of historical empirical studies by different scholars (Cronbach, 1971, 1988; Cronbach & Meehl, 1955; Messick, 1984).

Cronbach (1980a, b) formulated his thoughts regarding validity in relation to evaluative argument. Moreover, Cronbach has stressed the societal aspect together with valid arguments besides the role of availing structures that help in analysing and presenting data that are valid. The SEPM (1999) recommends that “. . . validation can be viewed as developing a scientifically sound validity argument to support the intended interpretation of test scores and their relevance to the proposed use” (AERA, 1999, p.9). Cronbach (1988) submitted that validity argument necessitates a holistic evaluative procedure that forecast the meaning and how test scores would be put to effective and efficient use. Additionally, it strives for a unified scrutiny of all the evidence that would be used in supporting or opposing the forecasted meaning and other possible unexpected meanings.

As a prerequisite of evaluating the forecasted meanings of test scores, one needs to identify precise and unambiguous statements of the assertions that are in line with the meaning and aims of the uses of the test scores that have been already forecasted. It is apparently true that validating an assessment instrument comes with its own challenges; however, it is highly impossible to

validate an assessment instrument if the meaning or interpretation is ambiguous. According to Kane (1994), and Crooks, Kane and Cohen (1996), the projected meaning could be quantified in relation to interpretive argument which outlines the interconnections linking the deductions derived from the test scores and the decisions that ought to be made subsequently. Essentially, interpretive argument aims to as clearly as possible, assume and infer from the meanings of the test results.

Interpretive argument builds a solid background necessary for advancing a valid argument. It is preferable that we begin with a strong declaration of the projected meaning with respect to a specified unambiguous interpretive argument. Facts and careful scrutiny would be applied in reading and setting expectations in interpretive argument, while strengthening the less developed aspect of the argument.

The all-inclusive idea of validity combines the reflections of content, standard and possible outcomes into a contextual concept to analytically test coherent assumptions or theories regarding the meaning of test scores and how they are used (Messick, 1984). Basically, validating assessment scores or instruments is analytically evaluating the interpretation and possible outcomes of measurement. Thus, validation unites systematic investigation with logical argument to either provide an acceptable explanation or quash the meaning of a test score or what it is used for.

Matters surrounding validity in-terms of interpreting test result, uses, relevance and societal outcomes are multidimensional and interrelated. There is usually the problem of extricating them analytically, hence, the recent observation of the concept as a coherent one (Messick, 1989). For instance,

societal outcomes offer facts that contribute immensely to the connotations of test scores and the uses of both valid facts together with value outcomes. The crux of a coherent validity in most of the time is the suitability, significance and expediency of interpretations of test scores seem impossible to separate as the consolidative supremacy is driven from empirical observation that is rooted in clarification of test scores.

According to Messick (1989), referring to validity as a coherent idea is not necessarily an indication that the concept of validity cannot be meaningfully be segregated into diverse theoretical parts in emphasising matters and minor distinctions that one may perhaps disregard or treat with less seriousness, such as the societal implications of performance after assessment or the significance of the connotation of test scores and its application. The reason accorded to the segregation of the concept is providing ways of delivering illusive features of validity that aid in extricating certain available innate complications at the time of assessing the suitability, significance and expediency of the meaning of test scores.

Messick (as cited in Nitko, 2001) outlines four principles of validation that help in deciding the degree to assessment results are valid:

1. The explanations or connotations that teachers ascribe to their learners' assessment results prove to be valid provided they can have adequate facts that correctly back their claims.
2. The degree of utility that teachers ascribe to their learners' results demonstrates to be valid as long as they can have adequate facts that correctly back their arguments.

3. The connotations and usefulness of ones' assessment results prove to be valid as long as the implications of the test or score values are suitable.
4. The connotations and usefulness one ascribe to results of assessment prove to be valid provided the outcomes of these connotations and usefulness are in congruence with the suitable test or score values.

According to Nitko (2001), such principles must all be in place before a judgment on validity can be made. It follows that the assessment of validity is not based on a single piece of evidence. AERA, APA, NCME (1999), and Messick (1989) categorize validity evidence as content-based, criteria-based, and construct-based. Content validity fact refers to the satisfactory nature of the test's content along with the test responses from which interpretations might be constructed. Evidence of criterion-related validity addresses the empirical method of learning the connection between assessment scores and other measures along with some self-regulating external trials. Construct-related validity evidence is achieved when one is able to make inferences from certain ideas that require the application of psychologically driven concepts related to the test or assessment scores (Amedahe & Asamoah-Gyimah, 2015a).

It is important to note the four features of the recent views of validity. Each one of the views has rich historical contributions to the concept of validity. Firstly, validity encompasses the appraisal of the entire likelihood of the projected explanation or use of assessment or test results. In real sense, we validate the meaning or interpretation of the scores and not the test score itself. Changing from the former radical concept where the intended characteristics of measurement has been focused on the meanings or interpretations has been a

pre-existing one (Cronbach & Meehl, 1955; Messick, 1975), however, it has transformed into a more overt and reliable one.

Secondly, corroborating rising universal ideologies of construct validity, recent explanations of validity admit or side with the idea that the anticipated meanings would include a prolonged scrutiny of the readings and expectations as well as involving reasons for the forecasted meanings and any likely unforeseen meanings. The final judgment from the evaluation process mirrors the competence and suitability the meaning and how well the meaning is sufficiently backed by facts that are apt.

Moreover, as enshrined in Messick's (1989) episode and SEPM (AERA, APA, & NCME, 1999) validation could comprise the assessment of the outcome of the utility of the test as depicted in the following quote:

“Tests are administered in the expectation that some benefit will be realised from the intended use of the scores. A few of the many possible benefits are selection of efficacious treatments for therapy, placement of workers in suitable jobs, prevention of unqualified individuals from entering a profession, or improvement of classroom instructional practices. A fundamental purpose of validation is to indicate whether these specific benefits are likely to be realised” (AERA, MA, & NCME, 1999, p. 16).

Worries regarding the repercussions of validity is manifested in the explanation provided by Cureton (1951) regarding the strength of a test and the purpose for which the test was structured.

Additionally, validity is generally viewed as a universal or coherent assessment of the projected meaning of a test. Validity is merely not an

assortment of systems or apparatuses. The aspirations, tactics employed and principles upon which people make their judgment of validation are constant. The implications encompassing the explanations are to be defined. The implications as well as any required expectation need to be proven with available facts and any other likely meanings also warrant proper scrutiny. The precise expectation of determinants in validation would probably vary across contexts, however, the universal trait or formation of the actual happenings may be constant.

As part of their assignments, teachers construct tests, administer, evaluate and interpret the results. This study borders on the assessment of teacher knowledge regarding the validity and reliability of instructional assessment. In general, there are principles that teachers must follow in order to make meaningful and robust assessments of student learning. It has been empirically proven that when teachers follow these principles, the results of their students' assessments are valid and reliable (Agu et al., 2013). Background and problem statement found that teachers, especially in Ghana, do not follow the recommended principles when designing test items and practice assessment poorly in general. The implication of this is that their assessment results may have reduced validity and reliability. This study seeks to assess teachers' knowledge in validity and reliability in relation to classroom assessment.

Therefore, the prevailing argument is when teachers have inadequate knowledge in validity and reliability, they are more likely to engage in assessment practices that have reduced validity and reliability. In a similar way, when teachers have adequate knowledge in validity and reliability, they are more likely to engage in assessment practices that have high validity and



reliability. Evidence of reduced validity can include unclear definitions, poor sentence structure, ambiguous items, inadequate time limits, difficult tests, discrepancies between learning objectives and test items, a test with few items, incorrect ordering of items, and identifiable patterns of uttered responses, among others (Amedahe & Asamoah-Gyimah, 2015a).

If the above factors play a role, the validity of the test results will be reduced. That is, the interpretations and uses of the test results will be flawed. As noted by Amedahe (2014), most classroom-related tests are teacher-created tests because they are directly tailored to the classroom. The results of these tests reflect student abilities and teaching effectiveness. These results are widely used to guide several educational decisions. When the assessment process is flawed, the resulting decisions are baseless as the means by which such information was collected are problematic. Overall, if teachers do not have sufficient knowledge of the validity and reliability of the assessment, the quality of their students' assessment results suffers greatly from credibility.

### **Classical True Score Theory**

Historically, Classical True Score Theory developed from the early work of Edward Lee Thorndike in his first textbook on test theory in 1904. Almost all discussions of reliability testing begin with what is known as Classical Test Theory (CTT), also known as Classical True Score Theory. CTT is not a (refutable) scientific model, but a statistical model for test results.

From the theory, an observable test score could possibly be seen as a combination of two theoretical components, namely, a true score and a random score. The theory is modelled as:  $X = T + E$ ; where  $X$  represents the observable test score,  $T$  representing the individual's true score and  $E$  representing the error

of random. The observed score is what is seen on the test paper. The true score on the other hand is the expected value of the observed score after conducting the same test repeatedly at different times over a reasonable space of time. Again, the error score is the difference between individuals' observable score and their true score. Reliability is theoretically given as the ratio of the variance of the true and observable scores (Amedahe & Asamoah-Gyimah, 2015). This

is mathematically represented as:  $\rho^2_{xx} = \frac{\sigma^2_T}{\sigma^2_X}$

This implies that reliability tells the extent to which the variance of the observable score is explained or affected by the variance of the true score. For a test to be perfect and reliable, the true and observable score variances should be the same and then the reliability of such a test is +1. Based on this, one may deduce that a decrease in the error score leads to an increase in test reliability. The APA, AERA, and NCMUE (1985) define reliability as a situation where an individual is able to consistently reproduce similar test results after several assessments conducted at different times over a reasonable period of time with an acceptable margin of error. From the definition, it is clear that reliability is inversely related to errors of measurement.

In relation to this study, it can be said that, when teachers are knowledgeable in issues of reliability, they would engage in practices that enhance reliability of assessment results (Amedahe & Asamoah-Gyimah, 2015a).

## **Conceptual Review**

### **Concept of Assessment**

Assessment is an all-encompassing term that is used to describe all the activities a teacher or instructor uses in gathering the evidence regarding the

learners' learning outcomes through a carefully planned and an orderly executed medium, and subsequently make meaning out of their learners' progress in learning. Several scholars propose that the effectiveness in teachers' assessment practices would significantly improve provided they base their assessment on the content of the curriculum and how learning actions are designed (McMillan, 2001; Pilcher, 2001; Shepard, 2001; Stiggings, 2001). There is the contention that assessment methods and instructional strategies should be all-in-one since they both contribute positively to learners' academic progress.

According to Gronlund (2006), the improvement of learners' academic progress and their enthusiasm to learn is the ultimate goal for conducting classroom assessment. Shavelson et al. (2008) also concurred that assessment is a known means that is used to refine teaching and learning outcomes in the classroom setting. In the light of this, Gronlund concurred that classroom assessment that is apt requires that teachers or instructors conceive clearly every learning outcome they aspire to achieve and several assessment strategies that are of high relevance to the instructional procedures and fair across board. Additionally, Gronlund asserts that a comprehensive assessment needs the provision of principles for making decisions regarding efficacious achievement as well as prompt and thorough response to learners while encouraging or reinforcing positive performance outcomes and correcting their shortfalls.

It is a requirement on the part of instructors or teachers to design current and effective assessment procedures that scholars in educational assessment have endorsed to be used in the classroom setting. For instance, Stiggins and Chapuis (2005) submit that academics in assessment endorse that it is important

for learners to be given comprehensive information about their grading system well before hand and also actively participate in the process of assessment.

In recent times, assessment of learners' knowledge in academics has proven to be an inevitable part of the educational framework. According to Oyinka (2007), assessment presents a very difficult condition to teachers or instructors as they are expected to determine how best instructional outcomes have been accomplished, the suitability of the preferred instructional method as well as the capabilities of the teacher to effectively handle the instructional content. Hence, tests are critical to teachers in their everyday professional lives. Denga (2003) also reiterates the importance of test in teachers' professional career.

Compared to other fields such as the physical sciences which have standard instruments with calibrations, the behavioural science instruments require constantly standardising and validating instruments in ensuring their degree of validity and reliability. It is therefore impossible to reach the aspiration of validity in this sense unless the necessary processes in constructing test have been adhered, and subsequently observing the protocols of administration, scoring, and interpretation.

Miller and Parlett (1974) claim that compared to teaching, assessment has significant influence on learners' social behaviour and academic performance. Notably, assessment is arguably one of the dominant features of the formal education setup; hence, any careless and irresponsible approach may mar its positive influence in the teacher-learner relationship and the entire education system. Kellough and Kellough (1999) identified six assessment purposes which include: assisting students' acquisition of knowledge;

identifying students' capabilities and shortfalls; determining and improving teaching efficiency; ascertaining the effectiveness of a desired instructional method; determining and improving the success of curriculum programmes; and providing information useful for decision making.

These purposes indicate that the concept of assessment cannot be downplayed when it comes to teaching and student learning. Assessment practice is the set of steps and processes undertaken by teachers during an assessment. It includes the preparation, management, grading, recording and reporting of assessment information (Malone, 2013). The traditional testing methods used in assessment include multiple choice, matching, true/false, short answer, and essay questions. Alternatively, strategies such as observation techniques, conferences, portfolios, peer reviews and group assessments are equally used in assessing learners.

### **Teacher-made Assessment**

Classroom achievement tests (CATs) are examples of teacher-made assessment (McDaniel, 1994). Teachers usually design such tests to ascertain whether or not learners have attained a certain level of mastery of the content knowledge in a subject or their attainment following the delivery of a course, term or academic calendar (Amedahe, 1989). According to Mehrens and Lehmann (1991), teacher-made tests focus on just one subject area of a grade level.

Teacher-made assessment can be grouped in diverse forms. According to Mehrens and Lehmann (1991), one way of classifying teacher-made assessment is the formatting type used. That is essay-type versus objective-type. Classifying teacher-made assessment can also rely on the material used in

presenting the assessment instrument to learners that is verbally versus non-verbally. Moreover, other categorisations may rely on the intent of the assessment as well as the utility of the test outcome. That is criterion-referenced versus norm-referenced; accomplishment versus performance; and formative versus summative.

The most popular evaluation classification created by teachers among testing professionals is the classification based on the type of item format used, which divides tests into objective tests and essay-type tests (Etsey, 2004; Tamakloe, Atta, & Amedahe, 1996). The above testing experts have asserted that essay tests can be either the extended or the restricted answer types, while objective tests can take the form of short answer, true-false, agreement or multiple choice.

### **Assessment Practices**

Assessment of the value and extent of learning is an important characteristic of classroom learning across all schools. Essentially, setting, administering, scoring and grading examinations are ways by which instructors could employ to appreciate the learning progress or outcomes of their learners. Also, testing affords the instructor the opportunity to identify the capabilities and weaknesses of the learners for whom the test was designed. Tests also help in effectively observing the progress of learners and subsequently compare previous and current performances of learners in the classroom setup (Farrant, 2000). Assessment practices discussed in this study was looked at in terms of test construction, administration, scoring, and interpretation.

### **Test construction**

Adamu, Dawha, and Kamar (2015) assert that in order to construct good assessment instruments, teachers should be able to set out long-term goals, write

short-term aspirations, select the appropriate techniques for assessment, administer assessment and conduct an analysis of the gathered data and discuss the outcome with students or learners. Chidolue (cited in Agu et al., 2013) outlined the following competencies that classroom teachers should possess to construct good assessment instrument item: must be able to determine the intent of every single assessment routine; must be able to state specifically what the quantifiable educational objectives are; must be able to develop exceptional content summary; must be able to prepare assessment schedule which would direct the construction of test items; must have the ability to choose appropriate assessment instrument item arrangements; must have the ability to construct clear, accurate and explicit test items; must have the ability to construct test items that concentrate on gaining students' attention especially those from varied upbringing, about an idea; must be able to construct test items with acceptable error; must have the ability to develop a correct scheme for marking the test or assessment; must be able to perform appropriate analysis regarding their test items; must have the ability to develop assessment instrument that is characterised by monetary and time economy; must be able to give precise instructions regarding the administration of test and how it will be written; must be able to review the instrument with the intent of amending the frame of the test at the time it was constructed.

Koksal (2004) also identified certain characteristics of assessment instruments that are poorly constructed. Those characteristics according to Koksal are:

1. Failure to explicitly specify whom the assessment instrument was intended for, the academic domain that the assessment instrument seeks

to assess, duration of each item, and the probable grade for each appropriate response.

2. Failure to clearly identify or label different parts.
3. Assessment instrument with multiple likely responses as a result of a poor review.
4. Failure to indicate the duration of each of the sections on the assessment instrument.
5. Failure to acknowledge the academic strength of students in terms of grade level while developing the assessment instrument.
6. Ambiguous set of instructions on the assessment instrument.
7. When there is a disagreement between expected performance and actual performance of assessment instructions.
8. Failure to clearly articulate the ideas or concepts that one needs to assess on the assessment instrument.
9. College-production of some assessment instrument items.

When teachers follow the aforementioned principles or procedures when developing test items, the quality test items are assured. Quality test would elicit information that can be relied on for useful decisions.

### **Test administration**

The most important aspect of administering a test is to give each candidate an equal opportunity to show their academic performance against what is being measured (Gronlund, 1988; Tamakloe et al., 1996). The need to maintain consistent test execution conditions cannot be overstated. This is particularly important for the test to provide consistent, reliable, and valid results without much influence from random errors.



The most essential value of administering a test in the classroom setting is to provide every candidate of the examination with the same level of opportunity to showcase their academic attainment as part of the entire learning plan or outcome. This suggests that candidates require a serene and safe environment that could help them in successfully taking the test with little or no interruption as well as a stable psychological state to facilitate the kind of learning outcome they anticipate. All the possible hindrances of a measurement's validity need to be adequately monitored. Given that the physical and emotional characteristics of the performance test can have a significant impact on the candidates' performance outcome, it is important that they are given the best testing or examination condition in order not to be disadvantaged in the examination. Even though the interferences that candidates encounter while writing examinations may not have any effect on their scores, those same hindrances may affect young people who are in the same examination condition (Gronlund & Linn, 1990; Linn & Miller, 2005).

It behoves on administrators and managers of test to master the values and conventions regarding test administration strategies and procedures in order to be certain that these values and conventions are religiously adhered to. When the administrators of test are alien to the conventions and values associated with testing, candidates as well as examiners always end up at the receiving end of the subsequent consequences. The examiner is the one who is solely responsible for a smooth administration of the examination by making sure the appropriate practices are followed. All periods of stages of the test including before, during and after the test should be handled with high degree of competence.

## **Test scoring**

Objective type tests are easy and simple to score since there are almost the same answers and only one best or correct answer. The issue with test scoring basically has to do with the constructed-response type. Response items that mandate candidates to provide verbal or non-verbal responses are also media for assessing candidates' efficiency in English language. Giving scores to multiple types of response items come with a number of setbacks compared to response items that are measured on a categorical scale (for example, yes and no items). It is critical to address certain important issues regarding the design of testing scores and scoring guides way beforehand as such action positively impacts testing practices (Educational Testing Service, 2005).

Numerous approaches are available for scoring assessment responses and assessment performances; hence, one needs to recognise the essence of employing the appropriate scoring approach. A scoring test that is systematic planned requires examiners to ensure the availability of certain unique features that ought or ought not to be in the test responses developed. A comprehensive scoring system makes use of an appropriate scale for scoring and samples for practical training sessions in order to direct all examiners in reaching a unanimous evaluation in general. A comprehensive English Language-ability speaking scale does not necessarily break English speaking into sub-sections. The scale entails vivid descriptions of every ability level and examiners would be in the position of assigning candidates to one of the speaking ability scores as specified by the scale. In assigning students' assessment scores, a scoring rubric should be used and constantly referred to. Rubrics are interpretations of the scores or performance of the candidate which the instructor uses in assigning

performance scores to candidates after performing an academic task (Educational Testing Service, 2007).

It is important to give greater consideration to the intent of the assessment, candidates' academic strengths and weaknesses as well as the difficulty level of the test to be performed when designing scoring scales or rubrics. Essentially, the rubrics of scoring ought to be in congruence with the expected task in order to be certain that examiners are adhering to the appropriate standards of scoring which is free from any scoring biases and confounding materials that could bring the test outcome into disrepute. After tests have been scored, they should be graded with a numerical classification or a letter grade.

### **Interpretation**

In order to ascertain the efficacy of an assessment which is to be conducted for examination candidates or any special group, the individual items and the entire test must be subjected to rigorous statistical analysis. Such statistical analysis should be preceded by the actual testing procedure and must be supervised by competent experts who could interpret and evaluate the results churned out by the statistical analysis for further actions (Educational Testing Service, 2005).

Also, computing the mean scores and the variability of the scores of the entire test performance of candidates is warranted. The analysis should consider both the individual candidates and the groups to which they belong. All these analyses help in identifying whether or not the assessment is of an acceptable level of difficulty to those to whom the assessment is designed, and to compare the performances of the different groups that took part in the assessment.

Analysing the variability of the test scores makes it possible for one to understand how the individual scores spread away or towards the mean or average score which signifies how individual candidates performed in relation to the average test score (Educational Testing Service, 2005).

Moreover, analysing the test items makes it possible for the assessor to have a fair idea of the number of individuals who are at the extreme ends of the score continuum that is those who have very high scores and very low scores when compared with the mean or average score. Such scores will help the assessor or test administrator in knowing what to do during any subsequent test or assessment process.

### **Validity and Reliability of Assessment**

Validity can be viewed as the manner in which factual observations and concepts agree with the explanations of tests as well as the projected use of those tests. This means that validity is a basic prerequisite of developing and appraising tests. In order to validate a test item, one needs to amass appropriate facts to deliver a comprehensive analytic projection of the meaning of the scores. One needs to note that assessors evaluate the meaning of test scores that have been projected and not merely the actual test. According to Messick (1989), validity is a cohesive and evaluative decision of the manner in which experiential facts and hypothetical thoughts agree with the suitability and relevance of actions and interpretations that hinge on test scores or other elements of assessment.

Reliability as a concept is usually used in two differing instances in the measurement domain. Firstly, reliability is conceptualised as the reliability coefficient of classical test concept, which is explained as the interconnection

that exist among scores derived from two separate corresponding tests, suggesting that going through one form of test has no significant bearing on the performance of the other. Secondly, reliability is also employed in a rather broader perspective to convey the meaning that, it is the manner in which test scores are consistent or reproducible after conducting the same test at different times or by different examiners or assessors with an acceptable margin of error.

Three identifiable elements can be said to convey some significant impact on validity and reliability on teacher-made assessments. These variables are the environmental setting, the test or assessment and the testee. The characteristics that are exhibited by the examinee has been identified to have probable effect on the validity and reliability of the test outcome. Cassel (2003) created a statistical measuring technique known as confluence score. It is a testing method that is used to ascertain the consistency (reliability) of an examinee's test score. The confluence score compares two paired items of a test to establish that examinees exhibit some degree of consistency in the way they answer questions. In employing the confluence score, the examiner ought to structure the test such that section of the questions would solicit for the same response in an opposing fashion. This means that consistency should be felt in examinees' responses. Hence, examinees who get different results for the same test are not reliable and should be excluded when validating the test (Cassel, 2003).

Regarding the environment in which the test is conducted, when the environment is unfavourable, the examinee would continuously provide inconsistent test scores so long as the assessment is conducted (Griswold, 1990). Irrespective of the environment and examinees' features, the validity and

reliability of test results are largely influenced by the quality of tests themselves. Another important thing to consider is the duration or quantity of the test items. Tests that are cautiously structured yield higher consistency as they often represent the characteristic being assessed. Three important things that lengthy tests do in maintaining validity are discussed as follow:

Initially, lengthy tests raise the quantity of the content that learners are expected to address, while safeguarding a precise image learners' knowledge. The second point is that, lengthy tests offset the influence of items that are defective through the provision of better and adequate number of tests. The third point to note is that, lengthy tests reduce unnecessary conjectures among examinees (Griswold, 1990). Besides the aforementioned factors, other factors have been identified as things to consider in safeguarding valid and consistent test results. One of such factors is that test questions should be free from ambiguity. Haphazardly constructed questions would compel examinees to guess about the responses, hence, reducing the reliability of the test results. Another point is that, the test should not be too difficult for the examinees to answer or respond (Griswold, 1990).

### **Sources of Validity Evidences**

Corroborating Messick's explanation of validity of been a unitary concept, AERA and NCME through the Standards, did outline five main thematic areas as evidences of validity. These are content, response processes, internal structure, relations to other variables, and consequences. Factual observations gathered from the aforementioned areas help make stronger arguments in favour of or against the propose interpretations or use of students' test scores.

## **Content-related evidence**

Generally, a focus on the content-related validity facts hinges on the content or components of the instruments' medium through which the construct of interest is measured. Within this framework two main things are examined. These are content relevance and representativeness. Basically, the issue of representativeness seeks to achieve evidence on how representative the assessment tasks represent the larger domain, where the domain is the large pool of content areas specified in the syllabus or curriculum. To judge whether results from a particular assessment instrument is valid, evidence must be gathered to show that the assessment tasks were enough in terms of representing the entire pool of content for the particular assessment. In addition to the representativeness of the content, the relevance of the content must be determined. In this regard, the assessment task must match the learning targets specified in the school's syllabus. Assessment tasks should not solicit information on trivial content areas.

Documenting validity facts that are associated with the content upon which the test was conducted is of utmost importance. In fact, a blueprint defines the framework or summary and plan of the test. The blueprint provides detailed and succinct descriptions of all sub-sections and further classifications the test details together with the particular test items specified in each sub-class and the difficulty of the items in terms of cognition.

Test content and its facts related to validity is accumulated after analysing test content and its relationship with the domains that the test purports to assess. Content evidence can come from, for example, expert judgments and systematic scrutiny of the manner in which the content of the test epitomises the

entire field. Assessing the suitability and appropriateness of an existing content area following the recent addition of intent to a test is worthwhile (Wedman, 2017).

### **Response processes**

As a source of validation, the response process may seem a bit odd or inappropriate. Response process is defined here as evidence of data integrity such that all sources of error related to test execution are controlled or eliminated to the greatest extent possible. The response process has to do with aspects of the assessment such as B. ensuring the accuracy of all responses to assessment prompts, the quality control of all data derived from assessments, the suitability of the methods used to combine different types of assessment results into a composite result, and the usefulness and accuracy of the Results reports provided.

For evidence of response process for the written comprehensive examination, documentation of all practice materials and written information about the test and instructions to students is important. Documentation of all quality-control procedures used to ensure the absolute accuracy of test scores is also an important source of evidence: the final key validation after a preliminary scoring – to ensure the accuracy of the scoring key and eliminate from final scoring any poorly performing test items; a rationale for any combining rules, such as the combining into one final composite score of MCQ, multiple true–false and short-essay question scores.

Other sources of evidence may include documentation and the rationale for the type of scores reported, the method chosen to report scores and the explanations and interpretive materials provided to explain fully the score report



and its meaning, together with any materials discussing the proper use and any common misuses of the assessment score data.

A response process that is characterised by facts of validity is required when a construct entails the assumption about one or more cognitive processes, such as reasoning, engaged in by test takers. Evidence often comes from analyses of individual responses but can also come from analyses of different subgroups. Such evidence can also provide information about construct irrelevant sources of variance, in which abilities and other confounding elements affect test performance.

### **Internal structure**

Factual validity that relies on an inner structure provides evidence on how relationships among test components are true reflections or aims of measuring test scores. Dimensionality analyses and analyses of differential item functioning can provide evidence of internal structure, which can then be analysed for conformity to the construct (Wedman, 2017). For a particular assessment results to be judged as highly valid or otherwise, there must be evidence to show relationships among the assessment tasks, and the interconnection among the entire score derived and the assessment tasks.

If an assessment instrument seeks to quantify only one domain such as mensuration or problem-solving, then those items should work together such that each contributes positively to the total score. Also, each item should differentiate among students in terms of geometry problem-solving ability. However, if the assessment seeks to describe more than one behavioural trait, then factual observations are warranted to buttress the argument that the various constructs are distinct. Evidence on internal structure relate to the statistical or

psychometric characteristics of the examination questions or performance prompts, the scale properties – such as reproducibility and generalizability, and the psychometric model used to score and scale the assessment. For instance, scores on test items or sets of items intended to measure the same variable, construct, or content area should be more highly correlated than scores on items intended to measure a different variable, construct, or content area.

Many of the statistical analyses needed to support or refute evidence of the test's internal structure are often carried out as routine quality-control procedures. Analyses such as item analyses – which computes the difficulty (or easiness) of each test question (or performance prompt), the discrimination of each question (a statistical index indicating how well the question separates the high scoring from the low scoring examinees) and a detailed count of the number or proportion of examinees who responded to each option of the test question, are completed. Summary statistics are usually computed, showing the overall difficulty (or easiness) of the total test scale, the average discrimination and the internal consistency reliability of the test.

Reliability as a concept is usually used in two differing instances in the measurement domain. Firstly, reliability is conceptualised as the reliability coefficient of classical test concept, which is explained as the interconnection that exist among scores derived from two separate corresponding tests, suggesting that going through one form of test has no significant bearing on the performance of the other. Secondly, reliability is also employed in a rather broader perspective to convey the meaning that, it is the manner in which test scores are consistent or reproducible after conducting the same test at different times or by different examiners or assessors with an acceptable margin of error.

### **Relationship to other variables (external structure)**

This origin of validity fact encompasses the manner in which assessment results are dependable compared to other assessment results for same students. Evidence in this regard can be seen as predictive or concurrent. In the case of the former, it has to do with providing evidence that students' current assessment results from an instrument are used to predict their performance on similar traits in the future. The latter, however, has to do with evidence on students' current results regarding a measure being quantified from the results of their present achievement using another assessment instrument.

