UNIVERSITY OF CAPE COAST

PERCEIVED SOURCES OF TEST SCORE POLLUTION IN JUNIOR HIGH SCHOOLS IN THE NORTHERN REGION, GHANA

UMAR SAFIANU ABUKARI

2021

UNIVERSITY OF CAPE COAST

PERCEIVED SOURCES OF TEST SCORE POLLUTION IN JUNIOR HIGH SCHOOLS IN THE NORTHERN REGION, GHANA

BY

UMAR SAFIANU ABUKARI

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

MAY 2021

DECLARATION

Candidate's Declaration

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

Candidate's Signature……. Date……

Name: ……

Supervisor's Declaration

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

Supervisor's Signature…… Date……

Name: ……

ABSTRACT

The impact of assessment results on teaching and learning development cannot be overstated. The performance of teachers, schools, and education systems is measured using test results. However there are concerns regarding dissatisfaction with achievement tests scores, misuse, and overuse of test results, high pressure to produce high test scores, high stakes nature of many tests uses, test scores used to determine educational improvements. The study assessed the perceived sources of test score pollution in Junior High Schools in two selected districts in the Northern Region, Ghana. A quantitative approach using descriptive survey was used for the study thus, the use of questionnaires to elicit responses from selected teachers. Descriptive statistics (means and standard deviations) and inferential statistics (Independent samples t-test) were used to analyze the data. A total of 265 teachers were sampled from the two districts (Yendi and Saboba). The findings confirmed the main sources of test scores pollution in the districts as teacher factor, situational factor (test administration), and external factor (parents and community). The study further revealed that the incidence of test scores pollution was higher in private schools than the public schools. Again, test scores pollution was revealed to be higher in Yendi municipality than in Saboba district. It was recommended that government should take steps to reduce the high stakes nature of standardized examinations in the country and again provide appropriate in-service training to Junior High School teachers on testing practices to help curb the situation where teachers spend much time in class in preparing students for tests to the neglect of meaningful learning.

KEY WORDS

Reliability

Validity

Test Score Pollution

Teachers Factors

Test administration

Parents Factor

ACKNOWLEDGMENTS

First and foremost, I want to express my gratitude to God, the most high for allowing me to complete this research. My heartfelt gratitude also extends to Dr. Kenneth Asamoah-Gyimah, my supervisor for the sacrifices, fatherly love, encouragement, guidance, and support throughout this study as well as the entire M. Phil programme. I am highly indebted to Dr. Ruth-Brew Annan for her expert advice and suggestions during this research work. I have learned a lot through my interaction with her.

My gratitude also goes out to all of the lecturers in the Department of Education and Psychology, particularly Professors F. K. Amedahe and Y. K. A. Etsey, for their various contributions to my completion of the M. Phil programme.

Finally, I want to express my gratitude to my wife, Mrs Memunatu Abukari, who has always believed in me and supplied me with the essential encouragement and faith in my capacity to finish what I had started.

DEDICATION

To my mum

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Background of the Study

Pupils' characteristics and behaviours that influence learning can be observed using several techniques (Amedahe, 2000). He further explained that Teachers can examine problem-solving skills, listen to students' responses to questions or remarks to detect their challenges, and then adjust their instruction accordingly, as well as discover probable misconceptions and provide solutions. He continued by stating that Teachers check at the conclusion of the lesson to see if the students have grasped the topics. The outcome will tell him if he should slow down, adjust his method, add additional instructional aids, go over past lessons again, re-teach a concept, or go on to the next topic. An assessment is the name for this procedure. It covers the entire range of processes for gathering data on students. According to Stiggins (1991), teachers spend a significant portion of their teaching time on assessment-related tasks. As a result, it is an integral element of the lesson. Assessment is the evidence of teaching (Akaranga & Ongong, 2013).

One of the primary instruments of assessment is testing. Tests are an important aspect of the educational assessment process. Tests play a vital function in assessing and certifying achievement. According to Anamuah-Mensah and Quagrain (1998), classroom tests have been regarded as one of the most tangible clues and the most crucial yardstick in determining the attainment of the objectives of any learning experience in the classroom.

Quagrain (1992) stated that achievement tests provide needed information for evaluation. Tests are also utilized to provide pupils with incentives and goals, as well as solutions for decision-making. This is because tests offer facts and data that help in understanding students and offer some measure of their capabilities. Tests have been the major instrument for admission to higher education and job placement in all industries since they are free of most systemic biases and give all candidates an equal chance. "Testing and measurement, because they are more standardized and objective than other assessment techniques reduce some of the inconsistency and subjectivity that influence evaluation" (Nitko, 2001, p. 7). The use of tests for selection and engagement is summative.