Factual observations regarding validity rely on the interaction of certain elements and the analysis of the interconnection among test scores as one or more elements may likely fall outside the domain of the test. Relationships with external variables can provide either convergent or discriminant evidence, and can take the form of either a predictive or concurrent study (Wedman, 2017).

A typical instance is that, collecting relational validity facts may be of great significance and may deliver a robust positive association with other elements of identical academic attainment or capability and factual observation depicting no relationship or association with certain elements that have been forecasted as measuring entirely different academic achievement or skill.

### **Consequences**

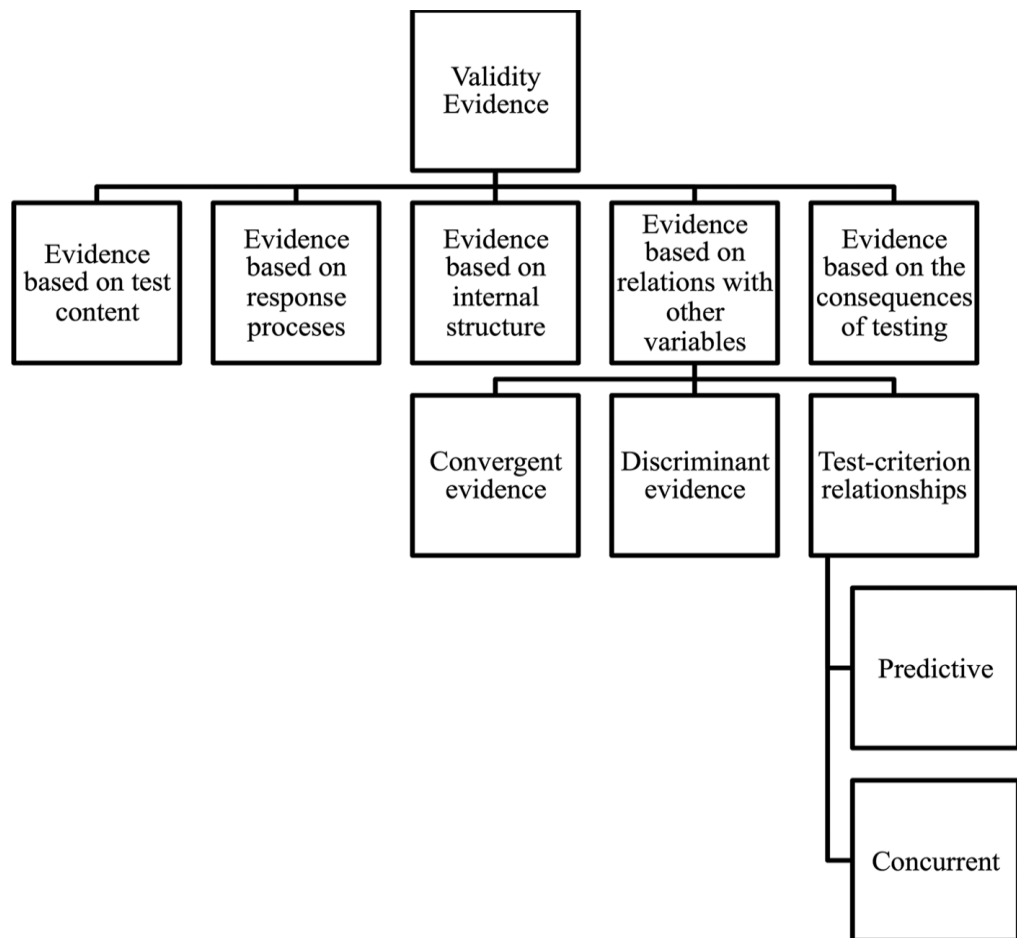
Validity as a consequence is a major feature of construct validity dispute. Factual observations that are related to the repercussions of testing and their subsequent effects is offered to suppose that destructions are not significant features of assessment and as a matter of fact, even if they happen the damage is mild on the assessment. Most of the facts are driven from personal

views than from other origins. Validity as a consequence can be viewed as the bearing of assessment scores, choices and outcomes on examinees' assessment and the entire teaching and learning outcomes. The repercussions of assessment on learners, academics or scholars and the entire society can be enormous as these repercussions or effects may be assumed or unassumed or convey either positive or detrimental outcomes (Downing, 2003).

Factual validity hinges on the repercussion of worries regarding test which are both assumed and unassumed repercussions and may or may not fall in the range of scores projected by the assessor. An unintended consequence of testing could be admitting fewer test takers from a particular group into a certain class or higher education program due to invaluable and irrelevance of the domain's variability source. Analyses of the potential repercussions are also important when a test is used for purposes other than that for which it was developed (Wedman, 2017).

In gathering evidence to validate students' assessment results, it is pertinent to look for evidence on both anticipated and unanticipated repercussions meanings and uses of students' results. All intended consequences of assessment results are those social and educational values associated with students' assessment results. Some intended consequences are improvement in students' learning, improvement in curriculum content and instructional strategies, and improvement in teachers' professional development and school's support for that development. Evidence on how well the aforementioned consequences are achieved contributes to the validation of the assessment results. Thus, if it is expected that students at a particular grade level should be able to read and write upon completion, in order to validate the results

of students who have passed that grade level there should be enough evidence to justify that the student could read and write. The five validity evidence classifications are facts that hinge on: test content; response processes; internal structure; relations with other variables; and repercussions of testing outcomes (AERA, APA, & NCME, 2014). Figure 1 depicts a pictorial illustration of the categories and subcategories of validity evidence.



*Figure 1-* Categories of evidence used to argue for the validity of test score interpretations and uses (AERA, APA, & NCME, 2014)

Reeves and Marbach-Ad (2016) stressed that observable facts from content validity is not enough attaining a validity that is in its highest form; it ought to collaborate with other factual observations to in order to elicit a robust validity that is characterised by sufficient, accurate and reliable facts.

## **Empirical Review**

### **Knowledge in Validity and Reliability in Assessment**

The National Education Association (2003) recounted that all education practitioners are mandated to exhibit competence and possess adequate knowledge about the principles of assessing their learners and also apply those principles in their professional career on a regular basis. A teacher is expected to have adequate and absolute knowledge in assessing their learners as their least trait. Accurate assessment is impossible until education professionals are knowledgeable on validity and reliability implications of assessment results. The Standards for Teacher Competence on Educational Assessment of Students (NCME, 1997) specify that teachers should be knowledgeable in choosing and developing assessment methods; administering, scoring and interpreting assessment results; using assessment results for decision making and grading; communicating assessment results; and recognizing unethical assessment practices.

Pipia (2014) examined the practical strategies teachers employ to enhance the degree of reliability and validity in assessment. This study was conducted among teachers in Georgia, US. Results of the study showed that the teachers always provide accurate information to students before administering the test, they also consistently refer to the scoring rubric when scoring the test papers of their students. It was further found that the teachers do not consider personal and confounding elements while assessing the test results. This implies that the teachers have knowledge in validity and reliability in terms of students' assessment.

Another study by Curriculum Research and Development Division [CRDD] (1999) examined student assessment procedures in Junior High Schools (JHSs) in 11 Ghanaian districts. It was also found that teachers lacked sufficient training regarding how assessment practices are managed. It was reported that approximately a little above half (55%) of the teachers who were interviewed had the conviction that they lacked the appropriate confidence needed to test and measure their learners as a result of not have undergone any form of training or professional development module.

Ankomah, Amedahe, and Cobbinah (2020) assessed test construction skills among senior high school teachers in Sekondi-Takoradi Metropolis. The study employed the descriptive cross-sectional survey design. The study made use of 346 teachers using the census method. Information was gathered from the teachers through the use of questionnaire. The results of the study indicated that the teachers had higher knowledge in test construction.

In a similar study, Ankomah and Nugba (2020) examined the level of test construction skills possessed by senior high school teachers in Cape Coast Metropolis. The design for the study was cross-section survey. Through the cluster sampling technique, 583 teachers were selected from senior high school teachers in Cape Coast Metropolis. It was revealed from the study that majority (46.1%) of the teachers had higher skills in test construction, 33.8% also had moderate skills in test construction, while 20.1% had a lower level of skills in test construction. It can, therefore, be said that respondents, generally, have higher skills in test construction.

## **Adherence to Practices that Enhance Validity and Reliability of Assessment Results**

In a similar study, Oduro-Kyireh (2008) examined the testing practices of Senior Secondary School (SSS) English Language, Core Mathematics and Integrated Science teachers in relation to the construction, administration and assessment of their classroom achievement tests and the interpretation of the results of these Testing. The study aimed to find out whether the procedures used by teachers in constructing, administering and scoring class achievement tests and interpreting the results of these tests conformed to the principles and guidelines prescribed by measurement specialists. A sample of 10 districts was randomly selected from the 21 districts in the Ashanti region. Using the cluster sampling method, 26 SSSs were sampled out of a total of 56 SSSs from the 10 districts. The sample for the study included 265 teachers who taught the three subjects. The questionnaire was designed to collect data from respondents.

The study found that out of 10 principles of test construction, teachers followed seven, including: defining test intent; reference of the short-term teaching goals of the subjects to experiments; Choosing the appropriate test format for testing the stated objectives; Writing the test questions weeks before the test (at least two weeks) to make room for reviews and other eventualities; Preparation of the marking scheme once the test items are written; Checking the test items after putting them aside for a few days; and to write clear and concise instructions for the entire test and parts of it. Principles not commonly used by teachers include: using test specification tables to determine the items included in the test; preparing more items than needed for the test; and evaluation of the test as a whole according to the criteria of clarity, practicality, validity,



efficiency and fairness. It is clear from the results that the trainers lack the necessary skills to create tests. This therefore suggests that teachers' knowledge of validity and reliability is limited in the sense that all principles of test construction rest on ensuring validity and reliability in assessment.

Quansah, Amoako, and Ankomah (2019) examined the test construction skills of SHS teachers in the Cape Coast Metropolitan Assembly (CCMA). Through the use of qualitative document analysis, samples of final exam papers between 2015 and 2018 in Integrated Science, Nuclear Mathematics, and Social Studies were used in three randomly (lottery-scheduled) identified SHS in CCMA. Specifically, five sample questionnaires were selected based on each subject from each school. A total of 15 samples of examination papers were taken from each of the three schools. A total of 45 samples of examination papers from the three schools were sampled. Researchers in the field of educational measurement and evaluation have critically questioned the model instruments. The results of the study indicated that the teachers were not sufficiently qualified in the preparation of examination questions at the end of the semester. It was evident that there were issues with the relevance of the rating, consistency and unfairness in how the rating was rated.

In another study, Marmah and Impraim (2013) examined the competence of university lecturers in creating multiple-choice test questions. 63 lecturers from five departments of the University of Education, Winneba-Kumasi Campus (UEW-K) served as a sample. Questionnaires were introduced to collect relevant data from respondents. Data were also collected through a content analysis of selected end-of-semester exams with multiple-choice items constructed by the lecturers for the 2010-2012 end-of-semester exams. The

analysis was based on (1) appropriate use of multiple-choice items, (2) format of the test, (3) quality of the items, and (4) appropriateness of the options. The results showed that most of the time the teachers did not follow the principles established by measurement experts when developing the multiple choice item. For example, as part of the task planning, the instructors did not create a table of test specifications and did not emphasize or capitalize negative phrases or phrases in the text.

Ing, Musah, Al-Hudawi, Tahir, and Kamil (2015) examined the content validity of teacher-conducted assessment in three Chinese elementary schools in Johor, Malaysia. The study by Ing et al. further examines teachers' understanding of the specification table in the selected schools. A 10-item questionnaire was administered to 30 teachers as a means of collecting data on the specification table. Responses about teachers' knowledge of the specification table were collected with items 1-4, while items 5-10 asked for information on the validity of teacher-generated tests. The results of the study indicated that examiners or teachers showed little knowledge of the specification table.

Moreover, a careful scrutiny revealed that a greater proportion of the teachers never got involved in any organised course to upgrade their knowledge regarding specification tables, hence, they lacked the competence to develop the appropriate specification table for their respective subject areas of teaching. Finding of the study further revealed that teacher-made assessments conformed to content validity. Moreover, majority of the teachers never consulted the specification table when developing their assessment instruments, indicating

that the lack of awareness and knowledge on the teachers' part regarding the essence of specification table and how they are used.

Cooper, Pittman, and Womack (2016) examined teachers' understanding of validity and reliability in assessment. The study sampled 40 multiple-binary choice items administered to 258 students in the Business Essentials classrooms at Douglas County High School. The data gathered from the student responses were analysed with MS Excel. The criterion that was employed in analysing the data include KR-20, difficulty, and discrimination indices. Results from the study showed that qualities of the test items were good, a few items were determined to be potentially problematic. The assessment in alignment with classroom curriculum and criterion validity was established by aligning the assessment with state standards. Based on this, Cooper et al. indicated that the teachers have adequate knowledge in validity and reliability in classroom assessment, hence the quality of their test items.

Anhwere (2009) examined the management of assessment practices in terms of fundamental principles in the construction, administration and marking of classroom or teacher tests with particular reference to the marking of essay test items at the teacher training colleges in Ghana. A descriptive sample survey was carried out at 20 selected teacher training colleges with 310 respondents, 230 of whom were men and 80 women. A questionnaire was used to collect data from respondents. The study found that respondents failed to follow basic testing principles in the creation of placement tests and therefore viewed the administration of college assessment practices as an added burden to their profession. It was concluded from the study that teacher training college tutors

as a whole had limited skills and competences in design knowledge, conducting classroom/teacher tests and scoring essay-type tests.

Koskal (2004) similarly studied teachers testing English Teaching Skills (ELT) skills among English teachers in Turkey. The data were collected during the academic years 2001-2002 and 2002-2003. Fifty-six different student assessment examples (achievement test, formative test or quiz and progress test) were retrieved from schools where teachers completing the ELT academic program gained their practical experience. Evaluation of those sampled tests revealed, among others, the following flaws:

1. The tests failed to identify or state the target audience, the domain the test purported to assess, duration of the test and grade points that were allocated to each item.
2. The test items were characterised with ambiguous responses and having numerous possible responses as they were not stated in context. Illustratively, the examinees were instructed to choose from one of the following:

*“I decided/have decided to move to a better job.*

*I worked/ have worked in an office in Japan.”*

In the two questions, both options are correct since the test items were not stated in context.

3. The duration for each task was not specified on the assessment instrument and majority of the instructions were ambiguous and imprecise.
4. Most items were engulfed with grammatical and typographical errors. Some items were also not based on the learning objectives.

5. Some items failed to serve the purpose for which they were intended.

For instance, the items assessed speaking skills instead of writing skills, as they were instructed to fill the gaps of conversational sentences.

Based on the findings, Koskal (2004) concluded that the teachers had limited skills in test construction and therefore recommended that teachers be trained in testing and assessment. In their study, Hamman-Tukur and Kamis (2000), examined three categories of students' (level 200, 300 and 400) examination questions in University of Maiduguri designed for biochemistry students. The finding indicated that most of the examination questions assessed simple learning outcomes of knowledge and comprehension categories of the cognitive domain at the expense of learning outcomes that call for synthesis and evaluation. Based on the finding, the study concluded that teachers have limited test skills. The study recommended that teachers be sensitised about the essence of designing questions that assess comprehensive learning outcomes of learners.

Boothroyd, McMorris, and Pruzek (1992) studied 41 seventh and eighth grade science and mathematics teachers. These teachers represented 25 public and private districts from many geographic regions in New York. They examined a sample of approximately 350 multiple-choice and completion items submitted by the teachers. They found that teachers' knowledge of validity and reliability in measurement was not adequate. They attributed this deficiency to inadequate training in measurement given at the pre-service teacher education level.

Räisänen, Tuononen, Postareff, Hailikari and Virtanen (2016) examined teachers' experiences of the validity and reliability of students' assessment results at a university in Finland. Using an interview guide, teachers were asked

why they asked the specific exam questions and the type of knowledge and understanding required in the exam. They were also asked to recall how they rated the students' learning outcomes in each test answer. Teachers were also asked what kind of evaluation criteria they used for different grade levels and why they gave certain points for certain answers. Teachers' responses to their practices indicated that grades did not always reflect students' learning outcomes. There were also problems with the alignment of the evaluation and the clarity of the evaluation criteria. For these and many other reasons, the assessment was considered unfair and biased.

In Kenya, Kinyua and Okunya (2014) examined the factors that influence the manner in which teacher-made tests are valid and consistent in the Nyahururu District of Laikipia County. Kinyua and Okunya surveyed 57 participants, where specifically, 42 and 15 teachers and key informants respectively were selected from occupants of various academic positions and performing varied responsibilities in their schools in the Nyahururu District. The study adopted a mixed descriptive survey research design. Questionnaires were used to gather data from the teachers, while interview guide was used to get information from the key informants. The study unravelled that the experience of teachers, training on test construction and analysis, level of education, use of Bloom's taxonomy, moderation of tests and length of tests have an effect on validity and reliability of the tests. The aforementioned elements were identified to have deferring consequence on validity and consistency of assessment tests that are constructed by classroom teachers. Experienced teachers who may have been given some level of practical education or knowledge and apply this knowledge when designing test exhibit

competence and confidence as against their colleagues who have no such knowledge or skills. Subsequently, a conclusion that teacher-made tests have higher validity and consistency was made.

### **Demographic Characteristic and Knowledge in Validity and Reliability**

Kinyua and Okunya (2014) found that the experience of teachers, level of education, use of Bloom's taxonomy, moderation of tests and length of tests have an effect on validity and reliability of the tests. Experienced teachers who may have been given some level of practical education or knowledge and apply this knowledge when designing test exhibit competence and confidence as against their colleagues who have no such knowledge or skills. Subsequently, a conclusion that teacher-made tests have higher validity and consistency was made.

Zhang and Burry-Stock (2003) examined teachers' appraisal practices across instructional levels and content domains, and teachers' self-perceived appraisal abilities as a function of classroom experience among teachers in two school districts in a southeastern state. Data from 297 teachers from the Assessment Practices Inventory were analyzed in a MANOVA design. It has been found that as the length of instruction increases, teachers rely more on objective testing for classroom assessment and become more concerned about the quality of assessment. Alkharusi (2009) found that student teacher assessment knowledge tends to vary by gender. In particular, in a survey of 211 prospective teachers, Alkharusi found that, on average, men tend to have a higher level of knowledge in educational assessment than women. Also, Yan (2016) examined the impact of key demographic variables of gender, school level, and goal orientation on students' self-assessment practices, including self-

directed feedback seeking (SDFS) and self-reflection (SR). A total of 8,843 Hong Kong students were surveyed, ranging from 4th to 6th secondary school. Results indicated that female students demonstrated higher levels of self-assessment practices than male students, including SDFS and SR.

In another study, Asamoah, Songnalle, Sundeme, and Derkye (2019) examined gender differences in formative assessment knowledge of senior high school teachers in the Upper West Region of Ghana. The descriptive survey design was used in the study. A simple random sample was used to select a sample of 295 high school teachers from a population of 1139 teachers. The questionnaire was used to collect data for the study and the data collected were analyzed using an independent t-test. The results of the study showed a significant difference in the formative assessment knowledge of male and female high school teachers and that male high school teachers perform better in their formative assessment knowledge than their female colleagues.

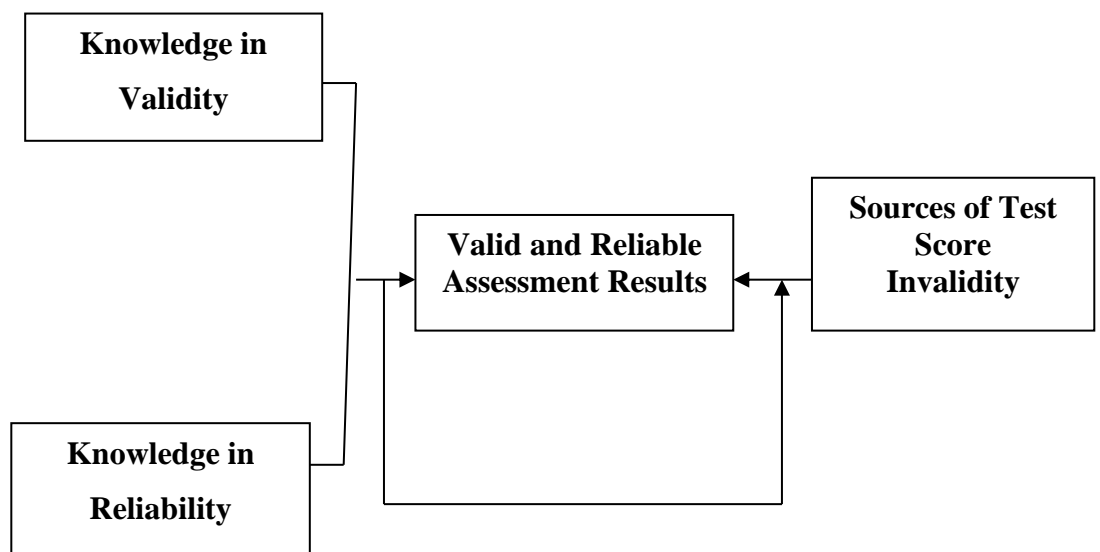
Alkharusi (2011) used a cross-sectional descriptive survey research design in his study to compare male and female teachers in training and active duty in terms of their self-perceived assessment skills. The convenience sampling method was used to select 330 prospective and part-time teachers. To collect data from the participants, a survey was developed in the form of a questionnaire. The results of the study showed significant gender-specific differences with regard to test design and administration as well as the communication of assessment results. Compared to women, men had higher skills in test construction and administration, and in communicating assessment results. Summarizing from the above empirical studies, apart from Pipia (2014); (Ankomah et al., 2020); (Ankomah & Nugba, 2020), the other studies found



that teachers had limited skills in assessment and that their practices also fell short of the recommended guidelines for conducting assessment. These indicate a lack of teachers' knowledge of validity and reliability, which are part of the basis of assessment. In addition, the validity of the evaluation results of these teachers is questionable in terms of their validity and usefulness.

### **Conceptual Framework**

Based on the literature reviewed, a flow chart has been developed to explain the possible interaction that exists between teachers' knowledge in validity and reliability regarding classroom assessment practices (Figure 2).



*Figure 2-* Flow chart showing the link between teachers' knowledge in validity and reliability, and valid and reliable assessment results

From Figure 2, teachers having adequate knowledge in validity and reliability of classroom assessment are more likely to practice the appropriate assessment practices which would elicit meaningful interpretations and

usefulness of students' assessment results. The diagram, as shown in Figure 2, shows that teachers' knowledge in validity and reliability in classroom assessment practices would lead them into engaging in recommended assessment practices. Adherence to the recommended assessment practices such as test construction, administration, scoring, and interpretation result in the achievement of students' assessment results that are both valid and reliable.

It is also conceptualised that, certain practices such as crafting test items with unclear definitions, poor sentence structure, ambiguous items, inadequate time limits, difficult test, mismatch between learning objectives and test items, a test with few items, improper arrangement of items, and identifiable patterns of answers, poor invigilation, poor examination conditions, and hodgepodge scoring of tests, among others when prevalent, may hinder the validity and reliability of students' assessment results. However, with teachers' adequate knowledge in validity and reliability, the impact of the aforementioned factors is buffered.

### **Chapter Summary**

This chapter was devoted towards the review of related literature. The chapter reviewed literature relative to validity and classical true score theory, the theories that form the bases of validity and reliability, respectively. The concepts of assessment, teacher-made assessment, assessment practices, validity and reliability of assessment, sources of validity evidences were reviewed. The review also contained empirical review of studies in this regard. Further, a flow chart was developed as a conceptual framework to explain the link between teachers' knowledge in validity and reliability, and the credibility of their students' assessment results.



## **CHAPTER THREE**

### **RESEARCH METHODS**

The present study focused on investigating the knowledge of SHS teachers in STMA in validity and reliability in classroom assessment. The current chapter presents the research methods that were used in carrying out the investigation. Basically, this encompassed the research design, population, procedures of sampling, instruments for collecting data, and data processing and analysis.

#### **Research Design**

The research design provides the framework on which the entire research activity is executed. The research design clearly explains the actual plans that are precise and explainable which the principal investigator wants to employ in conducting the study. Amedahe and Asamoah-Gyimah (2015) defined research design as the overall set of intended actions that help the researcher in finding answers or solutions to the identified research questions or testing the hypotheses that guide the study. Essentially, the research design serves as the researcher's blueprint that directs the entire research activity since it entails vivid information that are easily explainable. There are different types of research design, however, the choice of a particular design mainly depends upon the uniqueness of the research problem identified, research questions and/or hypotheses involved, and the study group among, others.

The various research designs have their roots in the philosophies which underpin them. These philosophies are the bedrock of the various research

paradigms. Basically, there are three research paradigms, namely: qualitative, quantitative, and mixed methods (Creswell, 2014). This study adopted the descriptive survey design. It was utilised in this study because it aimed to understand the significance of a phenomena and describe its occurrence in a population (Dubin, 1978; Malhotra, 1998; Wacker, 1998). Descriptive survey design allows the researcher to collect both numerical (quantitative) and non-numerical (qualitative) data to provide you with relevant and accurate information. Since this study aimed at investigating the knowledge of Senior High School (SHS) teachers in Sekondi-Takoradi Metropolitan Assembly on validity and reliability of classroom tests, a descriptive survey design was used.

### **Study Area**

The present study was carried out in STMA. Sekondi-Takoradi is a twin city that is positioned at the south-eastern area of the Western region of Ghana. The metropolis is bounded by Ahanta West and Shama Districts to the west and east respectively. More so, the Atlantic Ocean and the Wassa East District bound the metropolis to the south and north respectively. STMA covers a total land area of 191.7 km<sup>2</sup> and serves as the Western region's administrative capital. Although the metropolis has the least land area in the region, it is the most advanced area regarding urbanisation compared to the other 21 remaining districts. The 2010 Population and Housing Census reported that the total population of STMA stood at 559,548, forming approximately 24% of the region's entire population. Of the total population, about 49% are males and the remaining 51% are females.

## Population

The study's population entails all public SHS teachers in STMA. Statistics from the Metropolitan Directorate of Education, Sekondi-Takoradi indicate that, there are 11 public SHS in STMA. Essentially, 981 public SHS teachers who teach in STMA were used as the population for the study. Out of the 981 teachers, 575 are males and the remaining 406 are females. Table 1 outlines the details of the population composition of teachers in STMA.

Table 1- *The Distribution of Public SHS Teachers in STMA based on Gender*

	School	Teaching Staff		
		Male	Female	Total
1	Methodist Senior High School	58	42	100
2	St. John's School	65	25	90
3	Ahantaman Girls Senior High School	50	38	88
4	Sekondi College	65	39	104
5	Diabene Senior High School	33	42	75
6	Fijai Senior High School	72	42	116
7	Ghana Secondary Technical School	53	13	66
8	Takoradi Senior High School	55	36	91
9	Archbishop Porter's Girls Secondary School	53	40	93
10	Bompeh Secondary Technical School	23	49	72
11	Adiembrah Senior High school	48	40	88
	Total	575	406	981

Source: GES, Sekondi-Takoradi Metropolitan Directorate (2019)

## **Sampling Procedures**

A sample of 278 was used for the study. The decision of using 278 was arrived at, having considered the guidelines by Krejcie and Morgan (1970). From Krejcie and Morgan's formula for the estimation of sample size, a population of 981 should have a minimum representative sample of 278. Based on the aforementioned, the sample of 278 by Krejcie and Morgan was used as the sample for the study. It is worthy to note that the use of larger samples in survey research cannot be overemphasised. Generally, when sample sizes are large, they get closer to the population, and for that matter, sampling error is reduced. In addition, larger samples produce more valid results compared to smaller sample sizes. Statistically, larger samples increase the power of the test, and therefore enhance the inferences that can be made from the sample to the population. I do acknowledge that the use of larger samples comes with its associated cost and difficulty in terms of data collection. In this study, however, these challenges were dealt with, and did not hinder the conduct of the study.

The stratified sampling method was, therefore, employed to enrol respondents for the study. The stratification variables were schools and type of school based on sex composition. In this study, therefore, the population was first stratified into three strata, namely, boys' SHSs; girls' SHSs; and mixed SHSs. In this case, the stratification variable was the type of school. Two of the schools were boys' and girls' SHSs each and 7 were mixed SHSs. Specifically, the disproportionate stratified sampling was used. In view of that, the two boys' and girls' SHSs (single sex SHSs) were selected. With the mixed SHSs, simple random (lottery) was used to select three schools out of the seven. Disproportionate stratified sampling was used because each of the categories of

schools had different numerical strength. Stratified proportionate sampling procedure was further used to select sample based on gender, and with this the simple random method was applied to derive a final sample in conducting the study. I used the stratified sampling in order to be certain that the sample was fairly represented in terms of the school type and gender. Table 2 shows the distribution of the sample.

Table 2- *Sample Distribution*

Selected Schools	Population	Sample		
		M	F	T
Ghana Secondary Technical School	66	24	6	30
St. John's School	90	30	11	41
Ahantaman Girls Senior High School	88	23	17	40
Archbishop Porter's Girls Secondary School	93	24	18	42
Bompeh Secondary Technical School	72	11	22	33
Adiembra Senior High school	88	22	18	40
Fijai Senior High School	114	33	19	52
Total	611	167	111	278

Source: Field survey (2018)

From Table 2, out the sample of 278, 167 were males, while 111 were females.

### **Data Collection Instruments**

Questionnaire and observation checklist were designed for collecting data for the present investigation. I designed the questionnaire taking into consideration, the research questions and hypotheses. The questionnaire was made up of 74 items which were organised into four parts labelled 'A' to 'D'.



The first section, Section 'A' which comprised three items was used in soliciting information on respondents' demographic data. Section 'B' was made up of 20 items which solicited information on knowledge on validity. These items had dichotomous responses. Section 'C' was also made up of 20 items which bordered on knowledge on reliability. The dichotomous response format was used. Finally, Section 'D' was made up of 31 items measured on a 4-point Likert scale, with responses from never to always.

This study's observation checklist was made up of 18 items, which have to be checked during the observations. These items were organised under three main conditions pertaining to test administration: room conditions, invigilation, and testing conditions.

### **Pilot testing**

I pilot-tested the questionnaire in the Cape Coast Metropolis using 150 SHS teachers. Suggestions and feedback derived from the pilot test were used to refine the instrument for the final data collection. Some items were finally reframed based on feedback received from the respondents.

### **Validity and reliability**

I ensured that the instrument was valid and reliable before using it for the actual data collection phase. First, the questionnaire was given to my supervisors to vet, this ensured content-related validity evidence. In addition, the questionnaire was pilot tested using 150 SHS teachers in Cape Coast. Subsequently, I ascertained the reliability coefficients of the sub-sections with Cronbach's alpha reliability coefficient ( $\alpha$ ). A reliability coefficient of .70 or more, as indicated by Fraenkel, Wallen, and Hyun (2012), was deemed appropriate. Above all, throughout the conduct of the study, efforts were made

to eschew other practices which could hinder the validity and reliability of the results. The internal reliability estimates were obtained using KR-20 for knowledge on validity and reliability. Cronbach's alpha was used for practices that enhance validity. The reliability coefficient for knowledge on validity and reliability were .72 and .75. For the practices that enhance validity, a coefficient of .77 was achieved. These coefficients were estimated based on the final sample used for the study.

### **Data Collection Procedures**

I requested a letter of introduction from the head of the Department of Education and Psychology to introduce myself as a student assistant (Appendix H). Necessary ethical clearances and approvals were obtained from the appropriate authorities as required throughout the study (Appendix I). The letter of introduction was sent to the Metropolitan Directorate of Education, STMA, for permission to conduct the study. After approval was given, the specific dates and times for data collection were agreed with the school management. On the agreed dates, I and three other trained research assistants visited the various schools for data collection. All questionnaires were distributed personally and by hand to the selected teachers. After that, those who got a reply the same day were retrieved.

The entire data collection was done within three months. Two months were used for the data that were collected with the questionnaire, whereas, one month was used for the observation. The observations were done during the end of semester examinations.

### **Ethical consideration**

This study was strictly guided by the ethical codes that bind the conduct of research on humans. First, I received ethical approval from the University of Cape Coast (UCC) Institutional Review Board (IRB). In the area of data collection, ethical principles such as informed consent, voluntariness, anonymity, confidentiality, privacy and the right of withdrawal were observed. The data collected was treated confidentially, the names of the respondents or their school were not disclosed in any part of the work, rather pseudonyms were used. Data analysis was performed collectively as a group and therefore it was not possible to trace responses back to respondents. The collected data has been well managed and kept secret to avoid access by other people.

### **Data Processing and Analysis**

Quantitative and qualitative means of analysing data were used after collecting data. I used descriptive and inferential statistics in analysing the data. Frequencies and percentages were used in analysing data collected on Research Questions 1 and 2. The responses of the respondents were scored as '1' and '0' for 'right' and 'wrong', respectively. Total scores were then generated by summing the number right answers. Generally, the scores of the respondents ranged from 0 – 20. Scores between 0 and 10 were classified as low level of knowledge, while scores between 11 and 20 were also classified as high level of knowledge.

Moreover, means and standard deviations were used in analysing data collected on Research Question 3. The responses were scored as follows: 1 – Never; 2 – Often; 3 – Very often; 4 – Always. However, eleven of the items were negative items, and therefore they were reversed scored (items 3, 4, 6, 8,

12, 16, 19, 20, 26, 27, and 28). Mean scores were estimated for each of the practices. The mean scores range from 1.0 to 4.0. for the purpose of interpretation, mean scores from 3.0 to 4.0 were judged as mostly followed, whereas mean scores below 3.0 were judged as not followed. This was so because, ideally, teachers are expected to always engage in each of the practices where necessary.

However, I used content analysis in analysing data collected on Research Question 4. To answer this research question, samples of tests developed by the teachers were analysed. The test samples were in the following core subjects: English language, mathematics, and integrated science. The teachers were also interrogated in some instances. These tests were evaluated based on the following dimensions: content representativeness and relevance, thinking processes and skills represented, reliability and objectivity, fairness to different students, and practicality. In addition, the observation data were analysed using frequencies and percentages.

Data collected on Hypothesis 1 were tested using one-way multivariate analysis of variance (MANOVA). The independent variable is gender, which has two levels: male and female. The dependent variables are the mean scores in knowledge in validity and knowledge in reliability. Data collected on Hypothesis 2 were tested using one-way MANOVA. The years of teaching experience were categorised into five levels: less than 1 year, 1 – 4 years, 5 – 8 years, 9 – 12 years, and 12 years and above. Hypothesis 3 was tested using simultaneous multiple linear regression analysis. All the inferential analysis was done at 95% confidence level.

## **Chapter Summary**

This study adopted the embedded mixed method design. A sample of 278 was used for the study. The stratified sampling method was, therefore, employed to enrol respondents for the study. The stratification variables were schools and type of school based on sex composition. Questionnaire and observation checklist were designed for collecting data for the present investigation. Frequencies and percentages were used in analysing data collected on Research Questions 1 and 2. Moreover, means and standard deviations were used in analysing data collected on Research Question 3. Content analysis and frequencies and percentages were used in analysing data collected on Research Question 4. Data collected on Hypotheses 1 and 2 were tested using one-way MANOVA. Hypothesis 3 was tested using simultaneous multiple linear regression analysis.

## **CHAPTER FOUR**

### **RESULTS AND DISCUSSION**

The study focused on examining knowledge on validity and reliability of classroom assessment among Senior High School (SHS) teachers in Sekondi-Takoradi Metropolitan Assembly (STMA). The previous chapter dealt with the methodology employed to conduct this study. The current chapter presents the results of the data collected from the field. The chapter in addition, discusses the results of the study. Out of the 278 questionnaire administered, 255 were deemed valid and therefore used for the analysis. This translates to a 91.7% return rate. The results are presented in two parts. The demographic information of the respondents is presented in the first part, whereas the main results are presented in the second part.

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

This part presents the demographic characteristics of the respondents. The demographic characteristics include gender, professional background, and years of teaching. Details of the results are presented in Table 3.

As shown in Table 3, there were more males (58.4%) than females (41.6%). On the professional background, the majority of the respondents had degree with education (64.7%), only 2.4% and 17.6% possessed HND/Diploma and degree without education, respectively. Nearly 35% of the respondents indicated they had taught for 1 – 4 years, 30.2% had also taught for more than 12 years, whereas, 6.7% of the respondents appeared to have taught for less than a year. Generally, a vast majority of the respondents had taught for more than 4 years.

Table 3- *Distribution of Respondents based on Demographic Characteristics (N = 255)*

Variable	Frequency	Percentage (%)
<b>Gender</b>		
Male	149	58.4
Female	106	41.6
<b>Professional background</b>		
HND/Diploma with education	39	15.3
HND/Diploma without education	6	2.4
Degree with education	165	64.7
Degree without education	45	17.6
<b>Years of teaching</b>		
< 1 year	17	6.7
1 – 4 years	88	34.5
5 – 8 years	33	12.9
9 – 12 years	40	15.7
> 12 years	77	30.2

Source: Field survey (2020)

### **Research Question 1**

*What is the level of knowledge of teachers on validity?*

This research question sought to determine teachers' level of knowledge on validity of assessment. The respondents were asked to respond to 20 items eliciting factual knowledge on validity of assessment. The responses of the respondents were scored as '1' and '0' for 'right' and 'wrong', respectively. The responses of the respondents to each of the items are presented in Table 4.

Table 4- *Items on Validity*

No.	Items	Wrong		Correct	
		f	%	f	%
1.	When directions provided on tests are unclear, it may affect how students respond to the items.	14	5.5	241	94.5
2.	Test items should contain big vocabularies that can make students think for the meaning of the item.	78	30.6	177	69.4
3.	Students should be given a very short time within which they will complete a test or assessment.	40	15.7	215	84.3
4.	Test items developed to assess students should be very difficult, so that the teacher will know the students' ability.	79	31.0	176	69.0
5.	Results from a test with unclear instructions can be effectively used to make meaningful decisions.	85	33.3	170	66.7
6.	All learning objectives can be measured using a single type of test.	142	55.7	113	44.3
7.	Test items should always be based on the learning objectives.	54	21.2	201	78.8
8.	A test should contain very few items in order to measure the amount of students' learning.	122	47.8	133	52.2
9.	Items on a test should be arranged in a way that the difficult ones come first.	86	33.7	169	66.3
10.	Items on test should be arranged in a way that the correct answers form a systematic pattern for easy scoring or marking.	80	31.4	175	68.6
11.	A short test has limited uses and interpretations.	106	41.6	149	58.4
12.	Test items should closely examine what has been taught.	30	11.2	225	88.2
13.	The lighting and ventilation of the testing room can have a significant impact on the performance of students.	72	28.2	183	71.8
14.	When students cheat in an examination, their test result loses its trustworthiness.	34	13.3	221	86.7
15.	It is appropriate to conduct examinations when students are emotionally stable.	120	47.1	135	52.9
16.	Tests like essays can be scored effectively and consistently without necessarily following the marking scheme.	115	45.1	140	54.9
17.	It is necessary to offer some little assistance to students with low ability during examination.	58	22.7	197	77.3
18.	In scoring tests, weaker students can be given some scores to uplift their scores.	80	31.4	175	68.6



19.	It is important to score tests using personal knowledge of each student's past performance.	106	41.6	149	58.4
20.	Assessment procedures should be free of gender, ethnic, social class, and religious bias and stereotypes.	36	14.1	219	85.9

Source: Field survey (2020)

From the results in Table 4, an overwhelming majority of the respondents were able to correctly answer 14 out of the 20 items (thus, items 1, 2, 3, 4, 5, 7, 9, 10, 12, 13, 14, 17, 18, and 20), with percentage above 60%. However, 5 of the items (8, 11, 15, 16, and 19) were wrongly scored by most of the respondents, whereas 1 item (item 6: All learning objectives can be measured using a single type of test) had majority of the respondents getting it wrong. The respondents wrongly indicated that all learning objectives can be measured using a single type of test.

Among the items that most respondents had wrong include: “It is appropriate to conduct examinations when students are emotionally stable”; “Tests like essays can be scored effectively and consistently without necessarily following the marking scheme”; and “It is important to score tests using personal knowledge of each student's past performance”. Respondents' level of knowledge in validity was determined by summing their scores on all the 20 items. Details are presented in Table 5.

Table 5 presents the level of knowledge in validity. Generally, the scores of the respondents ranged from 0 – 20. Scores between 0 and 10 were classified as low level of knowledge, while scores between 11 and 20 were also classified as high level of knowledge. From Table 5, the majority of the respondents (81.6%) possessed high knowledge in validity. It can, therefore, be said that respondents have high knowledge in validity.

Table 5- *Level of Knowledge on Validity*

Level	Score range	Frequency	Percentage (%)
Low	0 – 10	47	18.4
High	11 – 20	208	81.6
Total		255	100.0

Source: Field survey (2020)

## Research Question 2

*What is the level of knowledge of teachers on reliability?*

The focus of this research question was to determine teachers' knowledge on reliability of assessment. The teachers were subjected to 20 factual statements on reliability of assessment in the quest to measure their knowledge on reliability of assessment. The responses of the respondents were judged as right or wrong. Table 6 presents the responses obtained.

Table 6- *Items on Reliability*

No.	Items	Wrong		Correct	
		f	%	f	%
1.	When assessing students on a particular content, it is important to use several forms of assessment.	34	13.3	221	86.7
2.	It is essential that more time is given to students to complete tasks given to them.	183	71.8	72	28.2
3.	Tests developed to assess students should contain more items.	97	38.0	158	62.0
4.	When scoring tests, more formal procedures should be used.	62	24.3	193	75.7
5.	Scorers should consistently refer to marking scheme.	81	31.8	174	68.2
6.	More than one person should mark or score each student's test (eg. essay or project), then the average score be used finally.	139	54.5	116	45.5
7.	It is important to score all responses to a particular question at a sitting without interruption.	96	37.6	159	62.4
8.	When assessing students, the assessment task should be tailored to each student's ability.	120	47.1	135	52.9

9.	In assessing students, teachers should use equivalent forms of tests.	84	32.9	171	67.1
10.	When grading students, teachers should primarily focus on students' scores on various assessments.	91	35.7	164	64.3
11.	Tests should be administered to students at any time even without prior notice.	109	42.7	146	57.3
12.	In scoring students' test, teachers should be very hard on students who miss classes and be generous to students who are regular in class.	67	26.3	188	73.7
13.	Test scores are said to be reliable when they accurately reflect the content taught.	216	84.7	39	15.3
14.	Adding more items to the test increases the reliability of the test results.	126	49.4	129	50.6
15.	When test results are reliable, they are also valid.	184	72.2	71	27.8
16.	The scoring of essay should be done script by script, but not item by item.	135	52.9	120	47.1
17.	Essay tests should be scored when the marker is physically and mentally alert.	53	20.8	202	79.2
18.	Previously scored items should not be kept out of sight when scoring the rest of the items.	156	61.2	99	38.8
19.	When making decisions about students' learning, it is important to use a combination of results from different assessment methods rather than a single assessment result.	44	17.3	211	82.7
20.	Teachers should select assessment tasks that differentiate best students from the least able students.	145	56.9	110	43.1

---

Source: Field survey (2020)

It can be deduced from Table 6 that, of the 20 factual items, the majority of the respondents were able to identify 10 items correctly, with percentages above 60% (these are, items 1, 3, 4, 5, 7, 9, 10, 12, 17, and 19). For seven items, the majority of the respondents had them wrong (these are, items 2, 6, 13, 15, 16, 18, and 20). The majority of the respondents, for example, wrongly indicated that test scores are said to be reliable when they accurately reflect the content taught (item 13), also they wrongly indicated that when test results are reliable, they are also valid (item 15). Three items nearly had a split in terms of the number of respondents who had them right or wrong (items 8, 11, and 14). To

elaborate, while 50.6% of the respondents were able to correctly indicate that adding more items to the test increases the reliability of the test results, 49.4% did not know that (item 14). Respondents' level of knowledge in reliability was determined by summing their scores on all the 20 items. Details are presented in Table 7.

Table 7- *Level of Knowledge on Reliability*

Level	Score range	Frequency	Percentage (%)
Low	0 – 10	80	31.4
High	11 – 20	175	68.6
Total		255	100.0

Source: Field survey (2020)

Table 7 presents the level of knowledge of respondents on reliability of assessment. As displayed in Table 7, the majority of the respondents (68.6%) had high knowledge in reliability of assessment. This implies that the teachers were generally knowledgeable on issues of reliability in assessment.

### **Research Question 3**

*What is the extent to which teachers engage in practices that enhance validity and reliability of assessment results?*

The aim of this research question was to identify the extent to which teachers engage in practices that enhance validity and reliability of assessment results. The respondents were asked to indicate the extent to which they engage in each of 31 practices outlined. Mean scores were estimated for each of the practices. The mean scores range from 1.0 to 4.0. For the purpose of interpretation, mean scores from 3.0 to 4.0 were judged as mostly followed, whereas mean scores below 3.0 were judged as not followed. This was so

because, ideally, teachers are expected to always engage in each of the practices where necessary. Table 8 presents results on the engagement of the practices.

Table 8- *Practices that Enhance Validity and Reliability*

Practices	M	SD	Remark
I provide clear instructions on how to answer test items.	3.33	.94	MF
Test items I develop do not contain big vocabularies that can hinder the meaning of the item.	2.94	1.04	NF
I give my students a very short time to answer tests that I administer to them.*	3.21	.97	MF
I use difficult test items to assess students, so that I will know the students' ability.*	3.02	.98	MF
I write test items in a way that they do not give students clues to the answer.	2.27	1.07	NF
I score students' tests using personal knowledge of each student's past performance.*	3.20	.97	MF
When assessing students, I make sure the assessment tasks are tailored to each student's ability.	2.31	1.07	NF
I use any type of test to assess students, irrespective of the learning objectives.*	3.20	.99	MF
I develop test items based on the learning objectives.	3.22	.94	MF
The tests I develop to assess students contain more items.	2.56	.83	NF
I arrange test items in order of difficulty, with the difficult ones first.	1.69	.95	NF
I arrange options of test items in a way that the correct answers form a systematic pattern for easy scoring or marking.*	3.33	.95	MF
When scoring, I score all responses to a particular question at a sitting without interruption.	2.35	1.11	NF
The tests I develop closely examine what has been taught.	3.22	.96	MF
I ensure proper room lighting and ventilation when students are being assessed.	3.11	1.02	MF
I allow students to help themselves during examinations. *	3.42	.96	MF
I conduct examinations when students are emotionally stable.	2.78	1.09	NF
I score tests like essays by consistently following the marking scheme.	2.81	1.03	NF
I offer some little assistance to students with low ability during examinations.*	3.34	.96	MF
When scoring tests, I give weaker students some scores to uplift their scores.*	3.30	.94	MF

The assessment procedures I use are free from gender, ethnic, social class, and religious bias and stereotypes.	3.29	1.04	MF
I use several forms of assessment when assessing students on a particular content.	2.71	1.07	NF
When scoring students' scripts test (eg. essay or project), I use more than one person (marker), then the average score be used finally.	1.89	.99	NF
In assessing students, I use equivalent forms of the tests.	2.75	.96	NF
When grading students, I primarily focus on students' scores on various assessments.	2.62	1.08	NF
I administer tests to students at any time, even without prior notice.*	3.14	.89	MF
When scoring students' tests, I am very hard on students who miss classes and generous to students who are regular in class.*	3.30	.96	MF
When scoring essay, I do it script by script, but not item by item.*	2.56	1.06	NF
I score essay tests when I am physically and mentally alert.	3.12	.97	MF
I keep previously scored items out of sight when scoring the rest of the items.	2.45	1.05	NF
When making decisions about students' learning, I use a combination of results from different assessment methods rather than a single assessment result.	3.11	.951	MF
Mean of means	2.89	0.99	NF

Source: Field survey (2020); \*Items reverse-scored; NF – Not Followed; MF – Mostly Followed

From Table 8, among the 31 practices outlined, the respondents did not engage in several practices that enhance validity and reliability of assessment, however, they mostly engaged in only 17. These, among others, include provision of clear instructions on how to answer test items ( $M = 3.33$ ,  $SD = 0.94$ ); development of test items based on the learning objectives ( $M = 3.22$ ,  $SD = 0.94$ ); development of tests based on content taught ( $M = 3.22$ ,  $SD = 0.96$ ); ensuring proper room lighting and ventilation when students are being assessed ( $M = 3.11$ ,  $SD = 1.02$ ); using assessment procedures that are free from gender, ethnic, social class, and religious bias and stereotypes ( $M = 3.29$ ,  $SD = 1.04$ );

not giving students a very short time to answer tests ( $M = 3.21$ ,  $SD = 0.97$ ); not using personal knowledge of student's past performance to score them ( $M = 3.20$ ,  $SD = 0.97$ ); and not allowing students to cheat during examinations ( $M = 3.42$ ,  $SD = 0.96$ ).

The practices that respondents did not follow, among others, include writing test items in a way that give students clues to answers ( $M = 2.27$ ,  $SD = 1.07$ ). This means that response on one item is influenced by another question. Also, the respondent did not arrange test items starting from less difficult ones ( $M = 1.69$ ,  $SD = 0.95$ ); not scoring all responses to a particular question at a sitting without interruption ( $M = 2.35$ ,  $SD = 1.11$ ); and not focusing primarily on students' scores on various assessments when grading ( $M = 2.62$ ,  $SD = 1.08$ ). This means that students' grades were influenced by some non-achievement factors. In sum, it can be said that respondents did not follow practices that enhance validity and reliability of assessment results (mean of means = 2.89,  $SD = 0.99$ ).

#### **Research Question 4**

*What are the sources of invalidity in test scores?*

The focus of this research question was to determine the sources of measurement invalidity in test scores. To answer this research question, samples of tests developed by the teachers were analysed. The test samples were in the following core subjects: English language, mathematics, and integrated science. In addition, observations were conducted during examinations in some of the schools within the study area.

## **Evaluation of Samples Tests**

This section presents results of the evaluation of the test samples used by the teachers in question. Generally, 12 samples of test were evaluated in all. Four samples were selected each from English language, mathematics, and integrated science (see Appendices C, D, and E). The teachers were also interrogated in some instances. These tests were evaluated based on the following dimensions: content representativeness and relevance, thinking processes and skills represented, reliability and objectivity, fairness to different students, and practicality.

### **Content representativeness and relevance**

Evaluation of all the tests examined students based on the topics stipulated in the national syllabus, yet there were no table of specification. Due to the fact that the assessors failed to use the table of specification in the construction of the items, the content representativeness and for that matter, content validity of the assessment instruments are questionable. The assessors of instrument E also failed to sample a wide range of content. Inferring from the number of items set, there was construct under representation of some topics in all the instruments (like Animal production and Waves for Test E, Nitrogen cycle, Cells, Air movement and Rocks for Test F, Measurement, Cells, Ecosystem, Dentition, Respiratory system; and Electronics for Test G).

### **Thinking processes and skills represented**

It is expected that an assessment instrument should represent thinking processes and skills stipulated in the school's curriculum and also allow students enough time to use complex learning skills. The dimensions for teaching, learning and assessment in English at the senior high school level,



assigns knowledge and understanding forty percentage (40%) weight and sixty percentage (60%) weight to application of knowledge. A careful analysis of all the instruments (Test A, B, C, D) revealed that the assessors failed to assign the correct weightings to the domains of instructional relevance due to their failure to use the table of specification in the construction of the items. Hence the thinking processes of students who wrote all the tests were not well represented.

It was found that most of the questions in the essay test also measured lower-level thinking skills (that is, remembering and understanding as stipulated in the Integrated Science syllabus) (see Appendix E), hence inappropriate percentage weightings were apportioned to the domains of instructional relevance. Gleaning from the points stated, the content validity of these assessment instruments was lowered.

The dimensions for teaching, learning and assessment for Mathematics at the Senior high school level, assigns remembering and understanding thirty percentage weight (30%) and seventy (70%) percentage weight to application of knowledge. These percentage weights indicated explains the relative emphasis teachers need to give to the various dimensions in teaching and testing. In view of this, the thinking processes and skills of all the examinees of the tests were not well represented.

### **Reliability and objectivity**

The reliability of assessment results was examined by looking at the test length, the testing conditions; the time allocated to the test, subjectivity in scoring and group variability. With regards to the test length, it can be concluded that the length of Test A was long enough in that, the assessment instrument consisted of 60 multiple choice items with four alternatives each for the paper

one. The paper two also consisted of three essay-type questions (of which only one essay was to be written on) ranging from an informal letter, an article and a short story, as well as a comprehension passage with several questions to be answered by the students. The length of Test B was also long because the assessment instrument consisted of eighty multiple choice items for the paper one and a comprehension passage with several questions, summary and a total of five essay questions in the paper two of which one essay was to be written on.

Tests C and D were also long enough in that, the assessment instruments consisted of 50 multiple choice items for the paper one. Paper two of Test C consisted of five essay-type questions (of which only one essay was to be written on), a comprehension passage with several questions to be answered and a summary question. Paper two of Test D, also comprised four essay-type questions (of which only one essay was to be written on), a comprehension passage with several questions to be answered and a summary question. Tests B, C and D can be likened to the WASSCE English language script since the assessors inculcated a third aspect of the paper two, which is, the summary.

In relation to the time allocated to test A, the entire testing period was two hours and the assessors clearly apportioned the time between the two sections such that, the paper two was to be covered within an hour and ten minutes whereas the multiple-choice items were to be answered within fifty minutes. As such, students might have worked efficiently within the time frame. Hence, the time allocation can be said to be fair enough to enable students to work efficiently. Though the three-hour period allotted to Test B was quite long, the assessors failed to apportion the period between the two sections. That is,

the time given to students to complete each section of the test was not specified. The omission of time specification may result in students wasting time on some aspects of the test without focusing on the other aspects of the paper. Hence, they might not have been able to complete the test on time. Students might have therefore resulted to guessing or cheating thereby introducing errors.

In spite of the fact that the two-hour period allotted to Test C and D was quite short, the assessors clearly apportioned one hour fifteen minutes to the Paper two and forty- five, minutes to the Paper one. The specification of time for the different sections of test C and D might have prompted students to work within the given time frame. Inferring from the time allocated to the tests, Test A yielded a high reliability whereas the reliability of Test B, C and D was lowered due to unclear directions with regards to time allocation (for Test B) and insufficient time (for Test C and D).

Most options of the multiple-choice items were arranged vertically in Test A, C and D and this bridged any source of hindrance to students' ability to easily identify the key to the various items whilst Test B had Horizontal arrangement of options which much as clearly arranged, overlapped each other which could have led to some confusion in the choices of the key.

With regards to the test length, it can be concluded that the length of instruments E, F and G were long in that, the instruments consisted of forty multiple choice items with four alternatives for each item in the Paper One. Though the length of Test E and G were also long, the assessors did not give the students enough time on practical. Instrument H was however short in length in that it comprised three compulsory essay type tests, one practical test and twenty multiple choice items.

### **Fairness to different students**

Considering the issue of fairness, deliberations with the assessors of all the tests revealed that, students were notified about the nature and purpose of the exam several weeks before its commencement. As a result, students were offered opportunity to prepare adequately for the tests. According to the assessors of Test A, C and D, some aspects of the syllabus were not evenly taught to all the classes. Much as the students had alternatives with regards to the essay aspect of the paper, the students were limited to the topics they were taught in terms of their choice of question to be answered.

The assessors also made mention that provisions were made for physically challenged students to write the exams without any form of impediment. Again, in ensuring fairness, the assessors made sure that the test items did not contain words or phrases that served as clues to some of the students, whilst being disadvantageous to the other students in other classes. Thus, the tests were fair with regard to wording, ethnicity and gender.

Regarding the integrated science paper (Appendix E), though students were made aware of the topics the tests would cover, some items (16, 24, 33, 35, 37, 38 and 40) of Test E and (25, 26, 31, 32, 33, 37, and 38) of Test F were based on topics which weren't covered by all the form one classes. The one hour, forty- five minute's period allotted to the Test F was inadequate and the assessors also failed to apportion the period between the two sections of the test. The students were supposedly treated equally without any form of discrimination in all the classes during the period of test administration and scoring.

## **Practicality**

The assessors were asked how best they ensured practicality and they alluded that, efforts were made to ensure adequate supply of answer booklets and other materials needed for the exams when it comes to Test A and D, whereas in the case of Test B and C, much as adequate supply of exam scripts and answer booklets were to be ensured, because the question papers were printed on the very day for writing, some classes were reported to have received their scripts much later. This instance could lead to leakage of questions to students who started later.

With regards to time, when the time allotted to a test is too short, students are likely not to complete the test which will affect their performance whereas when too much time is given, the students may resolve to cheat. The two-hour period allotted to Test A was good considering the number of items answered. On the other hand, the three hours allotted to Test B was questionable due to unclear directions. Also, the two hours period allotted to Test C and D was also quite short and this could affect the validity and use of the results.

The two hour and one hour thirty minutes period allotted to Test E and H respectively was good in comparison with the period allotted for the Integrated Science WASCE Examination. Whereas the one hour forty-five minutes period allotted to Test F was short, the two and half hour allotted to test G was too long. Students who answered instrument F and G might have guessed or cheated incorporating errors. Moreover, in view of the fact that the stipulated time was not apportioned between the two sections of Tests E, and F, the students might have not known the exact time allotted to each section to guide their responses. This could affect the validity and use of the results.

## Observation of Examination Procedures and Conditions

In addition to the evaluation of tests developed and used in assessing students, observation was conducted during the administration of the various assessments. For each of the seven schools used in the study, two observations were carried out. In all, 14 observations were carried out across all the schools. Basically, the observation covered three thematic areas namely, room conditions, invigilation, and testing conditions. Details of the observation are presented in Tables 9, 10, and 11.

Table 9- *Room Conditions (n = 14)*

Condition	Frequency	Percentage (%)
Arrangement of tables and chairs		
Appropriate	10	71.4
Not appropriate	4	28.6
Ventilation		
Good	8	57.1
Poor	6	42.9
Lighting		
Good	11	78.6
Poor	3	21.4
Sitting posture of examinees		
Appropriate	9	64.3
Not appropriate	5	35.7
Serenity of environment		
Good	6	42.9
Poor	8	57.1

Source: Field survey (2020)

As shown in Table 9, in terms of arrangement of tables and chairs (71.4%), ventilation (57.1%), lighting (78.6%), and sitting posture of examinees (64.3%) were seen to be good and appropriate. However, the majority of the cases observed indicate the serenity of the environment was poor (57.1%). Even though majority of the observations reported good and appropriate conditions, it is expected the aforementioned should be good and appropriate, which is not negotiable. Table 10 presents results on invigilation conditions.