In formative use of Assessment, diagnostic tests are administered to students to determine their weaknesses with the view to designing remedial programmes for their academic advancement. Continuously, state institutions including the Ghana Education Service (GES), the media, and school boards evaluate school and allocate money and other resources based on the results of standardized tests. For example, GES placed 470 schools in leagues in 2003. The test is again employed in defining the curriculum and structuring teaching and learning. In this vein, tests are regarded as bases to measure the effectiveness of educational policies. Thus, tests can be put to numerous uses. As Cronbach (1979) points out: "Tests are neutral; they serve those who want to maintain society without change, and they are a weapon available to those who want to criticize present institutions and create a society based fully on merit and self-determination" (p.39).

Achievement tests are generally classified into two, thus a teacher-made test and a Standardized test. The teacher-made test is constructed by the classroom teacher while the standardized test is developed by test expects with standard guidelines for administration and scoring. The focus of this study is on the pollution of a standard test, precisely the Basic Education Certificate Examination (BECE). Standard achievement tests and their results, interpretation and information are employed by many teachers or practitioners in the field. Different uses of standardized achievement test scores were presented by Haladyana, Nolen, and Haas (1991). They are of the view that teachers, researchers, test writers, departments of education, school board members, administrators, educational laboratories, universities, language supervisors, laypersons, journalists and parents are the users and consumers of achievement test information. In Ghana, test scores from standardized examinations had been used for mainly selection purposes, but recently the interest in test scores as a measure of quality as well as accountability has grown steadily (Tamakloe et al, 2005).

For instance in Ghana, with the results from the standardized test, teachers are transferred, schools are rated, students are promoted and graduated, and scholarships are awarded (Etsey, 2005). Test scores from the standardized test in Ghana are also used in making decisions about which senior high school (SHS) is better usually through league systems (Ghana Education Service, as cited in Anane, 2015). Stakeholders such as opinion leaders (e. g., chiefs) are encouraging that students' performance is being utilized to determine rewards and punishment for schools and their employees (Anane, 2015). Instructors and other educators, as well as their pupils, will be

motivated to work harder and more effectively to improve student learning if test scores are used to reward and sanction teachers or institutions (Nichols, Glass, and Berliner, as cited in Anane 2015).

However, Skeptics warned against high-stakes testing unforeseen consequences, these effects they identified include, a threat to test scores' validity, as well as perverse (Ryan, 2004) and dishonest educational practices (Jones & Hargrove, 2003; Nichols & Berliner, 2005 as cited in Anane, 2015). For some time now, test experts have cautioned that high-stakes testing could result in unethical forms of test preparation, inappropriate teaching practices such as cheating in examinations. This could result in loss of self-esteem, narrowed content teaching from the syllabuses (Anane, 2015). As the stakes of standardized achievement test increases, it results in pressure on teachers and students to raise these scores (Haladyna, 2002).

In Ghana, The West African Examinations Council (WAEC) administers exams at the end of the educational system. Examples of the WAEC conducted tests are the Basic Education Certificate Examination (BECE) and the Senior Secondary School Certificate Examination (SSSCE). Linn and Gronlund (1995) outlined the features of standardised achievement tests, which include the following:

1. The test items are of a very high technological standard. They are created by educational and testing experts, experimented (pretested), and chosen on the basis of difficulty, discriminating power, and adherence to a strict set of standards.
2. The guidelines for conducting and scoring the test are clearly established, ensuring that all test users follow the same methods.

3. Norms based on national samples of kids (pilot tested) in the grades where the test is intended for use are provided to aid in the interpretation of the scores. The examinations are offered in equivalent and comparable formats, as well as information on the degree to which they are comparable.
4. A test manual and other supporting resources are given as a reference for administering and scoring the test, assessing its technical features, and understanding and applying the results.

However, some critics are of the view that WAEC examinations are not standardized. For example, Oduro-Okyireh (2008), argues that WAEC examinations do not meet all the standard characteristics of standardized achievement tests. Others question the validity of these examinations. For instance, Anane (2008) points out that, while many people believe that testing benefits education in a variety of ways, most tests' validity and efficacy are questionable, especially the West African Examination Council's tests for both the junior and senior high schools are subjects of increasing debate. He explained further by stating that, even though it seems not much has been done in the form of research, recent public out-cries raises the question of whether improved test score performance indicates improved learning.