Table 10- *Invigilation (n = 14)*

Condition	Frequency	Percentage (%)
Number of invigilators per room		
Adequate	8	57.1
Not adequate	6	42.9
Invigilators making phone calls, reading, chatting		
Yes	9	64.3
No	5	35.7
Invigilators interrupting students		
Yes	10	71.4
No	4	28.6
Announcement of time remaining		
Frequently	3	21.4
Not frequently	11	78.6
Unnecessarily	-	-
Invigilators intermittently walking around the room during testing		
Frequently	10	71.4
Not frequently	4	28.6
Invigilators pestering students		
Yes	7	50.0
No	7	50.0
Position of invigilators.		
Appropriate	9	64.3
Not appropriate	5	35.7

Source: Field survey (2020)

From Table 10, the observation showed that on many cases it was found that the number of invigilators per room was adequate (57.1%); announcement of time remaining was no frequently done (78.6%); invigilators intermittently walking around the room during testing (71.4%); and the position of the invigilators were appropriate (64.3%). It was, however, observed on a number of occasions that invigilators were making phone calls, reading, and chatting from phones (64.3%); invigilators interrupted students (71.4%); and some cases, invigilators pestered students (50%). Table 11 presents results on conditions of the test.

Table 11- *Testing Conditions (n = 14)*

Condition	Frequency	Percentage (%)
Time allowed for the tests.		

Appropriate	10	71.4
Not appropriate	4	28.6
Start of test.		
Exactly on scheduled time	7	50.0
Not exactly on scheduled time	7	50.0
Stopping time.		
Exact time	6	42.9
Not exact time	8	57.1
Availability of supplementary answer booklets.		
Available and adequate	4	28.6
Available but not adequate	10	71.4
Not available	-	-
Supplementary question papers.		
Available	11	78.6
Not available	3	21.4
Emergency medical care		
Available	-	-
Not available	14	100.0

Source: Field survey (2020)

The results in Table 11 show that, regarding the testing conditions, it was observed that time allowed for the tests were appropriate in most cases (71.4%); the tests started on scheduled time (50%); and there were supplementary question papers available (78.6%). It was, however, observed that that in most cases the tests did not stop at the exact time (57.1%); supplementary answer booklets were available but not adequate (71.4%); and there were no emergency medical care services at all the schools observed (100%).

### **Descriptive Information on Knowledge on Validity, Knowledge on Reliability, and Adherence to Validity and Reliability Practices**

This section provides descriptive information on knowledge on validity, knowledge on reliability, adherence to validity and reliability practices. Details of the descriptive statistics are presented in Table 12.

Table 12- *Descriptive Statistics on Knowledge on Validity, Knowledge on Reliability, and Practices*

Parameters	Validity	Reliability	Practices
Mean	13.96	11.29	2.89



Standard deviation	3.26	2.30	0.35
5% Trimmed mean	14.02	11.35	2.90
Median	14.0	12.0	2.94
Skewness	-.308	-.521	-.304
Std. Error	.153	.153	.153
Z <sub>skewness</sub>	-2.01	-3.41	-1.99

Source: Field survey (2020)

From Table 12, the mean, 5% trimmed mean, and median for knowledge on validity, knowledge on reliability, and practices were approximately the same, suggesting that these data were normally distributed. Also, the  $Z_{skewness}$  for knowledge on validity, and practices were within the ranges of  $\pm 3.29$ , however, that of knowledge in reliability deviated a little. In addition, visual examination of the histograms and the normal Q-Q plots for each of the variables confirms normality of the distributions (see Appendix F).

### **Hypothesis 1**

*H<sub>0</sub>: There will be no significant gender difference in teachers' knowledge on (a) validity and (b) reliability of assessment.*

*H<sub>1</sub>: There will be a significant gender difference in teachers' knowledge on (a) validity and (b) reliability of assessment.*

This hypothesis was interested in testing differences in level of knowledge on validity and reliability on the basis of gender. This hypothesis was tested using one-way multivariate analysis of variance (MANOVA). The independent variable is gender, which has two levels: male and female. The dependent variables are the mean scores in knowledge on validity and knowledge on reliability. The dependent variables were measured on interval scale. Assumptions such as normality and linearity were checked and adhered to (see Appendix F). The results of the Box's test did not violate the equality of

variance-covariance assumption, Box's  $M = 2.03$ ,  $F(3, 4111819.30) = .67$ ,  $p = .570$ . Having met this assumption, Wilks' Lambda multivariate test was performed. Table 13 presents the results of the multivariate test.

Table 13- *Wilks' Lambda Multivariate Test for Gender Differences in Knowledge in Validity and Reliability*

Effect	Value	F	df1	df2	Sig.	Partial Eta Squared
Intercept	.030	4083.439	2	252	.000	.970
Gender	.996	.557	2	252	.574	.004

Source: Field survey (2020)

The result of the Wilks' Lambda multivariate test in Table 13 show no statistically significant gender difference on the combined knowledge in validity and reliability,  $F(2, 252) = .56$ ,  $p = .574$ , partial eta squared = .004. Separate univariate ANOVAs were performed at Bonferroni's alpha of .025. The Levene's test for equality of variance showed no violation of the assumption for validity ( $p = .784$ ) and reliability ( $p = .613$ ). Table 14 presents the univariate results.

Table 14- *Univariate Test for Gender Differences in Knowledge in Validity and Reliability*

Source	Dependent Variable	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	Validity	1	48113.590	4540.442	<.001	.947
	Reliability	1	31517.591	5961.205	<.001	.959
Gender	Validity	1	11.849	1.118	.291	.004
	Reliability	1	.461	.087	.768	<.001
Error	Validity	253	10.597			
	Reliability	253	5.287			
Total	Validity	255				
	Reliability	255				

Source: Field survey (2020)

In terms of knowledge on validity, the result showed no statistically significant gender difference,  $F(1, 253) = 1.12, p = .291$ , partial eta squared = .004 (Table 14). The result implies that, practically, gender explained 0.4% of the variations in knowledge on validity. The results further showed no statistically significant gender difference in knowledge on reliability,  $F(1, 253) = .09, p = .768$ , partial eta squared  $< .001$ . The results generally, imply that male and female teachers do not vary in their level of knowledge on validity and reliability of assessment.

Based on the results gathered, the null hypothesis that “There will be no significant gender difference in teachers’ knowledge in (a) validity and (b) reliability in assessment” was failed to be rejected.

## **Hypothesis 2**

*H<sub>0</sub>: There will be no significant difference in teachers’ knowledge on (a) validity and (b) reliability of assessment with respect to years of teaching experience.*

*H<sub>1</sub>: There will be a significant difference in teachers’ knowledge on (a) validity and (b) reliability of assessment with respect to years of teaching experience.*

The focus of this hypothesis was to determine whether or not teachers differ in their level of knowledge on validity and reliability of assessment with respect to years of teaching experience. This hypothesis was tested using one-way MANOVA. The dependent variables are the mean scores in knowledge on validity and knowledge on reliability. These variables were measured on continuous basis. The independent variable was years of teaching experience. This variable has five levels: less than 1 year, 1 – 4 years, 5 – 8 years, 9 – 12 years, and 12 years and above. The normality and linearity assumptions were met (see Appendix F). The results of the Box’s test did not violate the equality

of variance-covariance assumption, Box's  $M = 7.66$ ,  $F(12, 42139.87) = .67$ ,  $p = .826$ . Upon meeting this assumption, Wilks' Lambda multivariate test was performed. Details of the results are presented in Table 15.

Table 15- *Wilks' Lambda Multivariate Test for Differences in Knowledge in Validity and Reliability with respect to Years of Teaching*

Effect	Value	F	df1	df2	Sig.	Partial Eta Squared
Intercept	.039	3092.879	2	249	<.001	.961
Years of teaching	.890	3.721*	8	498	<.001	.056

Source: Field survey (2020); \*Significant,  $p < .05$

Result from Table 15 show a statistically significant difference in the combined knowledge in validity and reliability in assessment,  $F(8, 498) = 3.72$ ,  $p < .001$ , partial eta squared = .06. Years of teaching accounted for 6% of the variations in knowledge in validity and reliability in assessment. The Levene's test for equality of variance showed no violation of the assumption for validity ( $p = .236$ ) and reliability ( $p = .737$ ). The results of the univariate tests are presented in Table 16.

Table 16- *Univariate Test for Differences in Knowledge in Validity and Reliability with respect to Years of Teaching*

Source	Dependent Variable	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	Validity	1	35763.671	3571.459	.000	.935
	Reliability	1	22549.233	4398.748	.000	.946
Years of teaching	Validity	4	47.343	4.728*	.001	.070
	Reliability	4	14.133	2.757	.028	.042
Error	Validity	250	10.014			
	Reliability	250	5.126			
Total	Validity	255				
	Reliability	255				

Source: Field survey (2020); \*Significant,  $p < .025$  (Bonferroni's alpha)

The univariate results are presented in Table 16, using Bonferroni's adjusted alpha of .025. With respect to years of teaching, there was a statistically significant difference in knowledge in validity,  $F(4, 250) = 4.73, p = .001$ , partial eta squared = .07. Years of teaching explained 7% of the variations in teachers' knowledge in validity. There was, however, no statistically significant difference in teachers' knowledge in reliability on the basis of years of teaching,  $F(4, 250) = 2.76, p = .028$ , partial eta squared = .04. Table 17 presents a multiple comparison using Tukey on knowledge in validity.

Table 17- *Tukey HSD Multiple Comparisons*

(I) Years of teaching	(J) Years of teaching	Mean Diff. (I-J)	S.E	Sig.
Less than 1 year	1-4 years	1.4799	0.84	.396
	5-8 years	.8057	0.94	.914
	9-12 years	-.0382	0.92	1.000
	Above 12 years	-.5623	0.85	.964
1-4 years	Less than 1 year	-1.4799	0.84	.396
	5-8 years	-.6742	0.65	.835
	9-12 years	-1.5182	0.60	.090
	Above 12 years	-2.0422*	0.49	<.001
5-8 years	Less than 1 year	-.8057	0.94	.914
	1-4 years	.6742	0.65	.835
	9-12 years	-.8439	0.74	.788
	Above 12 years	-1.3680	0.66	.233
9-12 years	Less than 1 year	.0382	0.92	1.000
	1-4 years	1.5182	0.60	.090
	5-8 years	.8439	0.74	.788
	Above 12 years	-.5240	0.62	.915
Above 12 years	Less than 1 year	.5623	0.85	.964
	1-4 years	2.0422*	0.49	<.001
	5-8 years	1.3680	0.66	.233
	9-12 years	.5240	0.62	.915

Source: Field survey (2020); \*Significant,  $p < .025$  (Bonferroni's alpha)

Among the multiple comparisons shown in Table 17, significant difference exists only between teachers who have taught for 1 – 4 years and those who have taught for more than 12 years ( $p < .001$ ). It was evident that

teachers who have served for more than 12 years ( $M = 14.97$ ,  $SD = 2.97$ ) had higher knowledge in validity than those who have taught for 1 – 4 years ( $M = 12.93$ ,  $SD = 3.21$ ). However, among all the others, there were no significant differences.

On the basis of the results this study, the null hypothesis that “There will be no significant difference in teachers’ knowledge in (a) validity and (b) reliability in assessment with respect to years of teaching experience” was rejected in favour of its alternative hypothesis.

### **Hypothesis 3**

*H<sub>0</sub>: Teachers’ knowledge on (a) validity and (b) reliability of assessment will not significantly predict the extent to which they engage in practices that enhance validity and reliability of assessment results.*

*H<sub>1</sub>: Teachers’ knowledge on (a) validity and (b) reliability of assessment will significantly predict the extent to which they engage in practices that enhance validity and reliability of assessment results.*

The aim of this hypothesis was to determine the influence of teachers’ knowledge on validity and reliability of assessment on the extent to which they engage in practices that enhance validity of assessment results. Simultaneous multiple linear regression analysis was performed to test this hypothesis. The predictor variables were teachers’ scores in knowledge on (a) validity and (b) reliability. These scores were on continuous bases. The criterion variable was engagement in practices that enhance validity of assessment results. This variable was also measured on continuous basis. Assumptions such as normality of residuals, linearity, and homoscedasticity were adhered to (see Appendix G). Table 18 presents the overall model summary.

Table 18- *Model Summary for Knowledge in Validity and Reliability*

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.509	.259	.253	.30596	1.807

Source: Field survey (2020); \*Significant,  $p < .05$ ;  $F(2, 251) = 43.81$ ,  $p < .001$

From Table 18, there was no autocorrelation since the Durbin-Watson result (1.81) is greater than 1.5 but less than 2.5. The overall model was statistically significant,  $F(2, 251) = 43.81$ ,  $p < .001$ , adjusted  $R^2 = .25$ . The result implies that knowledge in validity and reliability jointly accounted for about 25% of the variations in adherence to practices that enhance validity and reliability in assessment. The regression coefficients are presented in Table 19.

Table 19- *Regression Coefficients for Knowledge in Validity and Reliability*

Parameter	Unstd. Coef.		Std. Coef.	t	Sig.	VIF
	B	S.E.	Beta ( $\beta$ )			
(Constant)	1.888	.112		16.804	<.001	
Validity	.042*	.006	.383	6.745	<.001	1.09
Reliability	.037*	.009	.243	4.282	<.001	1.09

Source: Field survey (2020); \*Significant,  $p < .025$

The results in Table 19 show that both knowledge in validity ( $B = .042$ ,  $p < .001$ ) and reliability ( $B = .037$ ,  $p < .001$ ) are statistically significant predictors of adherence to practices that enhance validity of assessment results. Both knowledge in validity and reliability were positive predictors of adherence to practices that enhance validity in assessment. The results implies that a unit increase each in knowledge in validity and knowledge in reliability would lead .042 and .037, respectively, increase in adherence to practices that enhance validity in assessment.

Following the results of this study, the null hypothesis which states that “Teachers’ knowledge on (a) validity and (b) reliability of assessment will not significantly predict the extent to which they engage in practices that enhance validity and reliability of assessment results” was rejected in favour of the alternative hypothesis.

## Discussion

This section discusses the findings of the study. The discussion was done based on the following topical areas:

1. Teachers’ Knowledge on Validity and Reliability
2. Teachers’ Engagement in Practices that Enhance Validity and Reliability of Assessment Results



3. Sources of Test Score Invalidity
4. Demographic Characteristics and Teachers' Knowledge on Validity and Reliability

### **Teachers' Knowledge on Validity and Reliability**

It was found that overwhelming majority of the respondents were able to correctly answer 14 out of the 20 items, with a corresponding percentage above 60%. However, 5 of the items were wrongly scored by most of the respondents. Similarly, the majority of the respondents were able to identify 10 items correctly, with percentage above 60%. The majority of the respondents, for example, wrongly indicated that test scores are said to be reliable when they accurately reflect the content taught, also they wrongly indicated that when test results are reliable, they are also valid. The respondents further wrongly indicated that it is appropriate to conduct examinations when students are emotionally unstable; tests like essays can be scored effectively and consistently without necessarily following the marking scheme; and it is important to score tests using personal knowledge of each student's past performance. The study, generally, revealed that the majority of the respondents possessed high knowledge in validity and reliability.

The findings of the current study agree with several other studies (Ankomah et al., 2020; Ankomah & Nugba, 2020; Pipia, 2014). Pipia (2014), for instance, found that the teachers provide accurate information to students before administering the test, they also consistently refer to the scoring rubric when scoring the test papers of their students. It was further found that the teachers do not consider personal and confounding elements while assessing the test results. This implies that the teachers have knowledge in validity and

reliability in terms of students' assessment. Similarly, studies by Ankomah et al. (2020); and Ankomah and Nugba (2020) have reported high test construction skills among teachers in Sekond-Takoradi and Cape Coast, respectively. It is worthy to note that though the findings of the current study agree with the aforementioned, the cases of Ankomah et al. (2020); and Ankomah and Nugba (2020) were only limited to skills in test construction, but not assessment in general. There is the likelihood that in the cited studies, the teachers were not knowledgeable in assessment in general.

It was revealed that both knowledge on validity and reliability were positive predictors of adherence to practices that enhance validity in assessment. The findings of this study have implications on the validity and reliability on the practice of assessment by these teachers. The findings are in harmony with the National Education Association [NEA] (2003) that a teacher is expected to have adequate and absolute knowledge in assessing their learners as their least trait. The NEA further adds that accurate assessment is impossible until assessors are knowledgeable on validity and reliability implications of assessment results. Also, the Standards (1999) specify that teachers should be knowledgeable in choosing and developing assessment methods; administering, scoring and interpreting assessment results; using assessment results for decision making and grading; communicating assessment results; and recognizing unethical assessment practices. This was, however, not the case in the current study.

The finding of the current study disconfirms that of Ing et al. (2015) who found that assessors or teachers displayed little knowledge about the specification table. Moreover, a careful scrutiny revealed that a greater

proportion of the teachers never got involved in any organised course to upgrade their knowledge regarding specification tables, hence, they lacked the competence to develop the appropriate specification table indicating limited knowledge on the teachers' part regarding the essence of specification table and how they are used. Also, findings by Anhwere (2009) revealed that teachers failed to adhere to basic testing principles in constructing assessment test, and thus consider managing college assessment practices to be extra load for their profession. Anhwere, therefore, concluded, on the whole, tutors in the teacher training colleges had limited skills and competence in the knowledge of construction, administration of classroom/teacher made tests and scoring of essay-type tests.

### **Teachers' Engagement in Practices that Enhance Validity and Reliability of Assessment Results**

It was revealed that the respondents did not engage in several practices that enhance validity and reliability of assessment, however, they mostly engaged in only seventeen. In sum, it can be said that respondents did not follow practices that enhance validity and reliability of assessment results. Ideally, it is expected that teachers engage in all practices stipulated to regulate and bind the conduct of assessment in general. Doing so minimises the errors that are likely to influence test scores, hence making them not reliable. Similarly, adherence to these practices provides more evidence to support the soundness and meaningfulness of the interpretation of such results.

The results of the current study confirm previous studies (Oduro-Kyireh, 2008; Risnen et al., 2016; Anhwere, 2009). Oduro-Kyireh's study aimed to find out whether the procedures used by teachers in constructing, administering and

evaluating class achievement tests and interpreting the results of these tests conformed to the principles and guidelines prescribed by measurement specialists. The study found that the following principles are not commonly used by teachers: the use of test specification tables to determine the items included in the test; preparing more items than needed for the test; and evaluation of the test as a whole according to the criteria of clarity, practicality, validity, efficiency and fairness. Oduro-Kyireh concluded that the instructors lack the requisite skills needed to construct tests. This, therefore, suggests that teachers' level of knowledge on validity and reliability is limited in the sense that all the principles regarding test construction are founded on ensuring validity and reliability in assessment.

In another study, Marmah and Impraim (2013) examined the competence of university lecturers in creating multiple-choice test questions. The results showed that most of the time the teachers did not follow the principles established by measurement experts when developing the multiple choice item. For example, as part of the task planning, the instructors did not create a table of test specifications and did not emphasize or capitalize negative phrases or phrases in the text.

The results of the current study, on one hand, found that teachers had high level of knowledge on validity and reliability of assessment. On the other hand, the teachers did not adhere to a number of recommended practices, though they had high knowledge in validity and reliability. Obviously, these results appear to be a paradox. Ankomah et al. (2020) found that attitude towards test construction played a significant role as far as adherence to recommended principles are concerned. The authors explained that teachers may possess high

level of skills, but when they have a negative attitude, they are less likely to adhere to recommended principles.

### **Sources of Test Score Invalidity**

Findings of the study revealed some issues regarding content representativeness and relevance, thinking processes and skills represented, reliability and objectivity, fairness to different students, and practicality of the test samples evaluated. In addition, the observations revealed that the conditions of test administration, namely, room conditions, invigilation, and testing conditions were not good and appropriate. The findings of this study are in support of that of Cooper et al. (2016) who examined teachers' understanding of validity and reliability in assessment. Results from their study showed that qualities of the test items were good, a few items were determined to be potentially problematic. Based on this, Cooper et al. indicated that the teachers have adequate knowledge in validity and reliability in classroom assessment, hence the quality of their test items.

Similarly, Ing et al. (2015) Finding of the study further revealed that teacher-made assessments conformed to content validity. Moreover, majority of the teachers never consulted the specification table when developing their assessment instruments, indicating that the lack of awareness and knowledge on the teachers' part regarding the essence of specification table and how they are used. Quansah et al. (2019) explored the test construction skills of SHS teachers in the Cape Coast Metropolitan Assembly (CCMA). In total, 15 samples of examination papers were taken from each of the three schools. Findings of the study indicated that teachers were inadequately skilled in constructing end-of-terms examination questions. It was obvious as there were problems with the

relevancy of the assessment, consistency, and unfairness regarding the evaluation of the assessment.

Koskal (2004), found, among others, the following flaws: the tests failed to identify or state the target audience, the domain the test purported to assess, duration of the test and grade points that were allocated to each item. The duration for each task was not specified on the assessment instrument and majority of the instructions were ambiguous and imprecise. Most items were engulfed with grammatical and typographical errors. Some items were also not based on the learning objectives. Some items failed to serve the purpose for which they were intended. For instance, the items assessed speaking skills instead of writing skills, as they were instructed to fill the gaps of conversational sentences.

In their study, Hamman-Tukur and Kamis (2000), found that most of the examination questions assessed simple learning outcomes of knowledge and comprehension categories of the cognitive domain at the expense of learning outcomes that call for synthesis and evaluation. These reduce the validity of the test results produced by those tests.

### **Demographic Characteristics and Teachers' Knowledge on Validity and Reliability**

The results further showed no statistically significant gender difference in knowledge in validity and knowledge in reliability. These, generally, imply that male and female teachers do not vary in their level of knowledge in validity and reliability of assessment. In other words, gender does not necessarily matter as far as knowledge on validity and are concerned. The finding of the current study is in sharp contrast to a couple of studies (Alkharusi, 2009; Alkharusi,

2011; Yan, 2016; Asamoah et al., 2019). Alkharusi (2009), for example, who found that assessment knowledge of pre-service teachers tended to vary as a function of gender. Alkharusi found that males tended to have on average a higher level of knowledge in educational assessment than females.

Similarly, Yan (2016) found that this woman showed higher levels of self-appraisal practices than the man. In another study, Asamoah et al. (2019) found a significant difference in the formative assessment knowledge of male and female high school teachers and that male high school teachers performed better in their formative assessment knowledge than their female peers.

Alkharusi (2011), in his study, revealed a significant gender difference in terms of test construction and administration, and also communication of assessment results. Males had higher skills in test construction and administration, and also communication of assessment results compared to the females.

With respect to years of teaching, there was a statistically significant difference in knowledge on validity, however, no statistically significant difference in teachers' knowledge in reliability on the basis of years of teaching. It was evident that teachers who have served for more than 12 years had higher knowledge in validity than those who have taught for 1 – 4 years. The finding of the current study agrees with Kinyua and Okunya (2014) whose study unravelled that the experience of teachers has an effect on validity and reliability of the tests. It was found that experienced teachers who may have been given some level of practical education or knowledge and apply this knowledge when designing test exhibit competence and confidence as against their colleagues

who have no such knowledge or skills. Subsequently, a conclusion that teacher-made tests have higher validity and consistency was made.

### **Chapter Summary**

The chapter presented results and discussion of the findings. It was found that majority of the respondents' possessed high knowledge in validity and reliability. It was also revealed that the respondents did not engage in several practices that enhance validity and reliability of assessment, however, they mostly engaged in only seventeen. The results further showed no statistically significant gender difference in knowledge in validity and knowledge in reliability. With respect to years of teaching, there was a statistically significant difference in knowledge in validity, however, no statistically significant difference in teachers' knowledge in reliability on the basis of years of teaching. Finally, it was revealed that both knowledge in validity and reliability were positive predictors of adherence to practices that enhance validity in assessment.



## **CHAPTER FIVE**

### **SUMMARY, CONCLUSION AND RECOMMENDATION**

This chapter presents a summary of the entire study as well as the conclusions and recommendations. The summary was looked at in terms of the overview of the study and the key findings that emerged.

#### **Summary**

##### **Overview of the study**

The present study focused on investigating knowledge in validity and reliability in classroom assessment among SHS teachers in Sekondi-Takoradi Metropolitan Assembly. The study was guided by four research questions, which are:

1. What is the level of knowledge of teachers on validity?
2. What is the level of knowledge of teachers on reliability?
3. What is the extent to which teachers engage in practices that enhance validity and reliability of assessment results?
4. What are the sources of invalidity in test scores?

In addition, the study tested the following three hypotheses:

1.  $H_0$ : There will be no significant gender difference in teachers' knowledge on (a) validity and (b) reliability of assessment.  
 $H_1$ : There will be a significant gender difference in teachers' knowledge on (a) validity and (b) reliability of assessment.

2. H<sub>0</sub>: There will be no significant difference in teachers' knowledge on (a) validity and (b) reliability of assessment with respect to years of teaching experience.

H<sub>1</sub>: There will be a significant difference in teachers' knowledge on (a) validity and (b) reliability of assessment with respect to years of teaching experience.

3. H<sub>0</sub>: Teachers' knowledge on (a) validity and (b) reliability of assessment will not significantly predict the extent to which they engage in practices that enhance validity and reliability of assessment results.

H<sub>1</sub>: Teachers' knowledge on (a) validity and (b) reliability of assessment will significantly predict the extent to which they engage in practices that enhance validity and reliability of assessment results.

This study adopted the embedded mixed method design. Essentially, 981 public SHS teachers who teach in 11 public schools STMA were used as the population for the study. Of the 981 teachers, 575 were males and the remaining 406 were females. A sample of 278 was used for the study. The stratified sampling method was, therefore, employed to enrol respondents for the study. The stratification variables were schools and type of school based on sex composition. Questionnaire and observation checklist were designed for collecting data for the present investigation. Frequencies and percentages were used in analysing data collected on Research Questions 1 and 2. Moreover, means and standard deviations were used in analysing data collected on Research Question 3. Content analysis and frequencies and percentages were used in analysing data collected on Research Question 4. Data collected on

Hypotheses 1 and 2 were tested using one-way MANOVA. Hypothesis 3 was tested using simultaneous multiple linear regression analysis.

### **Key of findings**

The following emerged from the study:

1. Results indicate that majority of the respondents correctly answered 14 out of the 20 items, with a corresponding percentage above 60%. However, 5 of the items were wrongly scored by most of the respondents. The majority of the respondents possessed high knowledge in validity.
2. It was further revealed that majority of the respondents (68.6%) had high knowledge in reliability of assessment.
3. It was revealed that the respondents did not engage in several practices that enhance validity and reliability of assessment, however, they mostly engaged in only seventeen. In sum, it can be said that respondents did not follow practices that enhance validity and reliability of assessment results.
4. Findings of the study revealed some issues regarding content representativeness and relevance, thinking processes and skills represented, reliability and objectivity, fairness to different students, and practicality of the test samples evaluated. In addition, the observations revealed that the conditions of test administration, namely, room conditions, invigilation, and testing conditions were not good and appropriate.
5. The results further showed no statistically significant gender difference in knowledge in validity and knowledge in reliability. These, generally,

imply that male and female teachers do not vary in their level of knowledge in validity and reliability of assessment.

6. With respect to years of teaching, there was a statistically significant difference in knowledge in validity, however, no statistically significant difference in teachers' knowledge in reliability on the basis of years of teaching. It was evident that teachers who have served for more than 12 years had higher knowledge in validity than those who have taught for 1 – 4 years.
7. Finally, it was revealed that both knowledge in validity and reliability were positive predictors of adherence to practices that enhance validity in assessment.

### **Conclusion**

It can be deduced from the results of this study that senior high school teachers in Sekondi-Takoradi are knowledgeable on issues of validity and reliability of assessment, and this has the tendency in helping them improve their adherence to practices that enhance validity and reliability of assessment results. It can also be concluded that though the teachers were knowledgeable on validity and reliability of assessment, they did not engage in several practices that enhance validity and reliability in assessment, and this could partly be due to factors such as commitment, attitude, and motivation, among others. It can also be concluded that the soundness of the interpretations and uses of the end of semester results of the schools in this study for certain purposes are questionable.

### **Recommendations**

Following the findings and conclusions, the following are recommended:

1. The Metropolitan Directorate of Education, STMA, is encouraged to continue and intensify their workshops for teachers, particularly in the area of assessment.
2. The Ministry of Education (MoE), Ghana Education Service (GES), and headmasters of the senior high schools in STMA are encouraged to sensitize teachers on the need to strictly adhere to practices that enhance validity and reliability of assessment results.
3. Senior high school teachers in STMA are entreated to embrace practices that have been recommended in the conduct of assessment so as to reduce errors and also improve the validity of students' assessment results.
4. It is recommended that examination boards of the senior high schools in STMA engage the services of professionals in assessment so as to scrutinise tests developed to assess students, and also guide them in the conduct of the examinations so that they can engage in practices that would enhance the validity and reliability of their assessment results.
5. The headmasters and teachers of senior high schools in STMA are entreated in the provision and acquisition of materials to aid in the conduct of examinations.

#### **Suggestions for Further Studies**

1. A sequential mixed methods design is recommended for future investigations so as to provide possible explanations on the results of this study.

2. Future studies are encouraged to replicate or incorporate teachers from other regions in the country so as to provide a broader perspective on teachers' knowledge in validity and reliability from the national view.

## REFERENCES

- Adamu, G. G., Dawha, J. M., & Kamar, T. S. (2015). A scheme for assessing technical teachers' competencies in constructing assessment instruments in technical colleges in Gombe State. *ATBU, Journal of Science, Technology & Education (JOSTE)*, 3(2), 1-8.
- Agu, N. N., Onyekuba, C., & Anyichie, A. C. (2013). Measuring teachers competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. *Educational Research and Reviews*, 8(8), 431-439.
- Alkharusi, H. (2009). Correlates of teacher education students' academic performance in an educational measurement course. *International Journal of Learning*, 16, 1-15.
- Alkharusi, H. (2011). Self-perceived assessment skills of pre-service and in-service teachers. *Jurnal Pendidikan Malaysia*, 36(2), 9-17.
- Amedahe, F. K. (2014). Test construction practices in secondary schools in the Central Region of Ghana. *The Oguaa Educator*, 2, 52-63.
- Amedahe, F. K., & Asamoah-Gyimah, K. (2015). *Introduction to educational research*. Cape Coast: UCC Printing Press.
- Amedahe, F. K. (1989). Testing practices in secondary schools in the Central region of Ghana. (Unpublished master's thesis). University of Cape Coast, Cape Coast, Ghana
- Amedahe, F. K., & Asamoah-Gyimah, K. (2015a). *Introduction to measurement and evaluation*. Cape Coast: UCC Printing Press.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999).

Standards for educational and psychological testing. Washington, DC: American Psychological Association.

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

Ankomah, F., & Nugba, R. M. (2020). Validation of Test Construction Skills Inventory through the lens of Item Response Theory (IRT). *American Journal of Creative Education*, 3(2), 86-100.

Ankomah, F., Amedahe, F. K., & Cobbinah, A. (2020). The issue of non-adherence to test construction principles: Do teachers' commitment or attitude by the teacher really matter? *International Journal of Social Sciences & Educational Studies*, 7(4), 109-121.

Asamoah, D., Songnalle, S., Sundeme, B., & Derkye, C. (2019). Gender difference in formative assessment knowledge of senior high school teachers in the Upper West Region of Ghana. *Journal of Education and Practice*, 10(6), 54-58.

Assessment and Learning Research Synthesis Group [ALRSG]. (2004). *A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes*. London: Evidence for Policy and Practice Information and Coordinating Centre.

Boothroyd, R. A., McMorris, R. F., & Pruzek, R. (1992, April). *What do teachers know about measurement and how did they find out?* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.



- Cassel, R. (2003). Confluence is a primary measure of test validity and it includes the creditability of test taker. *College Student Journal*, 37:348-353.
- Cooper, T., Pittman, A., & Womack, S. (2016). Using reliability, validity, and item analysis to evaluate a Teacher-Developed test in international business. Retrieved from <http://www.experiment-resources.com/validity-and-reliability.html>
- Creswell, J. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4<sup>th</sup> ed.). London: SAGE Publications, Inc.
- Crocker, L., & Algina, J. (2008). *Introduction to classical & modern test theory*. Mason, OH: Cengage Learning.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed), *Educational measurement* (2nd ed.) (pp. 443-507). Washington, DC: American Council on Education.
- Cronbach, L. J. (1980a). Validity on parole: How can we go straight? New directions for testing and measurement: Measuring achievement over a decade. *Proceedings of the 1979 ETS Invitational Conference* (pp. 99-108). San Francisco: Jossey-Bass.
- Cronbach, L. J. (1980b). Selection theory for a political world. *Public Personnel Munagement*, 9(1), 37-50.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

- Crooks, T., Kane, M., & Cohen, A. (1996). Threats to the valid use of assessments. *Assessment in Education*, 3, 265-285.
- Cureton, E. E. (1951). Validity. In E. F. Lingquist (Ed.), *Educational measurement*. Washington, DC: American Council on Education.
- Curriculum Research and Development Division (CRDD). (1999). *Investigation into student assessment procedures in public junior secondary schools in Ghana*. Accra: Ghana Education Service.
- Deneen, C., & Boud, D. (2014). Patterns of resistance in managing assessment change. *Assessment & Evaluation in Higher Education*, 39(5), 577-591.
- Denga, I. D. (2003). *Educational Test, Measurement and evaluation*. Calabar: University of Calabar, Printing Press.
- Downing, S. M. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education*, 37(9), 830–837.
- Educational Testing Service. (2005). *Guidelines for constructed-response and other performance assessments*. Princeton, NJ: Author.
- Educational Testing Service. (2007). *ETS international principles for fairness review of assessments*. Princeton, NJ: Author.
- Etsey, Y. K. A. (2004). *Educational measurement and evaluation*. Lecture notes on EPS 203. Unpublished document, University of Cape Coast, Ghana.
- Farrant, J.S., (2000). *Principles and Practice of Education*, (New ed.). London: Longman.
- Fives, H., & DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment, Research & Evaluation*, 18(3), 1-7.

- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8<sup>th</sup> ed.). McGrawHill.
- Griswold, P. A. (1990) Assessing relevance and reliability to improve the quality of teacher-made tests. *NASSP Bulletin*, 76:18-24.
- Gronlund, N. E. (2006). Nature of student assessment. *Assessment of Student Achievement, 8th Edition, Pearson, Boston*, 14-30.
- Gronlund, N. E. (1988). *How to construct achievement tests*. Prentice Hall.
- Gronlund, N.E., and Linn, R.L., (1990). *Measurement and Evaluation in Teaching*. New York: Macmillan.
- Hamman-Tukur, A., & Kamis, A. B. (2000). Content analysis of B.sc Bio-chemistry examination questions implication for testing, teaching and development. *African Journal Research in Education*, 1, 59-62.
- Ing, L. M., Musah, M. B., Al-Hudawi, S. H. V., Tahir, L. M., & Kamil, N. M. (2015). Validity of teacher-made assessment: A table of specification approach. *Asian Social Science*, 11(5), 193-200.
- Kane, M. T. (1994). Validating interpretive arguments for licensure and certification examinations. *Evaluation and the health Professions*, 17, 133-159.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th Ed.). Westport, CT: Praeger Publishers.
- Kellough, R. D., & Kellough, N. G. (1999). *Secondary school teaching: A guide to methods and resources; planning for competence*. Upper Saddle River, New Jersey: Merrill.

- Kinyua, K., & Okunya, L. (2014). Validity and reliability of teacher-made tests: Case study of year 11 physics in Nyahururu District of Kenya. *African Educational Research Journal*, 2(2), 61-71.
- Koksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net. *Turkish Journal of Distance Education*. 5(1), 1-11.
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30, 607-610.
- Kubiszyn, T. & Orich, G. (1984). *Educational testing and measurement: Classroom application and practice*. New Jersey: Scott, Foresman and Company.
- Linn, R.L., and Miller, M.D., (2005). *Measurement and Assessment in Teaching*, (9th ed.). New Jersey: Pearson.
- Malhotra, M. K., & Grover, V. (1998). An assessment of survey research in POM: from constructs to theory. *Journal of Operations Management*, 16(4), 407-425.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrast between testers and users. *Language Testing*, 30(3), 329-344.
- Marmah, A. A., Impraim, A. K. (2013). University lecturer's competence in the construction of multiple choice test items: A case study of Coltek-Kumasi. *The International Journal of Humanities & Social Studies*, 1(4), 1-9.
- McMillan, J. H. (2001). *Classroom assessment: Principles and practice for effective instruction* (2nd ed.). Boston: Allyn & Bacon.

- Mehrens, W. A., & Lehman, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Messick, S. (1975). The standard program: Meaning and values in measurement and evaluation. *American Psychologist*, 30, 955-966.
- Messick, S. (1984). The psychology of educational measurement. *J Educ Measure*, 21, 215-237.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.). *A nation at risk: The imperative for education reform*. Washington, DC: U.S. Government Printing Office.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York: American Council on Education and Macmillan.
- Miller, C. M. I., & Parlett, M. (1974). *Up to the mark: A study of the examination game*. Guildford: Society for Research into Higher Education.
- National Council on Measurement in Education. (1997). *Competency standards for student assessment for educational administrators*. Retrieved December 21, 2019.
- National Education Association. (2003). *Balanced assessment: The key to accountability and improved student learning*. Washington D.C.: National Education Association.
- Nitko, J. A. (2001). *Educational assessment of students*. New Jersey: Prentice Hall.

- Oduro-Kyireh, G. (2008). *Testing practices of senior secondary school teachers in the Ashanti Region of Ghana*. Unpublished master's thesis, University of Cape Coast, Cape Coast, Ghana.
- Oyinka, N. B. (2007). *Measurement and evaluation in education*. Owerri: Springbelt Publishers.
- Pipia, E. (2014). Practical strategies for teachers to enhance the degree of reliability and validity in assessment. *Journal of Education*, 3(1), 47-52.
- Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation*, 38(3-4), 84-92.
- 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.
- Räisänen, M., Tuononen, T., Postareff, L., Hailikari, T., & Virtanen, V. (2016). Students' and teacher's experiences of the validity and reliability of assessment in a bioscience course. *Higher Education Studies*, 6(4), 181-189.
- Reeves, T. D., & Marbach-Ad, G. (2016). Contemporary Test Validity in Theory and Practice: A Primer for Discipline-Based Education Researchers. *CBE—Life Sciences Education*, 15, 1-9.
- Sadler, R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, 34, 807-826.
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., ... & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between

- curriculum and assessment developers. *Applied measurement in education*, 21(4), 295-314.
- Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed). Washington, DC: American Educational Research Association.
- Shillingburg, W. (2016). Understanding validity and reliability in classroom, school-wide, or district-wide assessments to be used in teacher/principal evaluations.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology*, 18, 161-169.
- Spearman, C. (1913). Correlations of sums and differences. *British Journal of Psychology*, 5, 417-426.
- Stiggins, R. J. (2001b). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-15.
- Stiggins, R., & Chappuis, S. (2005). Putting testing in perspective: It's for learning. *Principal Leadership. High School Ed.*, 6(2), 16-20.
- Stringer, N. (2008). *Contemporary validity theory and the assessment context in England*. Centre for Educational Research and Policy. Retrieved from [www.cerp.org.uk](http://www.cerp.org.uk).
- Tamakloe, E. K., Atta, E. T. & Amedahe, F. K. (1996). *Principles and methods of teaching*. Accra: Black Mask Ltd.
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). *SAGE handbook of mixed methods in social and behavioral research* (2nd ed.). Thousand Oaks, CA: Sage.

- Thurstone, L. L. (1955). The criterion problem in personality research. *Educational and Psychological Measurement, 15*, 353–361.
- Trochim, W. M. K. (2015). *Research methods knowledge base*. Retrieved <http://www.atomicdog.com/trochim>
- Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of Operations Management, 16*(4), 361-385.
- Wedman, J. (2017). *Theory and Validity Evidence for a Large-Scale Test for Selection to Higher Education*. Umeå, Sweden: UmU Printing Service, Umeå University.
- Yan, Z. (2016). Student self-assessment practices: The role of gender, school level and goal orientation. *Assessment in Education: Principles, Policy & Practice, 1-17*.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*, 323-342.



## APPENDICES

**APPENDIX A**  
**QUESTIONNAIRE**

Dear Respondent,

This questionnaire seeks to gather information about your experiences and practices in testing as a teacher. The information you provide in this questionnaire is strictly for academic purposes. The information provided will be treated as coming from a group, and for that matter, you will not be associated with any of the responses. Participation in this study is not compulsory. Any information you provide will be kept secretly and confidentially. Please provide responses as frank as possible.

**Instruction:** Please check (√) or write where necessary.

**SECTION A – Demographic Information**

1. Gender
  - a. Male [      ]
  - b. Female [      ]
2. Professional background
  - a. HND/Diploma with education [      ]
  - b. HND/Diploma without education [      ]
  - c. Degree with education [      ]
  - d. Degree without education [      ]
3. Years of teaching experience
  - a. Less than 1 year [      ]
  - b. 1 – 4 years [      ]
  - c. 5 – 8 years [      ]
  - d. 9 – 12 years [      ]
  - e. Above 12 years [      ]

### SECTION B – Validity

Please read carefully the following statements and indicate to the best of your knowledge whether the statements are True or False.

No.	Items	True	False
1.	When directions provided on tests are unclear, it may affect how students respond to the items.		
2.	Test items should contain big vocabularies that can make students think for the meaning of the item.		
3.	Students should be given a very short time within which they will complete a test or assessment.		
4.	Test items developed to assess students should be very difficult, so that the teacher will know the students' ability.		
5.	Results from a test with unclear instructions can be effectively used to make meaningful decisions.		
6.	All learning objectives can be measured using a one type of test.		
7.	Test items should always be based on the learning objectives.		
8.	A test could contain very few items in order to measure the amount of students' learning.		
9.	Items on a test could be arranged in a way that the difficult ones come first.		
10.	Items on test should be arranged in a way that the correct answers form a systematic pattern for easy scoring or marking.		
11.	A short test has limited uses and interpretations.		
12.	Test items should closely examine what has been taught.		
13.	The lighting and ventilation of the testing room can have a significant impact on the performance of students.		

14.	When students cheat in an examination, their test result loses its trustworthiness.		
15.	It is appropriate to conduct examinations when students are emotionally stable.		
16.	Tests like essays can be scored effectively and consistently without necessarily following the marking scheme.		
17.	It is necessary to offer some little assistance to students with low ability during examination.		
18.	In scoring tests, weaker students can be given some scores to uplift their scores.		
19.	It is important to score tests using personal knowledge of each student's past performance.		
20.	Assessment procedures should be free of gender, ethnic, social class, and religious bias and stereotypes.		

### **SECTION C – Reliability**

Please read carefully the following statements and indicate to the best of your knowledge whether the statements are True or False.

<b>No.</b>	<b>Items</b>	<b>True</b>	<b>False</b>
1.	When assessing students on a particular content, it is important to use several forms of assessment.		
2.	It is essential that more time is given to students to complete tasks given to them.		
3.	Tests developed to assess students should contain more items.		
4.	When scoring tests, more formal procedures should be used.		
5.	Scorers should consistently refer to marking scheme.		

6.	More than one person should mark or score each student's test (eg. essay or project), then the average score be used finally.		
7.	It is important to score all responses to a particular question at a sitting without interruption.		
8.	When assessing students, the assessment task should be tailored to each student's ability.		
9.	In assessing students, teachers should use equivalent forms of tests.		
10.	When grading students, teachers should primarily focus on students' scores on various assessments.		
11.	Tests should be administered to students at any time even without prior notice.		
12.	In scoring students' test, teachers should be very hard on students who miss classes and be generous to students who are regular in class.		
13.	Test scores are said to be reliable when they accurately reflect the content taught.		
14.	Adding more items to the test increases the reliability of the test results.		
15.	When test results are reliable, they are also valid.		
16.	The scoring of essay should be done script by script, but not item by item.		
17.	Essay tests should be scored when the marker is physically and mentally alert.		
18.	Previously scored items should not be kept out of sight when scoring the rest of the items.		
19.	When making decisions about students' learning, it is important to use a combination of results from different assessment methods rather than a single assessment result.		
20.	Teachers should select assessment tasks that differentiate best students from the least able students.		

**SECTION D – Practices that Enhance Validity of Assessment Results**

Please read carefully the following statements and indicate how frequently you engage in each of the following practices, using the following scale: **N – Never; O – Often; VO – Very often; A - Always.**

No.	Practices	N	O	VO	A
1.	I provide clear instructions on how to answer test items.				
2.	Test items I develop do not contain big vocabularies that can hinder the meaning of the item.				
3.	I give my students a very short time to answer tests that I administer to them.				
4.	I use difficult test items to assess students, so that I will know the students' ability.				
5.	I write test items in a way that they do not give students clues to the answer.				
6.	I score students' tests using personal knowledge of each student's past performance.				
7.	When assessing students, I make sure the assessment tasks are tailored to each student's ability.				
8.	I use any type of test to assess students, irrespective of the learning objectives.				
9.	I develop test items based on the learning objectives.				
10.	The tests I develop to assess students contain more items.				

11.	I arrange test items in order of difficulty, with the difficult ones first.				
12.	I arrange options of test items in a way that the correct answers form a systematic pattern for easy scoring or marking.				
13.	When scoring, I score all responses to a particular question at a sitting without interruption.				
14.	The tests I develop closely examine what has been taught.				
15.	I ensure proper room lighting and ventilation when students are being assessed.				
16.	I allow students to help themselves during examinations.				
17.	I conduct examinations when students are emotionally stable.				
18.	I score tests like essays by consistently following the marking scheme.				
19.	I offer some little assistance to students with low ability during examinations.				
20.	When scoring tests, I give weaker students some scores to uplift their scores.				
21.	The assessment procedures I use are free from gender, ethnic, social class, and religious bias and stereotypes.				
22.	I use several forms of assessment when assessing students on a particular content.				
23.	When scoring students' scripts test (eg. essay or project), I use more than one person (marker), then the average score be used finally.				
24.	In assessing students, I use equivalent forms of the tests.				
25.	When grading students, I primarily focus on students' scores on various assessments.				

26.	I administer tests to students at any time, even without prior notice.				
27.	When scoring students' tests, I am very hard on students who miss classes and generous to students who are regular in class.				
28.	When scoring essay, I do it script by script, but not item by item.				
29.	I score essay tests when I am physically and mentally alert.				
30.	I keep previously scored items out of sight when scoring the rest of the items.				
31.	When making decisions about students' learning, I use a combination of results from different assessment methods rather than a single assessment result.				



## APPENDIX B

### OBSERVATION CHECKLIST

Name of School:.....

Subject:.....

CONDITIONS	JUDGMENT	REMARKS
<b>A. Room conditions</b>		
1. Arrangement of tables and chairs.	Appropriate [ ] Not appropriate [ ]	
2. Ventilation.	Good [ ] Poor [ ]	
3. Lighting.	Good [ ] Poor [ ]	
4. Sitting posture of examinees.	Appropriate [ ] Not appropriate [ ]	
5. Serenity of environment.	Good [ ] Poor [ ]	
<b>B. Invigilation</b>		
1. Number of invigilators per room.	Adequate [ ] Not adequate [ ]	
2. Invigilators making phone calls, reading, chatting, etc.	Yes [ ] No [ ]	
3. Invigilators interrupting students.	Yes [ ] No [ ]	
4. Announcements about time.	Frequently [ ] Not frequently [ ] Unnecessarily [ ]	
5. Invigilators intermittently walking around the room during testing.	Frequently [ ] Not frequently [ ]	
6. Invigilators pestering students.	Yes [ ] No [ ]	
7. Position of invigilators.	Appropriate [ ] Not appropriate [ ]	
<b>C. Testing conditions</b>		
1. Time allowed for the tests.	Appropriate [ ] Not appropriate [ ]	
2. Start of test.	Exactly on scheduled time [ ] Not exactly on scheduled time [ ]	

3. Stopping time.	Exact time [ ] Not exact time [ ]	
4. Availability of supplementary answer booklets.	Available and adequate [ ] Available but not adequate [ ] Not available [ ]	
5. Supplementary question papers.	Available [ ] Not available [ ]	
6. Emergency medical care	Available [ ] Not available [ ]	

## APPENDIX C

### ENGLISH LANGUAGE TEST SAMPLES

END OF SECOND SEMESTER EXAMINATION - AUGUST, 2019  
ENGLISH LANGUAGE- FORM ONE

DURATION: 2 HOURS

PAPER TWO  
ESSAY

TIME: 1 HOUR 10 MINS.

Answer only one question. Your essay should not be more than 350 words.

1. Your former classmate who recently travelled to further his/her studies abroad has heard about the new educational system in Ghana, write a letter giving three reasons why this new system is helpful.
2. As part of the school extra-curricular activities, your school organized an inter-house sporting competition. Describe what went on.
3. Write a story to illustrate the saying: The devil makes work for idle hands *It helps you*

#### READING COMPREHENSION

Read the passage carefully and answer the questions on it.

'It was not for nothing that I left the Ghana News Agency (GNA) to do full-time farming. I realized that I was born a farmer. That was why, when though I only had a meagre initial capital, I have been able to achieve my life's ambition'.

These inspiring words tumbled from the lips of Nana Gyedu, the newly crowned National Best Farmer. After completing tertiary education in 1998, he sought employment with the Ghana News Agency. Four years later, with the unquenchable controlling desire to enter farming still burning strongly in him, Nana took to farming alongside his work at the GNA.

'I bought ten acres of land for twenty thousand cedis and started a poultry farm with sixty laying birds', Nana said with a grin. But what forced Nana into voluntary retirement from his paid job at the youthful age of thirty-four? A farm assistant Nana had employed to take care of the farm accidentally poisoned fifty birds on the farm. This nearly disorganized him and almost derailed his hopes and plans. Nana said, with this setback, his thoughts began racing in his head. He finally decided to leave GNA and concentrated fully on farming.

Today, Nana has a total of one hundred and fifty acres of farmland comprising a palm plantation, cocoyam, cassava and sugar-cane farms. In addition, he has ten thousand birds, a piggery of fifty boars and sows, twenty beehives, twenty cattle, a hundred turkeys and two fish ponds.

In 2003, he was selected the third National Best Farmer. He redoubled his efforts and after seven years of hard work and meticulous work, Nana Gyedu eventually landed at the ultimate - his life's ambition - the 2010 National Best Farmer.

The strange thing about Nana's solid achievement is that he never took any bank loan nor received help from anybody. His only inspiration has been his wife and five children who have staunchly stood by him all these years. Nana Gyedu's edifying advice to the youth is that they should go into farming to save themselves from the problems of unemployment. He urges them not to bother themselves so much about bank loans which are not only difficult to come by but could also be risky.

A

Questions

1. What work did Nana engage alongside his regular career at the Ghana News Agency?
2. What was Nana's life ambition?
3. For what two reasons did Nana Gyedu advise the youth against taking bank loans?
4. What incident nearly robbed Nana of his life ambition?
5. "... his thought began racing in his head".  
What literary device has been used in the expression above?
6. "... the newly crowned National Best Farmers".
  - i. What is the grammatical name of the underlined?
  - ii. What is its grammatical function?
7. For each of the following words, use a word or phrase that means the same and can replace it in the passage.
  - i. Meagre
  - ii. Ambition
  - iii. Unquenchable
  - iv. Accidentally
  - v. Edifying

OBJECTIVES

TIME: 50 MINUTES

From the group of letters A to D, choose the one that is *nearly opposite* in meaning to the underlined word in each sentence.

1. The motorist said the splash of the water was \_\_\_\_\_ not deliberate.
  - a. accidental
  - b. incidental
  - c. avoidable
  - d. intentional
2. The computerized services of the new bank contrast with the \_\_\_\_\_ operations of the old ones.
  - a. mechanical
  - b. condensed
  - c. manual
  - d. electrical
3. He denied that he ever \_\_\_\_\_ committing the offense in his earlier statement.
  - a. argued
  - b. colluded
  - c. emphasized
  - d. admitted
4. Tunde says he always looks sober because there is nothing to be \_\_\_\_\_ about.
  - a. excited
  - b. happy
  - c. witty
  - d. serious
5. While the first battalion halted its march at night falls, the second \_\_\_\_\_ to the barracks.
  - a. set out
  - b. proceeded
  - c. turned up
  - d. returned
6. His father is an \_\_\_\_\_ driver but he is only a novice.
  - a. enlightened
  - b. experienced
  - c. aged
  - d. returned
7. The difference between Alariba and Awuni is that while she is industrious Awuni is \_\_\_\_\_.
  - a. slow
  - b. indolent
  - c. inactive
  - d. carefree
8. Instead of shunning Dufie, we should rather find a way of \_\_\_\_\_ her.
  - a. praising
  - b. honouring
  - c. encouraging
  - d. accepting

A

9. The manager wants to do the best with the company's scarce resources.
- a. many
  - b. beautiful
  - c. important
  - d. abundant
10. The police officer explained that pedestrians are also road users.
- a. beggars
  - b. settlers
  - c. motorists
  - d. athlete

Choose from the alternatives lettered A to D the one which is *nearest in meaning* to the underlined word or expression in each sentence.

11. How come Sebastian is so ill-mannered?
- a. arrogant
  - b. bad-tempered
  - c. rude
  - d. rough
12. What a lame excuse for arriving so late!
- a. tenable
  - b. plausible
  - c. cheap
  - d. weak
13. Elorm hit Aku intentionally.
- a. deliberately
  - b. seriously
  - c. harshly
  - d. repeatedly
14. The unforgettable learning experience has fortified my determination.
- a. mounted
  - b. promoted
  - c. boosted
  - d. assisted
15. Writing became the focal point of his life after his retirement.
- a. singular
  - b. central
  - c. actual
  - d. chosen
16. There is an almost endless supply of knowledge on the internet.
- a. limitless
  - b. unchecked
  - c. matchless
  - d. unbroken
17. Both meetings were held simultaneously.
- a. indoors
  - b. concurrently
  - c. systematically
  - d. lately
18. Government is bent on doing something about the fragile economy.
- a. booming
  - b. falling
  - c. weak
  - d. new
19. Convincing our opponents to agree with us is an arduous task.
- a. easy
  - b. difficult
  - c. unnecessary
  - d. impossible
20. After his fall, the general was posted to a remote township.
- a. distant
  - b. great
  - c. beautiful
  - d. near

After each of the following sentences, a list of possible interpretations of all or part of the sentence is given. Choose the interpretation you consider *most appropriate*.

21. After the third day, the mason realized that in accepting the job, he bit more than he could chew. This means
- a. he has to eat more food to do the job
  - b. accepting the job was a nice idea
  - c. he ate too much food when he accepted the job
  - d. the was too difficult for him

3

A

22. The hunter was admonished to pull himself together when he was highly infuriated by the woman. This means that he should
- shoot the woman
  - ponder over his action
  - control himself
  - contract himself
23. The linguist left no stone unturned when he was looking for his staff. This means he
- did everything possible in order to find his staff
  - looked under every stone in order to find his staff
  - refused to ask anybody about his staff
  - did not turned any stone when looking for his staff
24. I only got the message at the eleventh hour. This means
- the message took eleven hours
  - the message came at eleven o'clock
  - the message came almost too late
  - it was the only message received
25. The athlete started well but he fagged out before he reached his destination. This means that he was.....
- overtaken by his opponents
  - pushed out of the lane
  - exhausted towards the end
  - disqualified by the officials
26. A Christian won't change his lifestyle just to move with people. This means that he
- doesn't do what most people do
  - doesn't associate with others
  - doesn't love his friends
  - is not satisfied with life
27. Sampson is at daggers drawn with Osei. This means that Sampson
- uses daggers on Osei
  - quarrels with Osei everyday
  - is going to fight Osei
  - loves Osei
28. Emelia was beside herself with joy when she had distinction in the exams. This means that she was...
- disturbed
  - excited
  - mad
  - proud
29. In a democratic country, every citizen has the right to air their views. This means citizens could
- vote for their president anytime
  - insult each other freely
  - freely express their opinion
  - attend any social gathering
30. Growing up, my father advised me to call a spade a spade. This means he
- told me to learn the name of a spade
  - told me to speak the truth
  - told me to speak well
  - thought I didn't know a spade

From the alternatives lettered A to D, choose the one that best completes the sentences below

31. We should press \_\_\_\_\_ to victory.
- up
  - on
  - in
  - over
32. \_\_\_\_\_ have been friends for years.
- he and her
  - him and I
  - I and her
  - she and I
33. Prince is no more a friend of \_\_\_\_\_
- them
  - him
  - their
  - ours
34. How dare you \_\_\_\_\_ to me like that!
- to speak
  - spoke
  - speak
  - speaking
35. Kofi \_\_\_\_\_ all the day and now he is sick.
- has been studying
  - having studied
  - is studying
  - has studied
36. The man is accused of putting \_\_\_\_\_ an air of superiority.
- about
  - in
  - on
  - away

4

A

37. \_\_\_\_\_ all the contestants, the judges then announced the winner.
- to see
  - seeing
  - seen
  - having seen
38. \_\_\_\_\_ I joined the class late, I could pass the examination.
- since
  - as
  - despite
  - although
39. All your \_\_\_\_\_ like you very much.
- brother-in-law
  - brothers-in-law
  - brother-in-laws
  - brothers-in-laws
40. Carol wanted to drive to Accra \_\_\_\_\_ Ben insisted that they fly.
- but
  - and
  - since
  - or
41. The factory had to \_\_\_\_\_ many workers because of the fall in production.
- lay down
  - put away
  - lay off
  - put off
42. \_\_\_\_\_ you dislike the girl, you shouldn't beat her.
- since
  - as
  - even if
  - of course
43. \_\_\_\_\_ hearing the news, he jumped high for joy.
- over
  - with
  - in
  - on
44. Kofi will not meet his father at home if he \_\_\_\_\_ not here by 5:00pm.
- is
  - were
  - was
  - isn't
45. The handsome gentleman died \_\_\_\_\_ AIDS.
- at
  - with
  - of
  - through
46. It is every parent's wish that their children confide \_\_\_\_\_ them.
- with
  - in
  - on
  - at
47. The headmaster \_\_\_\_\_ to talk to all students after school.
- want
  - is wanting
  - wants
  - wanting
48. You know very well that it was no fault of \_\_\_\_\_.
- my
  - I
  - me
  - mine
49. I won't be able to attend the press conference because I am worn \_\_\_\_\_.
- on
  - through
  - down
  - out
50. I found out recently that one can enjoy \_\_\_\_\_ without drinking.
- myself
  - oneself
  - yourself
  - herself

In the short passage below the numbered gaps indicate missing words. Against each number in the list below the passage, four choices are offered in columns lettered A to D. For each numbered gap, choose from the options provided for that number, the word that is *most suitable* to fill the gap.

A

Our school excelled in sports last year. We 51 in most of the 52 when we met at the state's annual 53 games for secondary school. We collected the gold 54 in relays, sprints, high 55 and pole vaults.

- | A                | B            | C         | D           |
|------------------|--------------|-----------|-------------|
| 51. won          | lost         | equalled  | came in     |
| 52. plays        | occasions    | struggles | events      |
| 53. entertainers | championship | comics    | competition |
| 54. medals       | cups         | trophies  | hats        |
| 55. shot         | race         | jump      | hurdles     |

THERESA ENNIN: MAKOLA

56. "Move out of the way, move out, I say"  
The line above shows the cart pusher's
- a. delight
  - b. sadness
  - c. agitation
  - d. fear
57. "He refuses to be soothed and gives out a loud yell", a character in the poem does the above out of
- a. warmth
  - b. happiness and pressure
  - c. disappointment and agitation
  - d. hunger and tiredness
58. "The sweat runs down his face..."  
What literary device is used in the above line?
- a. metaphor
  - b. simile
  - c. personification
  - d. alliteration
59. Which group of people did the poem state as behaving heedlessly
- a. The crowd in the way of the cart pusher
  - b. The children playing "catch-me-if-you-can"
  - c. The passengers the drivers' mates were calling out
  - d. The tiny girl and her baby
60. "The runny-nose baby at her back is supported with a faded ATL cloth"  
What image does 'a faded ATL cloth' convey?
- a. A beautiful traditional cover
  - b. The need to patronize made in Ghana products
  - c. A reflection of a local art work
  - d. Scarcity and poverty

6



B

**END OF SECOND SEMESTER EXAMINATION – DEC 2020**  
**ENGLISH LANGUAGE**

SHS 2

3 Hours

Answer **three** questions in all: **one** question from section A and all the questions in sections B and C

**SECTION A (Essay)**  
{ 50 Marks }

Answer **one** question **only** from this section.

All questions carry equal marks. Your answer should not be less than 450 words.  
You are advised to spend about 50 minutes on this section.

1. As part of a campaign to increase the number of students in your school, your Headmaster has asked you as the school prefect to convince a gathering of junior high school pupils in your district to choose your school. Write the speech you would deliver.
2. You are the main speaker in an inter school debate on the motion: *Teachers should be made to answer for students' failure.* Write your arguments **for** or **against** the motion.
3. You have received information that your brother who is schooling in another part of the country is playing truancy and keeping bad company. Write a letter warning him to turn over a new leaf.
4. As a youth leader in your community, write a letter to the Minister of Youth and Sports identifying with reasons three areas where you think the government should concentrate efforts to bring about rapid development to the country.
5. Write a story to illustrate the saying: *Make hay while the sun shines.*

**COMPREHENSION**  
(20 Marks)

Read the following passage carefully and answer all the questions on it

When I was a little boy, I had been playing in the street of Kumasi one day with other boys about my own age. Full of cheerful bounce, we went scampering past the hospital right opposite the hospital. On the other side of the road, a convict was pulling down some clay - brick buildings under military supervision. The buildings were badly constructed too narrow and far too high - and were being demolished for their materials. It was a dangerous job for workers

B

who lacked both experience and proper equipment – country boys, most of them, who might have been gaoled only the week before.

They were using heavy boards of timber as battering-rams. Since they had no ladders, they attacked the walls at ground level knocking hole after hole through the clay till the entire wall came crashing down. We found them busy demolishing a building even taller and narrower than the rest. As soon as I took in the spectacle of these young boys toiling away with their rams against that puddled clay wall, I was struck all of a heap; I stood there, petrified, rooted to the spot. In this instant-despite the tumbled ruins on either side, which suggested that the prisoners had already demolished several other houses unharmed-I knew that something terrible was going to happen, something so serious that I would never forget it if I saw it. My friends first tried to drag me away, then gave up and stayed there with me.

As the work went on, a horrible impalpable fear rose and spread throughout my entire being, it was akin to the feeling you get when you are reading a detective story written by a cold, sadistic author, expert at twisting the last drop of suspense and horror out of his material. The walls were intact on the upper floors, but grew steadily weaker at their foundations. The beaten earth flaked away, revealing a core of puddled clay bricks. The whole building looked like a giant who was having the flesh torn from his bones leaving only a skeleton behind. Suddenly the guard shouted to the prisoners; "look out! Get away quickly!" The prisoners darted away into the court yard, bouncing like bullfrogs. It all happened in a flash. The wall came down with a dull roar and a crash; and instantly we heard an indescribable shriek. One of the men had not gotten clear in time.

What sticks in my memory, what I shall never forget as I live especially when I witness any accident or any more than moment, any demonstration of unhappiness - is that shriek of a trapped man, calling out not so much for help as in protest against the whole ghastly injustice of life? That cry of mingled agony and defiance will always ring in my ears.

a. For each of the following words underlined in the passage, give another word or phrase that means the same and which can replace it in the passage.

- i. demolished
- ii. gaoled
- iii. a horrible
- iv. a cold
- v. expert

b. What was the feeling of the boys as they played in the streets?

c. Give one reason why the buildings were being pulled down?

d. State two reasons why according to the passage the job was dangerous.

e. Using two adjectives, describe the feeling of the writer when he saw the men destroy the building.

f. What happened when the building collapsed?

g. Full of cheerful bounce.....

B

- i. What is the grammatical name given to this construction?
- ii. What is its function?
  
- h. *the whole building looked like a giant* .....  
What figure of speech has been used in the above expression?

SUMMARY  
(30 marks)

Read the following passage carefully and answer all the questions on it

It is a well – established fact that some students perform creditably in school while others flop badly. Some students may be intelligent and highly motivated but still find it extremely difficult to climb the academic ladder. Such students face learning problems but may not necessarily have physical disabilities.

Among these problems, perhaps the most worrying is the bad attitude these students have towards their studies. This arises because wrong views have distorted the benefits of education. Many students actually doubt their own abilities to build a bright and secure future for themselves through education. Such students have not understood the aims and objectives of education to appreciate its numerous character and polished manners. Closely linked to the above problem is that students also have poor attitude to teachers, maybe because the teachers do not employ correct methods to make their teaching interesting.

Indeed, a more serious problem is that students sometimes find it difficult to grasp abstract ideas, especially in science and Mathematics. Reading the correct spelling of words and remembering of facts are all connected with this problem. Such students will do well not to fall into the trap of using drugs that can supposedly aid memory. All that drugs can do is to wreak havoc on the body.

Adequate and effective measures can always be taken to deal with learning problems in school. Prominent among these remedial measures is repetition of ideas or points which can aid the students in remembering. Repeating words can engage the senses of sight and hearing and leave deep impressions on the mind. If students learn Mathematical formulae and theorems, or scientific principles and theories by repeating them, it will aid them greatly.

A more practical approach for students to be determined to overcome their learning problems. Teachers have the know – how to keep students focused in their determination to better themselves. Since teachers are in touch daily with the students, they get to know them more intimately than their parents and are therefore in the best position to help them.

- a) In **three** sentences, **one** for **each**, state the learning problems students face in school.
- b) In **three** sentences, **one** for **each**, state the possible solutions for learning problems.







30. Eloquent politicians usually attract great crowds to their rallies.
- |                |               |
|----------------|---------------|
| a) Progressive | b) Pragmatic  |
| c) Charming    | d) Articulate |

## SECTION IV

After each of the following sentences, a list of possible interpretations of all or part of the sentence is given. Choose the interpretation you consider **most appropriate** for each sentence.

31. Halidu supported the team through thick and thin. This means he supported the team
- |                                       |                            |
|---------------------------------------|----------------------------|
| a) on and off.                        | b) as situations demanded. |
| c) even when there were difficulties. | d) to some extent.         |
32. When Anima received her examination result, she told her colleagues that the game was worth the candle. This means that
- |   |                              |
|---|------------------------------|
| a) the result did not reflect the effort. | b) the result was very good. |
| c) the examination was difficult.         | d) she failed totally.       |
33. Although prudent measures were put in place to salvage the economy, it still remained in the doldrums. This means that the economy
- |  |                                  |
|--|----------------------------------|
| a) lacked clear direction.               | b) performed beyond expectation. |
| c) depended on the importation of drums. | d) lacked any improvement.       |
34. I shall be with you in a jiffy. This means that I shall be with you
- |                          |                   |
|--------------------------|-------------------|
| a) at some future time.  | b) in due course. |
| c) as and when possible. | d) in a moment.   |
35. The mountain climbers were tired to the very marrow of their bones. This means they
- |                     |                    |
|---------------------|--------------------|
| a) were exercising. | b) slowed down.    |
| c) were fainting.   | d) were exhausted. |
36. There is danger looming on the horizon. This means that danger is
- |                |               |
|----------------|---------------|
| a) everywhere. | b) imaginary. |
| c) imminent.   | d) curtailed. |
37. I need money to keep body and soul together. This means that I need money to
- |             |              |
|-------------|--------------|
| a) grow.    | b) flourish. |
| c) succeed. | d) survive.  |
38. The minister's remarks were outrageous and clearly beyond the pale. This means that the remarks were
- |                             |                             |
|-----------------------------|-----------------------------|
| a) considered unacceptable. | b) beyond understanding.    |
| c) not focused.             | d) delivered too hurriedly. |
39. Homo's music mirrors the feelings of optimism in the country. This means that music
- |                           |                            |
|---------------------------|----------------------------|
| a) touches hearts.        | b) is worthy of emulation. |
| c) contains modern ideas. | d) reflects hope.          |

40. Adjoa Akoto was admonished to paddle her own canoe. This means that Adjoa Akoto should
- a) be in charge of the business.                      b) place her interests first.  
c) manage her own affairs.                                d) be more serious with life.

## SECTION V

In the following passage the numbered gaps indicate missing words. Against each number in the list below the passage, four choices are offered in columns lettered A to D. For each numbered gap, choose from the options provided for that number, the word that is **most suitable** to fill the gap.

The science of medicine has brought many benefits to man, the most important of which is the increase in ... 41 ... This highly desirable feat has resulted from the fact that certain diseases which defied all ... 42 ... in the past have been brought under ..... 43 ... Sensitive clinical instruments make accurate ..... 44 ..... possible. Powerful substances of ..... 45 ... value have been manufactured which bring relief to the sick and the suffering.

When one is sick, the best place to seek advice is a health centre where medical .... 46 ..... try to restore the body to good health. A good number of patients are ..... 47 ..... for a period of medication and observation. When their health condition improves they are ... 48 ..... from the hospital. Really, some people respond to ..... 49 ..... much more quickly than others. Those who are pronounced well are advised to report regularly for ..... 50..... This is sometimes cumbersome for those who live far away from health centres. It is the hope of all that the time will come when health care facilities will be accessible to all.

	A	B	C	D
41.	life expectancy	life cycle	life time	life style
42.	solutions	measures	remedy	cure
43.	care	control	authority	suppression
44.	diagnosis	assessment	discovery	investigation
45.	herbal	potent	medical	medicinal
46.	dispensers	personnel	operators	attendants
47.	restrained	kept	admitted	delayed
48.	relieved	dismissed	discharged	released
49.	medicine	treatment	care	prescription
50.	investigation	research	review	examination

## PROSE

AMA ATA AIDOO: *No Sweetness Here*

51. Tanchiw is an example of ...
- a. malapropism    b. anaphora  
c. imagery    d. metaphor



Read the following extract and answer questions 52 and 53

*He was beautiful, but that was not important. Beauty does not play such a vital role in a man's life as in a woman's, or so people think... Only an immodest girl like me would.....*  
Page 66

52. The underlined part of the extract refers to...
- |                     |                           |
|---------------------|---------------------------|
| a. Maami Ama        | b. Chicha                 |
| c. Kojo Fi's sister | d. Maami Ama's elder aunt |

53. Whose beauty is the speaker talking about?
- |              |               |
|--------------|---------------|
| a. Kojo Fi's | b. Nana Kum's |
| c. Kwesi's   | d. Old Kofi's |

*O you people, have you not heard of the German-people?*

54. Who is the speaker in the above statement?
- |                  |              |
|------------------|--------------|
| a. Hawa's father | b. Mma Asana |
| c. Issah         | d. Memunat   |

55. What commodity does Mma Asana goes to sell at the market?
- |              |                |
|--------------|----------------|
| a. Cow meat  | b. Smoked fish |
| c. Kola nuts | d. Tuo         |

56. Mma Asana says Issa leaves for the south in the night because...
- |  |
|--|
| a. he is hiding from Hawa                              |
| b. he does not want anyone to know about his departure |
| c. Fuseni was sick                                     |
| d. he wants to catch the government bus                |

*..... But this time I thought he was going too far. I rushed to rescue... and then it came wham!*  
Page 101

57. The underlined word in the extract above is an example of...
- |             |                 |
|-------------|-----------------|
| a. irony    | b. onomatopoeia |
| c. metaphor | d. hyperbole    |

58. Papa Kramo is also known as...
- |             |             |
|-------------|-------------|
| a. Mohammed | b. Marabout |
| c. Mallam   | d. Sheik    |

Read the extract below and answer questions 59 and 60.

*Please little sister. I am not trying to interfere in your private life. You said yourself that you want a man of your own. That man belongs to so many women already...*  
Page 109

59. Who is speaking and to whom?  
 a. suzzie to Connie  
 c. James to Mercy  
 b. Connie to Mercy  
 d. Connie to Suzzie
60. Who is the man in question?  
 A. James Arthur  
 c. Mensar- Arthur  
 b. Joe  
 d. Captain Ashley

DRAMA

YAW ASARE: Ananse in the Land of Idiots

61. The drama can be classified as...  
 a. comedy  
 c. farce  
 b. tragedy  
 d. satire
62. What punishment does the priestess give to Ananse?  
 a. Imprisonment  
 c. caning  
 b. Banishment  
 d. death
63. By granting Ananse's request, King Dosey shows that he is.....  
 a. wise  
 c. gullible  
 b. kind  
 d. brave

*You go to sleep snoring like a pregnant pig.* pg 30

64. This is an example of... a.  
 a. personification  
 c. an apostrophe  
 b. a hyperbole  
 d. a simile

Read the following extract and answer questions 65 to 68

*Simple. That's where an active brain comes handy. Are you ready to accept the bargain? Think of yourself with half a chieftain- an autonomous empire with pretty maidens...think of all these pleasures.*  
 Page 31

65. Who made this statement?  
 a. Akpala  
 c. Mbisila  
 b. Ananse  
 d. Odudu
66. Who was the addressee?  
 a. Ananse  
 c. Potagyiri  
 b. Odudu  
 d. Akpala
67. The underlined part of the extract refers to...  
 a. dispatching Odudu to the underworld  
 c. sending the princess away  
 b. posing as Potagyiri  
 d. killing Potagyiri

68. The extract reveals that the speaker is....
- |              |            |
|--------------|------------|
| a. brave     | b. cunning |
| c. confident | d. concern |
69. One of the major themes in the play is....
- |                  |               |
|------------------|---------------|
| a. compromise    | b. corruption |
| c. individualism | d. wisdom     |
70. Which of the following words aptly describes Ananse?
- |               |             |
|---------------|-------------|
| a. wise       | b. crafty   |
| c. forthright | d. tolerant |

POETRY  
KWESI BREW: *Lest We Should Be the Last*

71. The attitude of the personae in the poem is...
- |                 |               |
|-----------------|---------------|
| a. indifference | b. disrespect |
| c. reverence    | d. anger      |
72. What did the persona and his people hope to receive?
- |                        |                   |
|------------------------|-------------------|
| a. a better livelihood | b. a fertile land |
| c. new homes           | d. fresh bans     |
73. "*Our hunger should be banished*" illustrates...
- |                    |             |
|--------------------|-------------|
| a. simile          | b. anaphora |
| c. personification | d. anagram  |
- ...*We left our corn in the barn*  
... *knowing that in your presence*  
*Our hunger would be banished...*
74. This extract is an example of....
- |            |               |
|------------|---------------|
| a. irony   | b. apostrophe |
| c. litotes | d. aphorism   |
75. .... is a major theme in the poem.
- |          |               |
|----------|---------------|
| a. Faith | b. corruption |
| c. Greed | d. Wealth     |
76. What does the image of the dog symbolizes....
- |                   |                             |
|-------------------|-----------------------------|
| a. a loyal friend | b. a victim of exploitation |
| c. a wild animal  | d. a greedy person          |

B

77. Which of the following is not a theme in the poem?
- |              |            |
|--------------|------------|
| a. injustice | b. revenge |
| c. love      | d. culture |
78. What does the phrase *perpetual scar* refer to?
- |                              |                               |
|------------------------------|-------------------------------|
| a. marriage in youthful days | b. wounds that won't go away  |
| c. resisting being put away  | d. wounds from the dog's bite |
- "... from cottage door to palace gate..."
79. The underlined portions of the above expression refer to...
- |                               |                            |
|-------------------------------|----------------------------|
| a. the sick and the healthy   | b. the weak and the strong |
| c. the ugly and the beautiful | d. the rich and the poor   |
80. The poem ends on the note of...
- |                   |          |
|-------------------|----------|
| a. joy            | b. peace |
| c. disappointment | d. hope  |

**APPENDIX D**

**MATHEMATICS TEST SAMPLES**

DECEMBER, 2020  
CORE MATHEMATICS  
FORM 2  
3 hrs

**1&2**

Name.....

Class.....

**END OF THE SECOND SEMESTER EXAMINATION  
CORE MATHEMATICS SHS 2**

**DEC, 2020**

SECTION A & B [100 marks]

**3hrs**

*Write your name and class in ink in the spaces provided above.*

*Answer forty objectives in section A & seven questions in section B. All the questions in Section B, Part 1 and three questions from Section B, Part 2.*

*In each question, all necessary details of working, including rough work, must be shown with the answer.*

*Give answers as accurately as data and tables allow.*

*Graph papers are provided for use in the examination.*

*The use of non – programmable, silent and cordless calculator is allowed.*

I

27777.7778  
222222.2222

SECTION B  
PART 1

[24 Marks]

1: 45 mins

Answer all the questions in this part. All questions carry equal marks.

1. (a) Without using calculator or tables, simplify:  $\frac{3}{7} \div 1\frac{1}{7} \times \left(2\frac{2}{7} - 1\frac{17}{63}\right)$   
(b) Simplify  $\frac{Q^{1/2} \times Q^{-3/4}}{Q^{-1/2} \div Q^{-1/4}}$
2. (a) Oboro and Ollie in Mathematics Dept. of Sekondi College formed a company and agreed that their annual profit will be shared in the ratio 5: 4 respectively. If at the end of the year, Oboro received Gh¢ 5,000.00 more than Ollie, how much was Ollie's share?  
(b) If  $20x = 16$ , find the value of  $x$ .
3. (a) Find three consecutive odd numbers such that the sum of the last two is 15 less than 5 times the first  
(b) The cost ( $c$ ) of producing number ( $n$ ) of bricks is given by  $c = 828 + \frac{2}{5}n$ , find the cost if the number of bricks is 5.
4. (a) Simplify  $\frac{3a-b}{ab} - \frac{2a+3b}{2ab} + \frac{1}{2}$   
(b) Evaluate without using tables  $\sqrt{27} + \sqrt{147} - 2\sqrt{75} + \sqrt{432}$ , leaving your answer in surd form

27777

500000  
4

3,915

PART II

[36 Marks]

Answer three questions only from this part. All questions carry equal marks.

5. (a) Copy and complete the table of values for the relation  $y = 2x^2 - 7x - 3$  for  $-2 \leq x \leq 5$

x	-2	-1	0	1	2	3	4	5
y	19				-9			

- (b) Using 2cm to 1 unit on x-axis and 2cm to 5 units on the y-axis, draw the graph of  $y = 2x^2 - 7x - 3$  for  $-2 \leq x \leq 5$ .

- (c) From your graph, find:

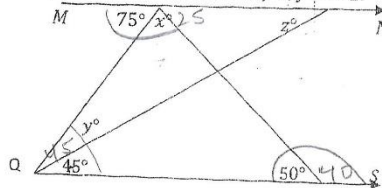
- (i) the least value of y and state the value of x it occurs
- (ii) the range of values of x for which  $2x^2 - 7x - 3$  increases
- (iii) the values of x for which  $2x^2 - 7x - 3 = x + 4$

6. (a) The table shows the mark distribution of sixty students in a test.

Marks	10	11	12	13	14	15
Frequency	2	5	20	23	y	3

Find (i) the value of y (ii) the mean, correct to one decimal place.

- (b) In the diagram,  $MN \parallel QS$ . Find the values of x, y and z.



Marks	Frequency	
10	2	20
11	5	55
12	20	240
13	23	320
14		
15		

I

7. (a) (i) Copy and complete the table of values for the relation  $y = 7 \cos x - 3 \sin x$  in one decimal place. 0.992

x	0°	30°	60°	90°	120°	150°
y	7.0			-3.0		

- (ii) Using a scale of 2cm to 30° on the x-axis and a scale of 2cm to 2 units on the y-axis draw the graph of  $y = 7 \cos x - 3 \sin x$  for  $0^\circ \leq x \leq 150^\circ$ . 6.99

- (iii) Use the graph to solve  $7 \cos x = 3.2 + 3 \sin x$  6.988

- (b) Given that  $\tan \theta = \frac{5}{12}$  where  $0^\circ \leq \theta \leq 90^\circ$ , evaluate  $\cos^2 \theta + \sin^2 \theta$

$$\frac{12}{24} + \frac{5}{25} = 1$$

$$1 + 0 = 1$$

8. (a) Kulio, Mildred and Esther entered into a business partnership. Their capital contributions were Gh₵ 8,000.00, Gh₵ 4,000.00 and Gh₵ 12,000.00 respectively. They agreed to share the profit according to the ratio of their contributions to the capital. If the business made a profit of Gh₵ 6,000.00, find the amount each receives from the profit.

- (b) Solve the equation  $\frac{64^x \cdot 2}{16^{1-x}} = 4^{2x}$  64

- (c) Factorize  $x^2 + ax - x - a$

9. (a) Without using calculator or tables, simplify:  $\frac{1}{2} \log_{10} \frac{25}{4} - 2 \log_{10} \frac{4}{5} + \log_{10} \frac{320}{125}$

- (b) R varies inversely as the third power of P. If  $R = 4$  when  $P = 3$ , find P when  $R = 1.6875$

- (c) Find the values of a and b in the equation  $\binom{a+3}{2} - \binom{b}{a+b} = \binom{2}{-1}$

$$\frac{80}{24} \times 60 = 200$$

$$\frac{40}{24} \times 60 = 100$$

$$\frac{120}{24} \times 60 = 300$$



I

SECTION A  
OBJECTIVE TEST  
[40marks]

1 : 15 Mins

Answer all the questions.

Each question is followed by four options lettered A to D. Find the correct option for each question and shade in pencil on answer sheet, the answer space which bears the same letter as the option you have chosen. Give only one answer to each question.

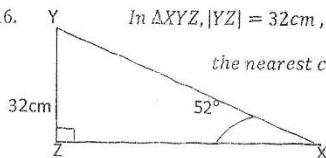
- U 1. Correct 0.04945 to two significant figures  
A. 0,05    B. 0.04    C. 0.050    D. 0.049
- A 2. Express as a ratio: the speed of 1 km/min to speed of 10 m/s  
A. 5:3    B. 5:4    C. 3:4    D. 3:5
- U 3.  $U = \{2,3,5,7\}$ ,  $P = \{2,5\}$  and  $Q = \{5,7\}$ . Find  $P^c \cap Q^c$   
A.  $\{2,3,7\}$     B.  $\{3,7\}$     C.  $\{2,3\}$     D.  $\{2,3,5\}$
- A 4. Solve:  $\frac{y+1}{2} - \frac{2y-1}{3} = 4$   
A.  $y = 29$     B.  $y = -29$     C.  $y = -19$     D.  $y = 19$
- E 5. Evaluate:  $\frac{\log_3 9 - \log_2 8}{\log_3 9}$   
A.  $-\frac{1}{2}$     B.  $\frac{1}{3}$     C.  $\frac{1}{2}$     D.  $-\frac{1}{3}$
- U 6. Express  $\frac{3^{\frac{1}{3}} - 2^{\frac{1}{2}}}{\frac{5}{12}}$  in its simplest form  
A. 2    B. -2    C. 3    D.  $-\frac{2}{3}$
- A 7. A binary operation is defined on the set of rational numbers  $Q$  as follows;  
 $a * b = a + b + 2ab$ . Evaluate  $\frac{1}{2} * \frac{3}{4}$   
A. 2    B. -2    C. 0    D. -1

(2,3,5,7)  
2,5  
1,2

I

8. If  $23y = 1111_2$ , find the value of  $y$ .  
 A. 7      B. 6      C. 5      D. 4
9. If  $3x - y = 5$  and  $2x + y = 15$ , evaluate  $x^2 + 2y$   
 A. 29      B. 30      C. 35      D. 42
10. Prisca gave a change of Gh¢ 1.15 instead of Gh¢ 1.25 to Franca.  
 Calculate the percentage error?  
 A. 8%      B. 8.7%      C. 9%      D. 9.2%
11. Evaluate:  $2\sqrt{28} - 3\sqrt{50} + \sqrt{72}$   
 A.  $4\sqrt{7} + \sqrt{2}$       B.  $4\sqrt{7} - 9\sqrt{2}$       C.  $4\sqrt{7} - 11\sqrt{2}$       D.  $4\sqrt{7} - 21\sqrt{2}$
12. If  $\tan x = \frac{3}{4}$ ,  $0 < x < 90$ , evaluate  $\frac{\cos x}{2 \sin x}$   
 A.  $\frac{2}{3}$       B.  $\frac{4}{3}$       C.  $\frac{3}{4}$       D.  $\frac{8}{3}$
13. Given that  $\vec{PQ} = (4\text{cm}, 060^\circ)$  it implies that  $-5\vec{PQ}$  is  
 A.  $(-20\text{cm}, -300^\circ)$       B.  $(20\text{cm}, 300^\circ)$       C.  $(20\text{cm}, 240^\circ)$       D.  $(-20\text{cm}, 0)$
14. Express  $\log 3 + 3 \log 2 - 3 \log 4$  as a single logarithm.  
 A.  $\log \frac{3}{16}$       B.  $\log \frac{3}{8}$       C.  $\log \frac{3}{2}$       D.  $\log \frac{8}{3}$
15. Write the following in standard form; 0.00000064  
 A.  $6.4 \times 10^7$       B.  $6.4 \times 10^{-7}$       C.  $6.4 \times 10^9$       D.  $6.4 \times 10^{-9}$

16. In  $\triangle XYZ$ ,  $|YZ| = 32\text{cm}$ ,  $\angle YXZ = 52^\circ$  and  $\angle XZY = 90^\circ$ . Find, correct to the nearest centimetre,  $|XZ|$



- A. 13cm      B. 20cm      C. 25cm      D. 31cm

$52^\circ$   
 $2104 + 1024$   
 $3128$   
 $61.05$

# I

- u 17. If  $P \propto \frac{1}{Q}$  and  $Q = r$ , find the equation connecting  $P$  and  $r$
- A.  $P = \frac{k}{r}$       B.  $P = kr$       C.  $P = \frac{r}{k}$       D.  $P = kr^2$
- u 18. If  $x:y = 2:3$ , find the value of  $\frac{x^2 + 2y^2}{y^2 - x^2}$
- A.  $2\frac{2}{5}$       B.  $3\frac{2}{5}$       C.  $4\frac{2}{5}$       D.  $6\frac{2}{5}$
- A 19. Linda is 11 years old and her brother, Asmah is 9 years old. They shared 60 oranges in the ratio of their ages. How many more oranges does Linda get?
- A. 6      B. 27      C. 34      D. 39
- u 20. Simplify  $\left(\frac{2}{3}\right)^{-3}$       A.  $\frac{8}{27}$       B.  $\frac{27}{8}$       C.  $-\frac{8}{27}$       D.  $-\frac{27}{8}$
- u 21. Solve the inequality  $3x - 4 < 5x + 4$
- A.  $x > 0$       B.  $x < -4$       C.  $x > -4$       D.  $x > -1$
- u 22. Find the standard deviation of the following numbers: 4, 6, 9, 10, 13, and 18.
- A. 13      B. 4.5      C. 4.6      D. 4.7
- A 23. Given that  $x = 3, y = -2$  and  $z = -1$ , evaluate  $\frac{2xy^2z}{3x - y}$
- A.  $\frac{23}{11}$       B.  $\frac{24}{11}$       C.  $-\frac{24}{11}$       D.  $\frac{24}{7}$
- A 24. Given that  $p^2 + q^2 = \frac{13}{36}$  and  $pq = \frac{5}{72}$ . Find  $(p + q)^2$
- A.  $\frac{18}{72}$       B.  $\frac{18}{2592}$       C.  $\frac{1}{2}$       D.  $\frac{8}{72}$
- A 25. Find the quadratic equation whose roots are  $-3$  and  $3$ .
- A.  $x^2 = 9$       B.  $x^2 + 9 = 0$       C.  $x^2 - 9x + 9 = 0$       D.  $x^2 - 9x = 6$
- u 26. In a class, 24 students read Physics, 21 read Mathematics and 9 read both Physics and Mathematics. If each student reads at least one of the two subjects, find the number of students in the class.
- A. 27      B. 45      C. 36      D. 54

11 + 9 = 20  
20 + 40 = 60

24 + 21 = 45

# I

- $\surd$  27. The result of adding 10 to a certain number is the same as multiplying the square of the same number by half. Which of the following equations represents this statement?
   
 A.  $x^2 - 2x - 20 = 0$     B.  $x^2 - x - 10 = 0$     C.  $2x^2 - x - 20 = 0$     D.  $x^2 - 2x - 10 = 0$
- $\surd$  28. Simplify  $\frac{a^2 - 4b^2}{a^2 + 4ab + 4b^2}$ 
  
 A.  $\frac{a+2b}{a-2b}$     B.  $\frac{a-4b}{a+4b}$     C.  $\frac{2a-b}{a+2b}$     D.  $\frac{a-2b}{a+2b}$
- $\Delta$  29. Given that  $a = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$ ,  $b = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$  and  $c = \begin{pmatrix} 8 \\ 4 \end{pmatrix}$  find  $2a - b + \frac{1}{2}c$ 
  
 A.  $\begin{pmatrix} -6 \\ 3 \end{pmatrix}$     B.  $\begin{pmatrix} 3 \\ 6 \end{pmatrix}$     C.  $\begin{pmatrix} -3 \\ -6 \end{pmatrix}$     D.  $\begin{pmatrix} 3 \\ -6 \end{pmatrix}$
- $\Delta$  30. If  $\log y = 3 \log 2 + \log 3 - \log 6$ , find the value of  $y$ .
   
 A. 8    B. 6    C. 3    D. 4
- $\Delta$  31. If  $8^x = 2\sqrt{2}$ , find the value of  $x$ .
   
 A.  $3/2$     B. 1    C.  $1/2$     D.  $-1/2$
- $\ast$   $\Delta$  32. Find the value of  $x$  in the equation:  $\cos(2x + 25) = \sin(x + 5)$  given that  $0^\circ \leq x \leq 90^\circ$ 
  
 A.  $90^\circ$     B.  $50^\circ$     C.  $20^\circ$     D.  $0^\circ$  and  $90^\circ$
- $\surd$  33. Simplify  $\frac{3}{14} - \frac{5}{21} + \frac{2}{35}$ 
  
 A.  $17/30$     B.  $17/210$     C.  $1/30$     D.  $1/210$
- $\Delta$  34. Factorize  $(x+6)^2 - 36x^2$ 
  
 A.  $(7x-6)(6-5x)$     B.  $(7x+6)(6x-5)$     C.  $(7x+6)(6+5x)$     D.  $(7x+6)(6-5x)$
- $\Delta$  35. A function is defined by  $h: x \rightarrow \frac{2x-6}{(x+2)(x-1)}$ . Find the value of  $x$  if  $h(x) = 0$ 
  
 A. 2 and 1    B. -2 and -1    C. -3    D. 3

36. Find the truth set of the quadratic equation  $x^2 + x = 20$

- A.  $\{x: x = -4, 5\}$  B.  $\{x: x = -5, -4\}$  C.  $\{x: x = -5, 4\}$  D.  $\{x: x = -4, 9\}$

At a rally attended by 520 people, 30% were Fantos, 25% Ewes, 15% Nzemas, 20% Gas and the rest Gonjas.

Use the information above to answer questions 37 and 38.

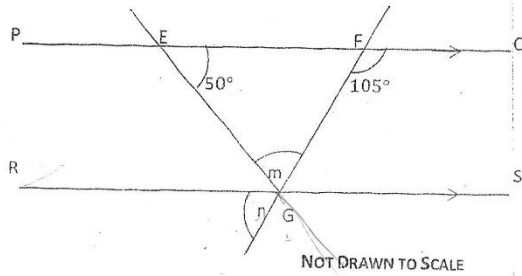
37. How many Gonjas were at the rally?

- A. 0 B. 10 C. 25 D. 52

38. How many more Fantos than Nzemas were at the rally?

- A. more than half B. 156 C. 15 D. 78

155  
130  
78  
104  
156  
468



105  
130  
176  
104

In the diagram  $\overline{PQ}$  is parallel to  $\overline{RS}$ ,  $\angle QFG = 105^\circ$  and  $\angle FEG = 50^\circ$ . Use the diagram to answer questions 39 and 40.

39. What is the value of the angle marked  $m$ ?

- A.  $55^\circ$  B.  $75^\circ$  C.  $105^\circ$  D.  $130^\circ$

40. Find the value of  $n$ .

- A.  $130^\circ$  B.  $75^\circ$  C.  $55^\circ$  D.  $40^\circ$

75  
130  
13

END OF SECOND SEMESTER EXAMINATION DECEMBER, 2020

CORE MATHEMATICS

EVEL: FORM TWO (2) GREEN AND GOLD

DURATION: 2HRS 15MINS

SECTION A (OBJECTIVE)

1. Evaluate, correct to two decimal places,  $75.0785 - 34.624 + 9.83$ 
  - a. 30.62
  - b. 50.28
  - c. 50.29
  - d. 30.60
2. If  $X = \{x: x < 7\}$  and  $Y = \{y: y \text{ is a factor of } 24\}$  are subsets of  $U = \{1, 2, 3, \dots, 10\}$ . Find  $X \cap Y$ .
  - a.  $\{1, 2, 3, 4, 6, 8\}$
  - b.  $\{2, 3, 4, 6, 8\}$
  - c.  $\{1, 2, 3, 4, 6\}$
  - d.  $\{2, 3, 4, 6\}$
3. Simplify:  $(5\frac{1}{3} - 2\frac{7}{8}) \div (2\frac{1}{3} + \frac{1}{6})$ 
  - a.  $2\frac{11}{24}$
  - b.  $1\frac{9}{60}$
  - c.  $\frac{59}{60}$
  - d.  $\frac{6}{15}$
4. Given that  $\tan y = -\frac{1}{4}$ ,  $90^\circ < x < 180^\circ$ , calculate  $\cos y$ .
  - a.  $\frac{4}{\sqrt{17}}$
  - b.  $\frac{1}{\sqrt{17}}$
  - c.  $-\frac{1}{\sqrt{17}}$
  - d.  $-\frac{4}{\sqrt{17}}$
5. If  $x = 3$  and  $y = -1$ , evaluate  $2(x^2 - y^3)$ 
  - a. 16
  - b. 20
  - c. 22
  - d. 24
6. Find the least value of  $x$  which satisfies the equation  $4x = 7 \pmod{9}$ .
  - a. 4
  - b. 5
  - c. 6
  - d. 7
7. Simplify  $\left[\left(\frac{16}{9}\right)^{-\frac{3}{2}} \times 16^{-\frac{3}{4}}\right]^{\frac{1}{3}}$ 
  - a.  $\frac{1}{4}$
  - b.  $\frac{3}{8}$
  - c.  $\frac{9}{16}$
  - d.  $\frac{3}{4}$
8. Given that  $\frac{\sqrt{3} + \sqrt{5}}{\sqrt{5}} = x + y\sqrt{15}$ , find the value of  $(x + y)$ .
  - a.  $\frac{1}{5}$
  - b.  $1\frac{1}{5}$
  - c.  $1\frac{1}{5\sqrt{5}}$
  - d.  $1\frac{1}{5\sqrt{15}}$
9. Solve  $\frac{1}{3}(5 - 3x) < \frac{2}{5}(3 - 7x)$ 
  - a.  $x < -\frac{7}{27}$
  - b.  $x > -\frac{7}{27}$
  - c.  $x < \frac{7}{22}$
  - d.  $x > \frac{7}{22}$

10. The expression  $\frac{5x+3}{6x(x+1)}$  will be undefined

- when  $x$  equals
- $\{-3, 0\}$
  - $\{-3, -1\}$
  - $\{0, -1\}$
  - $\{0, 1\}$

11. Make  $m$  the subject of the relation  $k =$

$$\frac{\sqrt{m-y}}{\sqrt{m+1}}$$

a.  $m = \frac{y-k^2}{1-k^2}$

b.  $m = \frac{y-k^2}{k^2+1}$

c.  $m = \frac{y+k^2}{1-k^2}$

d.  $m = \frac{y+k^2}{k^2+1}$

12. Express  $10010_{two}$  in base ten

- 36
- 34
- 18
- 10

13. Simplify  $\frac{a}{b} - \frac{b}{a} - \frac{c}{b}$

a.  $\frac{a^2-b^2-ac}{ab}$

b.  $\frac{a^2-b^2+ac}{ab}$

c.  $\frac{a^2-bc-ac}{ab}$

d.  $\frac{a-b+c}{ab}$

14. A fence  $2.4m$  tall, is  $10m$  away from a tree of height  $16m$ . Calculate the angle of elevation of the top of the tree from the top of the fence.

- $51.32^\circ$
- $52.40^\circ$
- $53.67^\circ$
- $76.11^\circ$

15. Fati buys milk at  $x$  per tin and sells each at a profit of  $y$ . If she sells 10 tins of milk, how much does she receive from the sales?

- $10(x+y)$
- $(10x+y)$
- $(x+10y)$
- $(xy+10)$

16. The interquartile range of a distribution is 7.

If the 25<sup>th</sup> percentile is 16, find the upper quartile.

- 9
- 23
- 30
- 35

17. The variance of a distribution is 9. Find the standard deviation.

- 81
- 27
- 9
- 3

18. The ages of three women are in the ratio of 3:4:5. If the difference between the ages of the eldest and youngest is 18 years, find the sum of the ages of the three women.

- 54 years
- 72 years
- 108 years
- 216 years

Marks	0	1	2	3	4
Frequency	4	6	4	2	4

Use the frequency distribution table to answer questions 19 - 21

19. Calculate the mean of the distribution.

- 3.6
- 2.0
- 1.8
- 0.5

20. Find the range of the distribution.

- 4
- 3
- 2
- 1

Find the median of the distribution.

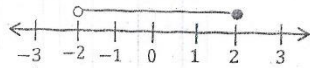
- a. 3.0
- b. 2.5
- c. 2.0
- d. 1.5

22. Express 4382.93 in standard form.

- a.  $438293 \times 10^4$
- b.  $43.8293 \times 10^2$
- c.  $4.38293 \times 10^4$
- d.  $4.38293 \times 10^3$

23. The cost of 12 note books is ₵54.84. Find the cost of one note book.

- a. GHC5.57
- b. GHC4.67
- c. GHC4.57
- d. GHC3.57



24. Which of the following inequalities is represented on the number line shown above?

- a.  $-2 > y > 2$
- b.  $-2 \leq y < 2$
- c.  $-2 \geq y > 2$
- d.  $-2 < y \leq 2$

25. Which of the following vectors is parallel to the vector  $\begin{pmatrix} 20 \\ 8 \end{pmatrix}$ .

- a.  $\begin{pmatrix} -25 \\ 10 \end{pmatrix}$
- b.  $\begin{pmatrix} -2 \\ -5 \end{pmatrix}$
- c.  $\begin{pmatrix} -15 \\ -6 \end{pmatrix}$
- d.  $\begin{pmatrix} -4 \\ 10 \end{pmatrix}$

26. Given that  $a = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ ,  $b = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ , and  $c =$

$\begin{pmatrix} -2 \\ 5 \end{pmatrix}$  evaluate  $a + 2b - c$

- a.  $\begin{pmatrix} 10 \\ 4 \end{pmatrix}$
- b.  $\begin{pmatrix} 7 \\ 10 \end{pmatrix}$
- c.  $\begin{pmatrix} 3 \\ 0 \end{pmatrix}$
- d.  $\begin{pmatrix} 6 \\ -4 \end{pmatrix}$

27. The bearing of a point Q from another point P is  $040^\circ$ . Find the bearing of P from Q.

- a.  $220^\circ$
- b.  $160^\circ$
- c.  $130^\circ$
- d.  $040^\circ$

28. A function  $g$  is defined by  $g: x \rightarrow \frac{2x-1}{x}$ . For what value of  $x$  is  $g$  undefined?

- a. 1
- b.  $\frac{1}{2}$
- c. 0
- d.  $-\frac{1}{2}$

29. Which of the following is/are rational?

I.  $P = \{2, \sqrt{3}, 5\}$

II.  $Q = \{7, \frac{1}{6}, \sqrt{9}\}$

III.  $R = \{-3 \leq x \leq 8\}$

- a. I only
- b. II only
- c. III only
- d. I and III only

30. Which of the following statements is false?

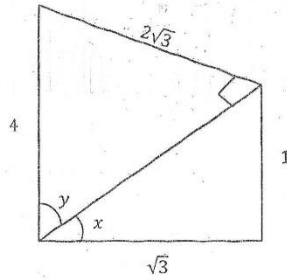
- a.  $x + 3 > 2 \Rightarrow x > -1$
- b.  $x > -5 \Rightarrow x + 1 > -5 + 1$
- c.  $x = 3 \Rightarrow x^2 = 9$
- d.  $x + 3 > 4 \Rightarrow x < 1$



31. Simplify  $\frac{4-x^2}{2x-x^2}$  where  $x \neq 0$  and  $x \neq 2$

- a.  $\frac{2-x}{x}$
- b.  $\frac{2+x}{x}$
- c.  $\frac{2}{x}$
- d.  $\frac{x-2}{x}$

32. In the diagram below, find the sum of  $x$  and  $y$ .



- a.  $120^\circ$
- b.  $60^\circ$
- c.  $90^\circ$
- d.  $45^\circ$

33.  $\vec{PQ} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$ ,  $\vec{RQ} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ . Find  $\vec{PR}$

- a.  $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$
- b.  $\begin{pmatrix} 3 \\ -1 \end{pmatrix}$
- c.  $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$
- d.  $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

34. Factorise  $2m - 2n + mx - nx$

- a.  $(2+x)(m-n)$
- b.  $(2+m)(n-x)$
- c.  $2x(m-n)(m+n)$
- d.  $(2+n)(m+x)$

35. Which of the following statements are true?

- I.  $\sqrt{4} \times \sqrt{9} = \sqrt{36}$
  - II.  $\sqrt{16} \times \sqrt{9} = \sqrt{15}$
  - III.  $\sqrt{100} \times \sqrt{25} = \sqrt{2500}$
  - IV.  $\sqrt{64} \times \sqrt{36} = \sqrt{100}$
- a. I and III only
  - b. I and IV only
  - c. II and IV only
  - d. II and III only

36. The least positive integer such that  $2x \equiv 1 \pmod{3}$  is

- a. 0
- b. 6
- c. 8
- d. 12

37. How many vertices has a cuboid?

- a. 4
- b. 6
- c. 8
- d. 12

38. If  $11y = (18)^2 - (15)^2$ , find the value of  $y$ .

- a. 1
- b. 8
- c. 9
- d. 10

39. List the integers within the interval  $7 < x \leq 14$

- a.  $\{8, 9, 10\}$
- b.  $\{10, 11, 12, 13\}$
- c.  $\{8, 9, 10, 11, 12, 13, 14\}$
- d.  $\{14\}$

40. The diagonal of a square is  $18\text{cm}$  long.

Which of the following gives, in  $\text{cm}$ , the length of the side of the square?

- a. 3
- b.  $3\sqrt{2}$
- c. 6.75
- d. 9

J

**Section B**

**Part I (answer all questions under this part)**

- Without using mathematical tables or calculators, evaluate  $\frac{0.015 \times 0.567}{0.05 \times 0.189}$ , leaving your answer in standard form.
- (a) Factorise completely:  $4b^2 - ab + (a + 9b)^2 - a^2$   
(b) Simplify  $\sqrt{28} + \frac{1}{\sqrt{7}}$
- Solve  $\frac{5x-7}{6} + \frac{2x-3}{4} = -\frac{2}{3}$
- Given that  $f: x \rightarrow 2x^2 - 8x + 5$   
 $g: x \rightarrow x - 2$   
Find;  
(i)  $f(-3)$   
(ii) The values of  $x$  such that  $f(x) = g(x)$
- If  $a = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ ,  $b = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$  and  $r = a + \frac{1}{2}(a + b)$ , find  
(i)  $r$   
(ii)  $|r|$

**Part II (answer any three questions under this part)**

- The points  $P, Q$  and  $R$  are located such that  $Q$  is  $15\text{ km}$  due south of  $P, R$  is  $10\text{ km}$  from  $P$  on a bearing of  $270^\circ$ . Calculate, correct to  
(i) two significant figures  $|\overline{QR}|$   
(ii) the nearest degree, the bearing of  $Q$  from  $R$
- The following data shows marks obtained by students in a Mathematics test.

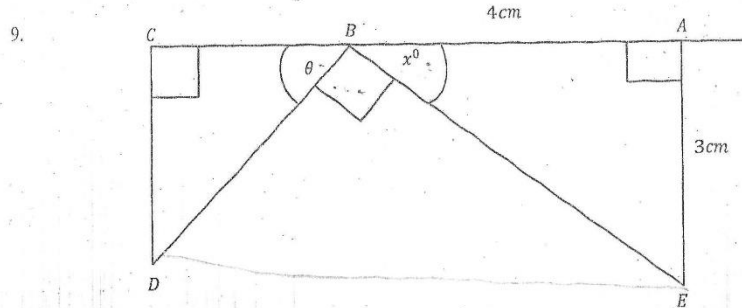
6	7	5	9	5	3
7	9	5	6	2	7
10	5	9	8	0	6
2	2	6	6	5	6
9	6	7	4	1	8

  
(a) Construct a frequency distribution table for the data  
(b) Calculate the mean for the distribution

8. (a) Copy and complete the table of values for  $y = 5 \sin x + 9 \cos x$  for  $0^\circ \leq x \leq 150^\circ$

$x$	$0^\circ$	$30^\circ$	$60^\circ$	$120^\circ$	$150^\circ$
$y$		10.3		-0.2	

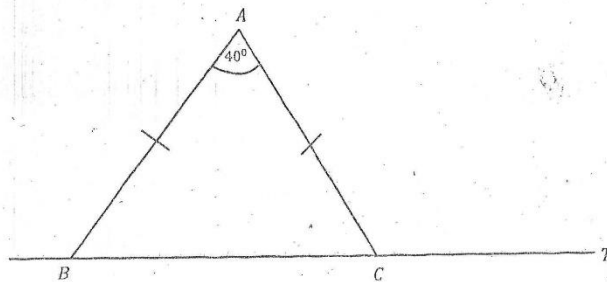
- (b) Using a scale of  $2\text{cm}$  to  $30^\circ$  on the  $x$ -axis and  $2\text{cm}$  to 2 units on the  $y$ -axis, draw the graph of  $y = 5 \sin x + 9 \cos x$  for  $0^\circ \leq x \leq 150^\circ$
- (c) Use your graph to solve the equation:  $5 \sin x + 9 \cos x = 0$



If  $\tan x = \frac{3}{4}$  and  $|BD| = 2\text{cm}$ , find, from the diagram above,

- (i)  $\sin x$   
 (ii)  $x^\circ$   
 (iii)  $|AC|$

10. (a) The height of an isosceles triangle is  $7\text{cm}$ . Find the length of the base, if the other sides are  $8\text{cm}$  long.
- (b) In the triangle,  $ABC$  is produced to meet  $T$ .  $|AB| = |AC|$ ,  $\angle BAC = 40^\circ$  and  $\angle ACT = (13x - 15)^\circ$ . Find the value of  $x$ .



**APPENDIX E**  
**INTEGRATED SCIENCE TEST SAMPLES**

**End of Second Semester Examination**  
**Integrated Science**  
**Form: SHS 1**  
**Time: 1 Hour 45 Mins.**

F

SECTION A

PART TWO

Answer three (3) questions only from this part. All questions carry equal marks

3a. what is binomial nomenclature?

b. Write the systematic names of the following

(i).  $\text{CaCO}_3$

(ii).  $\text{MnO}_4^-$

(iii).  $\text{HNO}_3$

c. Give four (4) reasons why we need to measure accurately

d. Briefly explain the industrial nitrogen fixation

e. Mention three (3) factors that affect active transport

4a. define each of the following

i. Active transport

ii. Osmosis

iii. Atomic number

iv. Mass number

b. state two(2) effects of temperature on weathering of rocks

c. explain three(3) ways through which nitrogen is lost so that crops cannot benefit from the nutrients

d. A body has a density of  $4000\text{kg/m}^3$  if the density of water is  $1000\text{kg/m}^3$ .

(i). Calculate the Relative density of the body

(ii). Will the body sink or float in the water? Why?

Sai. what is dilution?

ii. What is the concentration in  $\text{mol dm}^{-3}$  of  $35\text{g NaOH}$  dissolved in water to produce  $500\text{cm}^3$   $\text{NaOH}$  solution? [R.A.M.

$\text{H}=1.0, \text{O}=16, \text{Na}=23$ ].

b. Explain five (5) factors that affect the rate of diffusion

c. State the basic Units of measurement

1 | Page

$\frac{23}{16}$   
 $\frac{1}{40}$   
 $\frac{23}{16} \times \frac{1}{40}$

d. Distinguish between physical weathering and chemical weathering

6a. define the following

- i. Denitrification
- ii. Nitrification

b(i). Name the types of vertebrae found in mammals

(ii). State the region of the mammals body where each type of the vertebrae you have name in b (i) is found

c(i). Name the three basic particles of an atom

(ii). for each particle named in C (i) state the charge and also give its mass

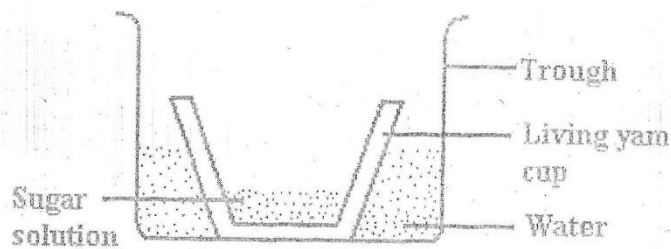
d. Tabulate three differences between mass and weight

### SECTION A

#### PART ONE PRACTICALS COMPULSORY

Answer all questions in this part

1. The figure below represents the beginning of an experiment to demonstrate osmosis in a living cell using yam tissue.



1a. draw and label the diagram to illustrate what will be observed if the setup is allowed to stand for 24 hours

b. what does the yam cavity represent?

c. explain the principles involved in the experiment

d. how would you set up a control experiment above?

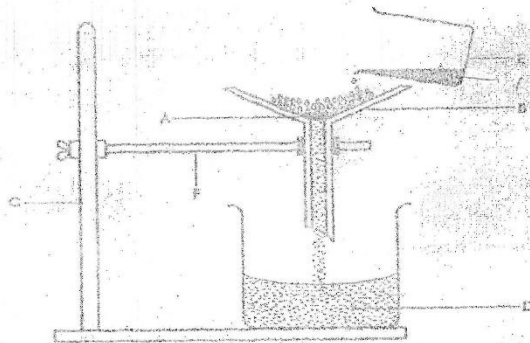
F

e. give one (1) example of the osmotic process in each of the things,

(i). flowering plant

(ii). Humans

2. The diagram below shows how a student separated a mixture. Use it to answer the questions that follows.



13

a. Name the parts labelled A, B, C, D, E and F

b. what is the method of separation which is used by the student?

c. What is E made up?

d. What name is given to the substances left in A?

### OBJECTIVES

1. Sexual reproduction in plants is through  
(a) bulbs      (b) corms      (c) seeds      (d) rhizomes
2. A fruit with a sticky covering is likely to be dispersed by  
(a) explosion      (b) mammals      (c) water      (d) wind
3. Which of these bones are in the lower arm of a mammal?  
(a) scapular and humerus      (b) humerus and ulna  
(c) radius and scapular      (d) radius and ulna

4. Which of the following vertebrae has the odontoid process?  
 (a) Atlas (b) Axis (c) Lumbar (d) Thoracic
5. A collection of sepals is called.....  
 (a) gynoecium (b) calyx (c) androecium (d) corolla
6. The head of the humerus is connected to the scapula by a  
 (a) hinge joint (b) gliding joint  
 (c) pivot joint (d) ball and socket joint
7. Which of the following units is not from temperature?  
 (a) Kelvin (b) Degree Celsius  
 (c) Pascal (d) Degree Fahrenheit
8. Which of the following units is equivalent to 1W  
 (a)  $1\text{Js}^{-1}$  (b) 1J (c) 1N (d)  $1\text{Ns}^{-1}$
9. A clinical thermometer is not sterilized in boiling water because  
 (a) a vector may contaminate it (b) it has a narrow bore  
 (c) water has lower boiling point (d) it has short range of temperature
10. The S.I unit of power is .....  
 (a) NmS (b)  $\text{NmS}^{-1}$  (c)  $\text{Nms}^{-1}\text{S}^{-1}$  (d)  $\text{Nm}^{-1}\text{S}$
11. Which of the following instruments can accurately measure a length of 0.001cm?  
 (a) Micrometer screw gauge (b) Measuring tape  
 (c) Meter rule (d) Vernier calipers
12. Which of the following ranks in classification shows the least resemblance among organism?  
 (a) Species (b) Class (c) Genus (d) Order
13. Dog is a common name. A correct way of writing a dog's scientific name is.....  
 (a) Canis Familiaris (b) canis Familiaris  
 (c) canis familiaris (d) Canis familiaris
14. Aristotle classified plants into  
 (a) trees, shrubs, herbs (b) trees, grass, weeds  
 (c) grass, shrub, herbs (d) weeds, shrubs, herbs
15. He introduced the scientific naming of organisms. He is .....  
 (a) Mendeleev (b) Aristotle  
 (c) Carolus Linnaeus (d) John ray
16. An organism having a definite cell wall, a nucleus but lacking chlorophyll is likely to belong to the Kingdom.....  
 (a) fungi (b) plantae (c) prokaryotea (d) protatista
17. Which of the following is not a non - living thing?  
 (a) Frog (b) Water (c) Man (d) Orange tree

H  
2019/2020 ACADEMIC YEAR

END OF FIRST SEMESTER EXAMINATION-GOLD TRACK

INTEGRATED SCIENCE

TIME: 30

SECTION A-OBJECTIVES

1. Atmospheric pressure is measured with a/an
  - a. Anemometer
  - b. Thermometer
  - c. Galvanometer
  - d. Barometer
2. The instrument used to measure current in an electric circuit is the
  - a. Ammeter
  - b. Anemometer
  - c. Barometer
  - d. Thermometer
3. Select the odd one out
  - a. Micrometer screw gauge
  - b. Callipers
  - c. Measuring Cylinder
  - d. Metre Rule
4. Which of the following instruments is **not** used for the purpose stated against it?
  - a. Chemical balance measures the weight of a body
  - b. Thermometer is used to measure the degree of hotness of a body
  - c. Vernier Callipers are used to measure the internal and external diameters of cylindrical objects
  - d. Micrometer screw gauge is used to measure diameters of thin wires
5. Which of the following instruments is best used to measure the radius of an egg
  - a. Spherometer
  - b. Micrometer screw gauge
  - c. Vernier Callipers
  - d. Opisometer
6. Safety devices to be used in chemistry laboratory include:
  - i. a Pair of goggles ✓
  - ii. nose guard ✓
  - iii. crash helmet
  - iv. hand gloves
  - a. i,ii and iii only
  - b. i,ii and iv only
  - c. ii,iii and iv only
  - d. i,ii,iii and iv
7. Which of the following symbols represents the S.I unit for electric current
  - a. C
  - b. V
  - c. K
  - d. A
8. Which of the following units is derived?
  - a. K
  - b. Pa
  - c. kg
  - d. s

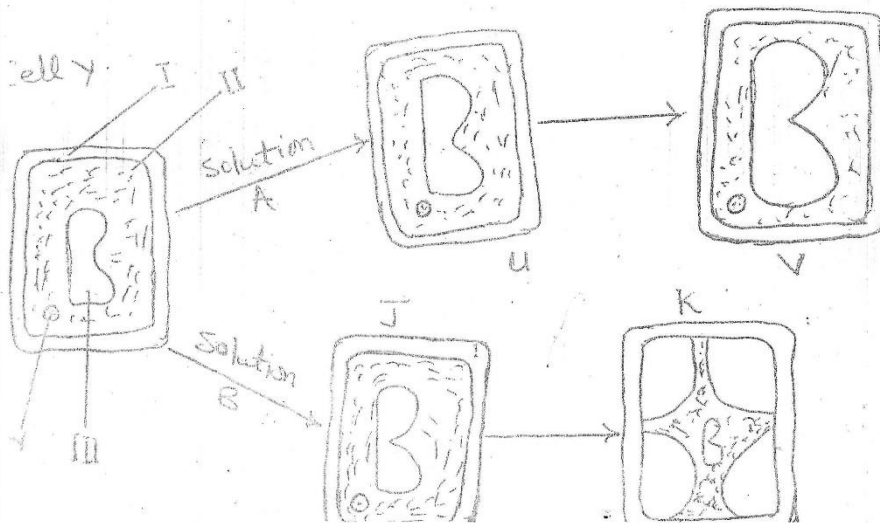
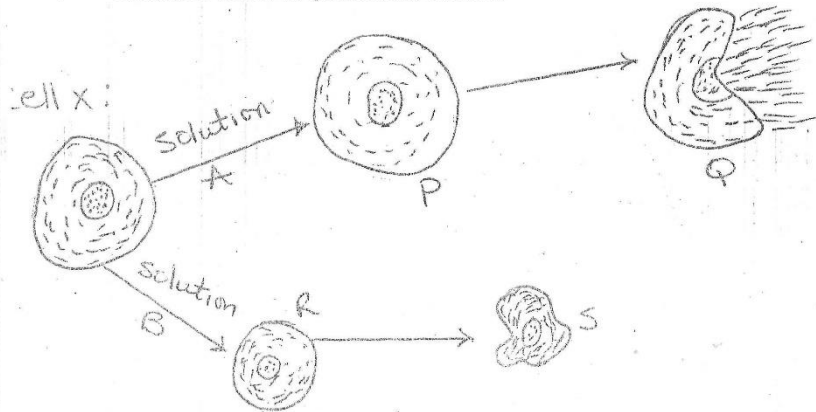


END OF SEMESTER EXAMINATION  
DECEMBER 2020  
SHS 2 INTEGRATED SCIENCE  
2½ HOURS

NAME:.....  
CLASS:.....  
DATE:.....

SECTION 'A' (1HR. 30MIN.)  
Answer question one (1) and any other two (2) from the rest

1. The diagrams below demonstrate a biological phenomenon in two (2) different types of cells. Use them to answer the questions that follow-a.



9. A liquid has relative density of 12.2. Calculate its density (Density of water =  $1.0 \times 10^3 \text{ kg m}^{-3}$ )
- $1.220 \text{ kg m}^{-3}$
  - $12.20 \text{ kg m}^{-3}$
  - $122.00 \text{ kg m}^{-3}$
  - $1200 \text{ kg m}^{-3}$
10. When a hydrometer floats in a liquid, it means that the
- weight of the liquid displaced is equal to weight of the hydrometer
  - density of the hydrometer is equal to the density of the liquid
  - upthrust on the hydrometer is less than the volume of the hydrometer
  - volume of the portion of the hydrometer in the liquid is equal to the volume of the liquid displaced
11. A body of mass 2kg floats on water. If the density of water is  $1000 \text{ kg m}^{-3}$ , Calculate the volume of body immersed
- $2.0 \times 10^{-3} \text{ m}^3$
  - $2.0 \times 10^{-2} \text{ m}^3$
  - $5.0 \times 10^3 \text{ m}^3$
  - $2.0 \times 10^2 \text{ m}^3$
12. A body has density of  $100.0 \text{ kg m}^{-3}$ , if its mass is 4.0kg, Calculate its volume
- $4.0 \times 10^3 \text{ m}^3$
  - $4.0 \times 10^{-3} \text{ m}^3$
  - $4.0 \times 10^3 \text{ m}^3$
  - $4.0 \times 10^{-2} \text{ m}^3$
13. For a body to float in a fluid, it must displace a quantity of the fluid equal to its
- Mass
  - Surface area
  - Volume
  - Weight
14. A small metal block is 10cm long, 5.0cm high and 3.0cm wide. If it has a mass of 1.2kg, Calculate its density
- $0.008 \text{ g cm}^{-3}$
  - $8.0 \text{ g cm}^{-3}$
  - $18.0 \text{ g cm}^{-3}$
  - $180.0 \text{ g cm}^{-3}$
15. The relative density of acid in a car battery can be measured with
- an ammeter
  - hydrometer
  - hygrometer
  - voltmeter
16. A floating body displaces an amount of fluid equal to its weight. This statement is the
- Archimedes' principle
  - law of floating
  - law of gravity
  - Newton's third law of motion
17. A floating body displaces a quantity of liquid equal to its own
- Density
  - Mass
  - Volume
  - Weight
18. The unit of velocity is
- ms
  - $\text{ms}^{-1}$
  - $\text{ms}^{-2}$
  - $\text{m}^2 \text{ s}^{-1}$
19. Scientific attitude does not include
- critical observation
  - asking questions
  - curiosity
  - superstitious beliefs

$\text{Density} = \frac{m}{V}$   
 $\text{Relative Density} = \frac{\rho_{\text{body}}}{\rho_{\text{fluid}}}$   
 $\rho_{\text{body}} = \text{RD} \times \rho_{\text{fluid}}$   
 $\rho_{\text{fluid}} = \frac{m}{V}$   
 $\rho_{\text{body}} = \frac{m}{V} \times \text{RD}$   
 $V = \frac{m}{\rho_{\text{body}}}$   
 $V = \frac{m}{\text{RD} \times \rho_{\text{fluid}}}$   
 $V = \frac{m}{\text{RD} \times 1000}$   
 $V = \frac{4.0}{100.0 \times 1000}$   
 $V = 4.0 \times 10^{-5} \text{ m}^3$

H

20. Linkages or inter relation of pure sciences studied together is known as
- a. Integrated Science
  - b. Applied Science
  - c. Technology
  - d. Computing

Answer: A

$2.5 \times 10^{-10}$   
m m cg dg m Dg hg kg  
0 0 0 0 0 1  
0000000  
M  
0.01

- 4
- ai) Identify the cells X and Y. (2mks)
- ii) Name the part of cell Y labelled I – IV. (2mks)
- iii) Give the types of solutions of A and B. (2mks)
- iv) Name the biological Phenomenon being demonstrated in both cells X and Y. (1mk)
- v) What special terms are used to describe the conditions labelled Q, S, V and K. (4mks) *guard cell*
- bi) State two (2) examples each of the phenomenon in  $\alpha$ ) animal  $\beta$ ) plants. *DIFFUSION*
- ii) Name other two (2) biological phenomena that are involved in movement of substances into and out of cells. (2mks)
- iii) State three (3) importance of movement of substances into and out of the cells. (3mks)
- 2a) Give the symbols for the following electronic devices
- i) LED      ii) Diode      iii) Amplifier
- iv) Variable capacitor      v) Inductor (5mks)
- b) Use diagram to describe how a diode can be used to change a.c to d.c (8mks)
- ci) Differentiate between intrinsic and extrinsic semiconductors. (2mks)
- ii) State three (3) uses of LED (3mks)
- iii) Name the two (2) types of transistors and their symbols. (2mks)
- 3ai) Write a balanced chemical equation for the production of glucose in green plants (5mks)
- ii) State three (3) importance of photosynthesis in the ecosystem. (3mks)
- iii) State five (5) factors that affect the rate of photosynthesis. (5mks)
- iv) What energy transformation occurs in the process of photosynthesis? (2mks)
- b) A student was given a green leaf which has been exposed to the sun for about six hours. Briefly describe how he will test for the presence of starch in the leaf in the laboratory. (5mks)
- Q4a) Describe the mechanism of inhalation in human. (5mks)
- bi) Name two (2) products of anaerobic respiration. (2mks)
- ii) State one function each of the following structures in the human body.
- $\alpha$ ) alveoli
- $\beta$ ) diaphragm (2mks)
- ci) Define energy
- ii) State the energy transformation the occurs in each of the following devices.
- $\gamma$ ) torch light
- $\beta$ ) moving motorcycle (6mks)
- d) An earth moving machine uses diesel oil which will burn to produce  $3.0 \times 10^9$  J of energy. The work performed by the machine is  $2.8 \times 10^8$  J
- i) What is the efficiency of the machine?
- ii) Explain why the efficiency of the machine is less than 100%. (5mks)



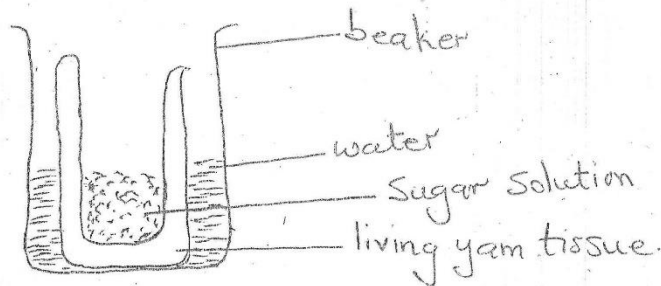
5

- Q5a) Define the following
- i) Molar mass
  - ii) Amount of substance concentration
  - iii) Avogadro's number
- (3mks)
- b) A student prepares a glucose solution by dissolving a particular mass of the glucose in 500cm<sup>3</sup> of distilled water. If the concentration of the solution is 0.5M. Calculate
- i) The mass of the glucose used
  - ii) The mass concentration of the solution (C=12, O=16, H=1)
- (5mks)
- cia) Name three (3) processes that are involved in the nitrogen cycle. (3mks)
- ii) State three importance of nitrogen fixation in the soil. (3mks)
- d) Briefly explain the following terms as applied to weathering of rocks.
- i) Hydrolysis
  - ii) Oxidation
  - iii) Reduction
- (6mks)



- G
10. The energy possessed by a student on top of a building is
    - a) heat energy
    - b) chemical energy
    - c) potential energy
    - d) kinetic energy
  11. The bronchi in the lungs divide into smaller branches which end into tiny sacs called
    - a) alveoli
    - b) bronchioles
    - c) spiracles
    - d) tubules
  12. The main process, by which living things release energy from food is known as
    - a) digestion
    - b) fermentation
    - c) oxidation
    - d) assimilation
  13. In flowering plants, mineral salts are transported through the
    - a) xylem
    - b) phloem
    - c) stomata
    - d) cambium
  14. One major role of water in plants is that it
    - a) keeps the plants green
    - b) allows for evaporation
    - c) prevents fungal attack
    - d) keeps the stem turgid
  15. Which of the following substances is the by-products of photosynthesis
    - a) carbon dioxide
    - b) glucose
    - c) oxygen
    - d) water
  16. Salting of fish for preservation uses the principle of
    - a) diffusion
    - b) osmosis
    - c) plasmolysis
    - d) active transport

Use the diagram below to answer questions 17 – 19



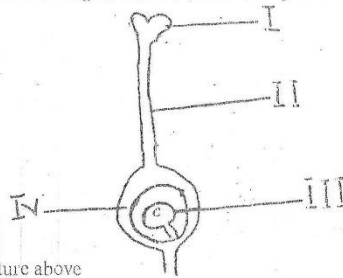
17. What phenomenon is being demonstrated in the diagram above?
  - a) Osmosis in a non-living tissue
  - b) Osmosis in a living tissue
  - c) Diffusion in a living tissue
  - d) Active transport in plants





- G
18. What name is given to the salt solution when compared to that water.
    - a) isotonic solution
    - b) hypertonic solution
    - c) hypotonic solution
    - d) diluted solution
  19. What will happen to the level of the salt solution after about two hours?
    - a) it will remain the same
    - b) It will rise
    - c) It will reduced
    - d) It will reduced initially and rise again
  20. Which of the following factor is a non-living factor in the ecosystem?
    - a) parasitism
    - b) predation
    - c) provision of shades by plant
    - d) pollution
  21. Movement of water into plant cell causes the cell to
    - a) become turgid
    - b) becomes flaccid
    - c) burst its cell wall
    - d) becomes plasmolysed

Use the diagram below to answer questions 22 and 23

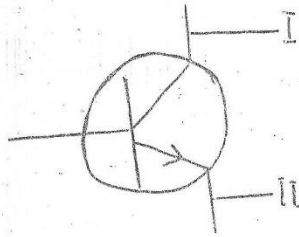


22. Identify the structure above
  - a) flower
  - b) pistil
  - c) carpel
  - d) stamen
23. The parts labelled I, II, III and IV are respectively
  - a) ovary, stigma, style and ovule
  - b) style, ovule, stigma and ovary
  - c) stigma, ovary, ovule and style
  - d) stigma, style ovule and ovary
24. A tissue is made up of cells which
  - a) performs different functions
  - b) have different structures
  - c) have similar structures and functions
  - d) look alike in shape
25. Which of the following factors will result in the formation of land and sea breeze?
  - a) convection current
  - b) the different phases of the moon
  - c) the revolution of the earth
  - d) volcanoes
26. Denitrifying bacteria reduce
  - a) ammonium ions to nitrate ions
  - b) nitrate ions to nitrite ions
  - c) nitrate ions into gaseous nitrogen
  - d) amino acids to ammonium salts

- 6
27. The process or phenomenon by which a plant or animal cell loses water by osmosis as a result of the cell being placed in hypertonic solution is
    - a) haemolysis
    - b) plasmolysis
    - c) photosynthesis
    - d) turgidity
  28. The movement of water into animal cause the cell to
    - a) become flaccid
    - b) become turgid
    - c) burst its cell contents
    - d) become plasmolysed
  29. A body of mass 2kg floats on water. If the density of water is  $1000\text{kgm}^{-3}$ , determine the volume of the body
    - a)  $2.0 \times 10^{-1}\text{m}^3$
    - b)  $2.0 \times 10^{-3}\text{m}^3$
    - c)  $2.0 \times 10^{-2}\text{m}^3$
    - d)  $2.0 \times 10^3\text{m}^3$
  30. A small metal block is 12.0cm long, 6.0cm high and 4.0cm wide. If it has a mass of 1.5kg, determine its density
    - a)  $0.2\text{gcm}^{-3}$
    - b)  $5.2\text{gcm}^{-3}$
    - c)  $192.0\text{gcm}^{-3}$
    - d)  $288.0\text{gcm}^{-3}$
  31. A ball and socket joint in the skeletal system allows
    - a) circular movement
    - b) gliding movement
    - c) side by side movement
    - d) up and down
  32. Which of the following units represent electric current?
    - a) Ampere
    - b) Kelvin
    - c) seconds
    - d) metre
  33. One of the following structures has mucus and cilia for trapping dust and germs
    - a) Alveolus
    - b) bronchiole
    - c) bronchus
    - d) trachea
  34. A transistor can be used as
 

I) switch	II) Amplifier	III) rectifier
a) I only	b) I and III	
c) I and III	d) I, II and III	
  35. Which of the following factors will not affect the rate of diffusion
    - a) pressure
    - b) temperature
    - c) surface area
    - d) concentration gradient
  36. The fluid which prevent friction at movable joint is the
    - a) plasma
    - b) synovial fluid
    - c) Amniotic fluid
    - d) Lymph
  37. The IUPAC name for  $\text{CO}_2$  is
    - a) carbon (IV) oxide
    - b) carbon (II) oxide
    - c) carbon (I) oxide
    - d) carbon oxide (IV)

Use the diagram below to answer questions 38 and 39

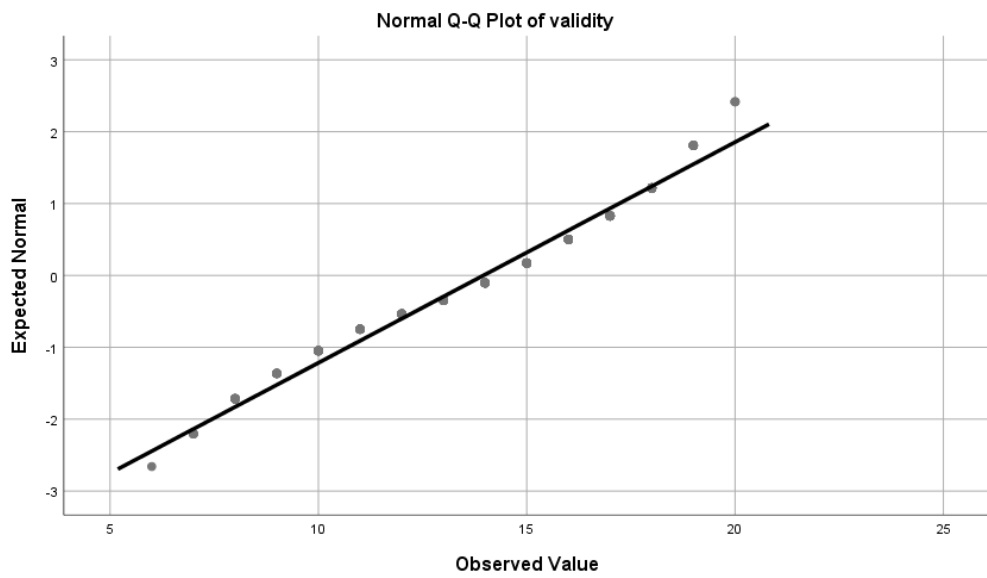
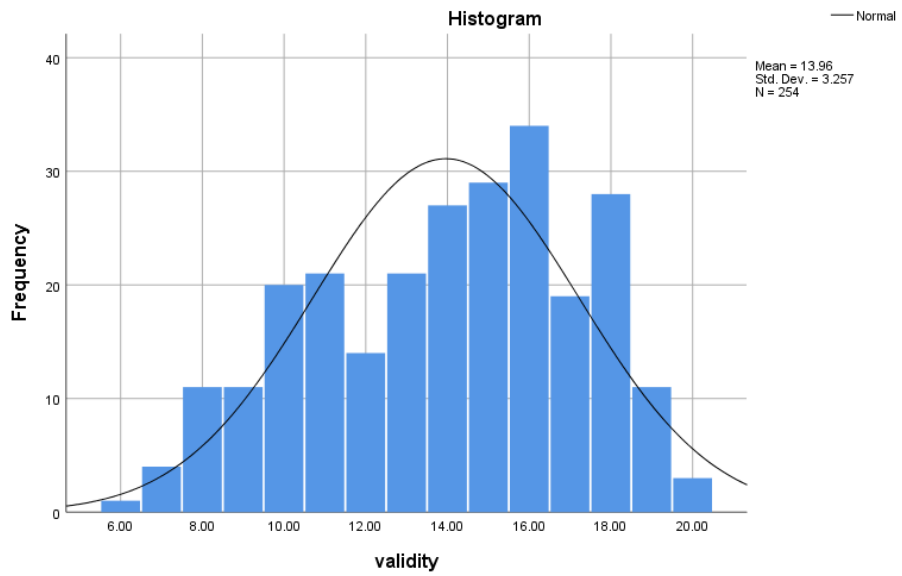


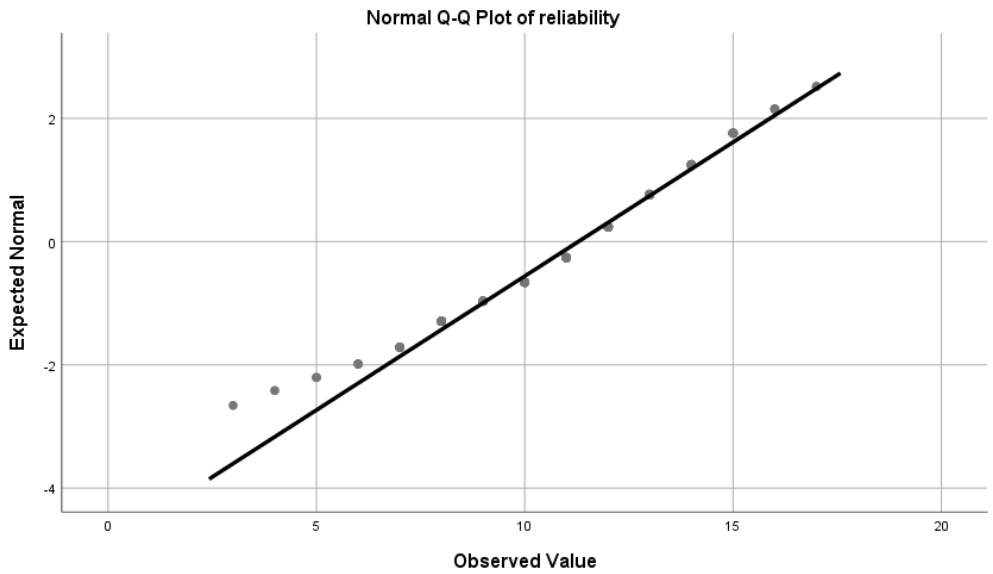
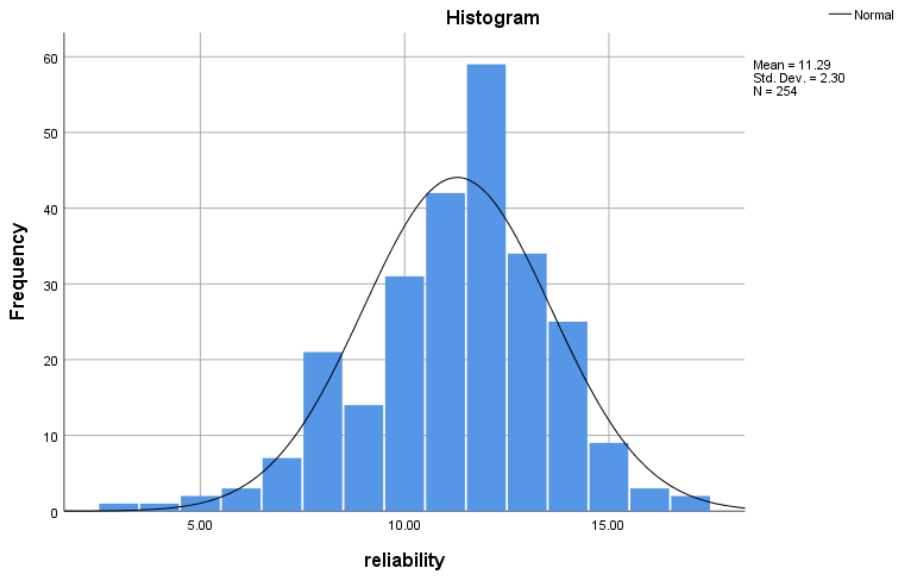
38. The part labeled I is
- a) base
  - b) collector
  - c) Emitter
  - d) Resistor
39. The function of the part labelled II is to supply
- a) protons
  - b) voltage
  - c) negative charges
  - d) positive charges
40. Which of the following is a plant pigment
- a) bile
  - b) chlorophyll
  - c) Haemoglobin
  - d) Melanin

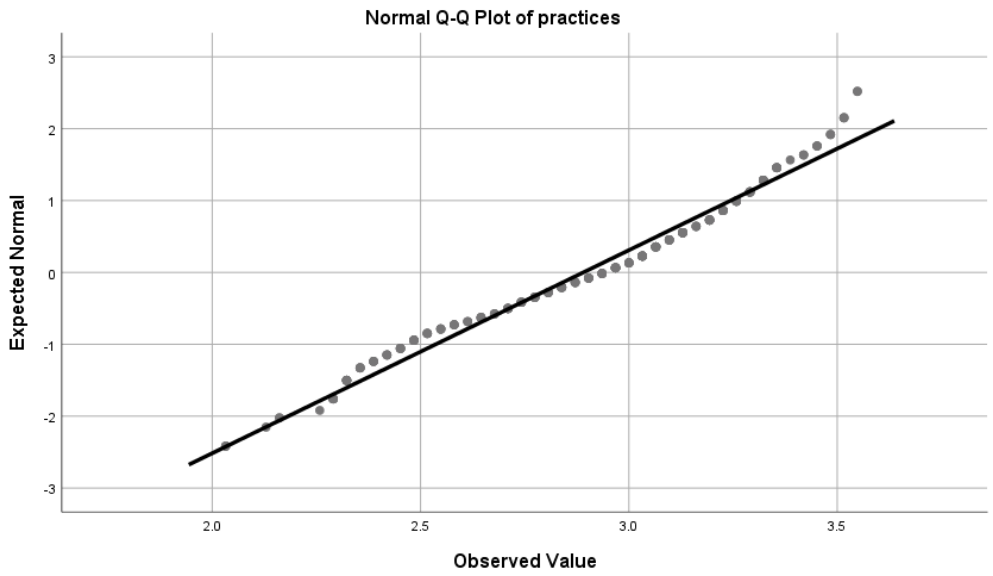
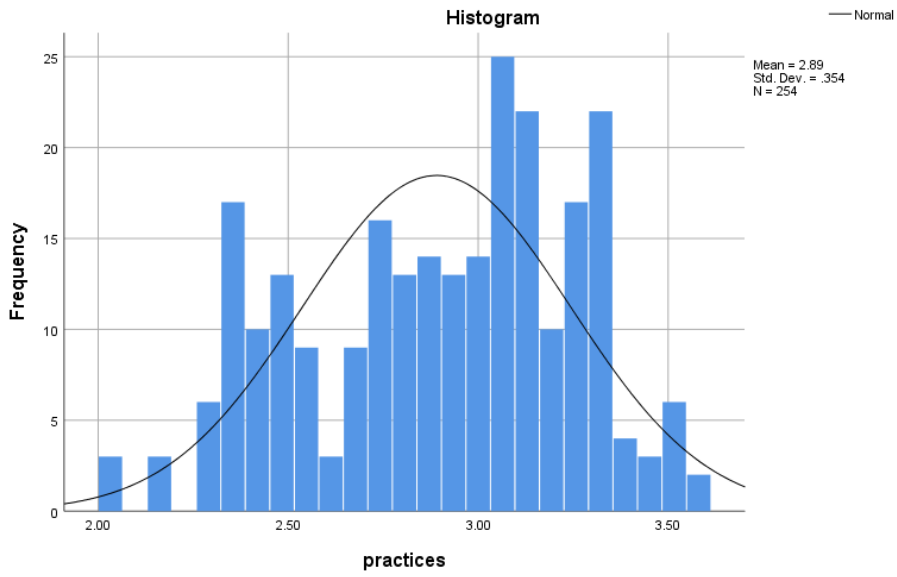
GOOD LUCK

# APPENDIX F

## NORMALITY TESTS

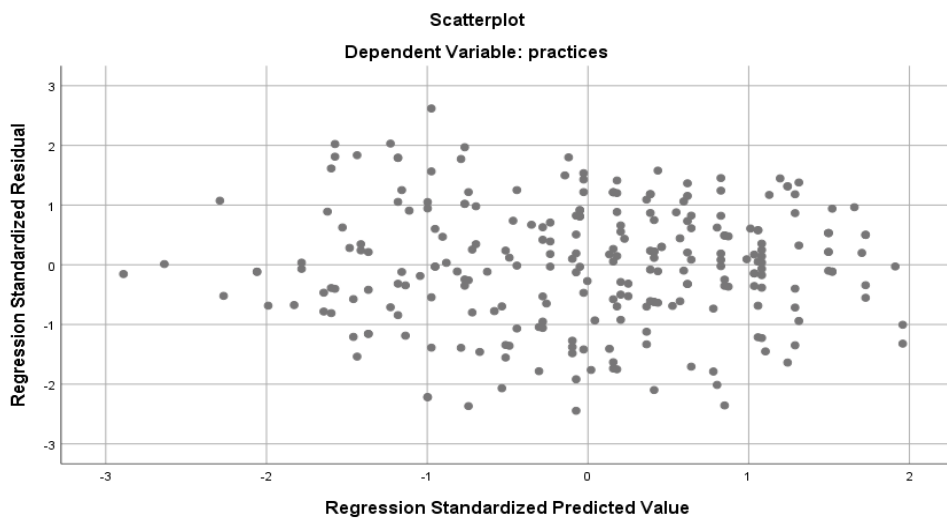
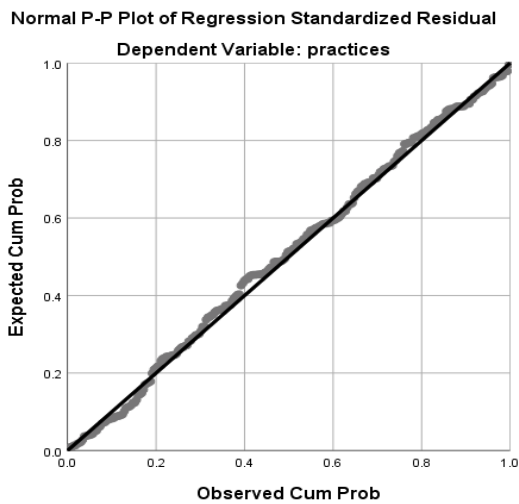
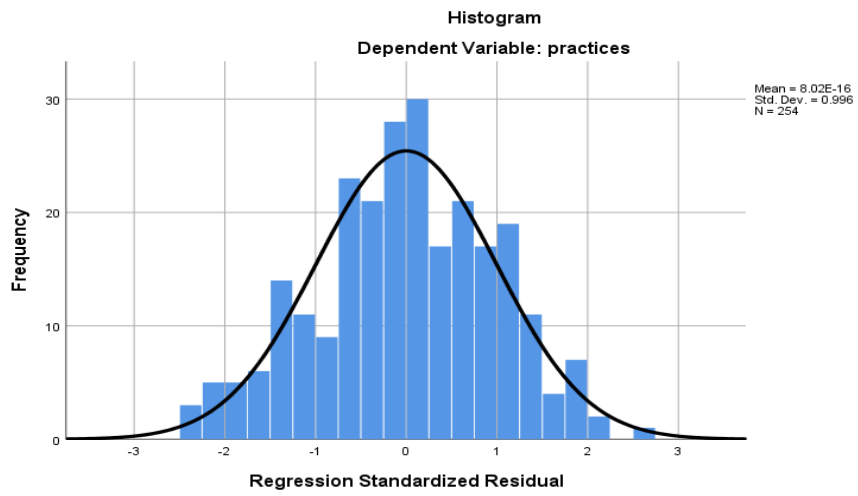






## APPENDIX G

### LINEARITY, HOMOSCEDASTICITY, NORMALITY OF RESISUALS





## APPENDIX H

### INTRODUCTORY LETTER

UNIVERSITY OF CAPE COAST  
COLLEGE OF EDUCATION STUDIES  
FACULTY OF EDUCATIONAL FOUNDATIONS

### DEPARTMENT OF EDUCATION AND PSYCHOLOGY

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



UNIVERSITY POST OFFICE  
CAPE COAST, GHANA  
5<sup>th</sup> March, 2020

Our Ref:

Your Ref:

**TO WHOM IT MAY CONCERN**

Dear Sir/Madam,

**THESIS WORK  
LETTER OF INTRODUCTION  
MS. ABIGAIL ABA TAWIAH MENSAH**

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

Ms. Mensah is researching on the topic:

**“TEACHERS’S KNOWLEDGE OF VALIDITY AND RELIABILITY OF CLASSROOM TEST: EVIDENCE FROM SENIOR HIGH SCHOOLS IN SEKONDI-TAKORADI METROPOLIS.”**

She has opted to gather data at your institution/establishment for her thesis work. We would be most grateful if you could provide her the opportunity and assistance for the study.

Any information provided would be treated strictly as confidential. We sincerely appreciate your co-operation and assistance in this direction.

Thank you.

Yours faithfully,

Theophilus A. Fiadzomor  
**Principal Administrative Assistant**  
For: **HEAD**

APPENDIX I

ETHICAL CLEARANCE

UNIVERSITY OF CAPE COAST  
COLLEGE OF EDUCATION STUDIES  
ETHICAL REVIEW BOARD

UNIVERSITY POST OFFICE  
CAPE COAST, GHANA



Our Ref: CES-ERB/UCC-EDU/14/20-42  
Your Ref: .....

Date: 22nd June, 2020

Dear Sir/Madam,

ETHICAL REQUIREMENTS CLEARANCE FOR RESEARCH STUDY

Chairman, CES-ERB  
Prof. J. A. Omoshe  
jomotoshe@ucc.edu.gh  
0243784739

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

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

The bearer, Abigail Tawiah Mensah, Reg. No. EF/mep/18/0014 is an M.Phil. / ~~Ph.D.~~ student in the Department of Education and Psychology..... in the College of Education Studies, University of Cape Coast, Cape Coast, Ghana. ~~He~~ / She wishes to undertake a research study on the topic:

Teachers' knowledge in validity and reliability of classroom assessment: The case of Senior High School teachers in the Sekondi-Takoradi Metropolis.

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

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

Thank you,  
Yours faithfully,

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