Anane (2015), pointed out that, as society continues in pursuit of better schools and higher-achieving pupils, testing has become the focal point. He continued by explaining that, good test scores have become a prominent educational goal in today's culture. As a result, there is a lot of pressure, especially in schools with a lot of low-achieving children, to show academic growth and achievement through better test results. Etsey, (2005) stated that

in Ghana, testing has primarily consisted of traditional or formal exams of fundamental abilities and topic knowledge, with the results serving as a gauge to take actions such as transfer of teachers, ranking of schools, promotion of student, graduation and scholarships. Anane (2015), stated that terminal examinations conducted at the end of Basic and Senior High Schools (West African Senior Schools Certificate Examination, WASSCE and Basic Education Certificate Examination, BECE) administered in West Africa are becoming more and more high stakes, as results from such examinations are used for determining the quality of a senior high school, and for selection and placement in tertiary institutions and remedial classes respectively.

Recently, pupils who scored low marks in the core subjects during BECE are denied placement into senior high schools. The stakes are indeed high. Again, the Ministry of Education (2002) stated that high-stakes exams like the West African Senior Schools Certificate Examination, WASSCE and Basic Education Certificate Examination, BECE where results from these examinations are used as indicators of quality continue to wreak havoc on the education system. The use of test scores to take crucial national decisions including teacher transfer, school rating, student promotion and graduation and scholarships has brought about practices that pollute or contaminate test scores. These phenomena are so serious that most teachers engage in teaching students on how to pass examinations rather than encouraging meaningful learning.

The social limitations of test scores cannot be overlooked. One of the flaws opponents point out is that test result leads to rigorous grouping methods in the sense that it brings about rigid classification, also known as

categorizing, labelling, or grading. According to Worthen and Spandel (1991), classifying students can be condescending and rude, as well as damaging to kids who have low test scores. This has ramifications that could do more harm than good. They went on to say that one of the most major indictments of the test's use is that most assessments favour pupils from higher socioeconomic backgrounds than those whose socioeconomic background is lower. Even well-intentioned usage of tests, they claim, might disadvantage people who are unfamiliar with the majority culture's ideas and language. The anxiety that comes with taking exams and obtaining test results stems from a fear of the tests' degrading social connotations. In the worst-case scenario, worried examinees develop phobias that result in biological diseases such as "examination fever," "examination diarrhoea," and transient bladder weakness. For the reasons stated above, Gronlund (1976) claimed that all tests are harmful to students since they might cause anxiety and undermine their self-concept. Examinees may react to the negative consequences of testing's social implications, such as disruption of academic progress, loss of professional growth and promotion, and the shame of being labelled a non-achiever. As a result, many examinees engage in a variety of test-taking malpractices, commonly referred to as Examination malpractice. Exam malpractice is defined as intentional wrongdoing that is done in violation of official examination guidelines in order to provide a candidate with an undue advantage or disadvantage (Wilayat, 2009). This influences the test scores. Messick (2004) called these influences “contaminants”

Several researchers have exposed some ways by which test scores can be polluted. For instance, Onyibe, Uma and Ibina (2015) identifies collusion

among candidates, impersonation, 'giraffing,' inscription, and unusual behaviours within and outside the examination venues as instances of examination misconduct. Exam malpractice is caused by a variety of factors, according to Ushie and Ishanga (2016). These include a faulty value system that calls for a major pursuit of certification rather than knowledge and skills, laziness, lack of preparation or inadequate preparation for the examination, lack of self-confidence, socio-economic factors, political overtones, privatization and commercialization of education, poor invigilation, and weak parental function (lack of or insufficient examination hall). The main reasons for examination malpractices in Nigerian schools are indiscipline, non-implementation of the examination malpractices decree, and a lack of proper student monitoring during examinations (Adeyemi 2010). If the objective of the test is for selection or placement, the phenomena should not be condoned because it threatens society's integrity and can lead to freaks being selected for vital and sensitive jobs. According to Azizeh and Mansoor (2010), test score pollution has a significant impact on the accuracy of test score interpretations and casts questions on the logic of many of the test score applications.

In recent years, misinterpretation and misuse of test scores have received a lot of attention within the community of testing specialists' education. It is critical that test users exercise caution when it comes to the misuse and misinterpretation of test scores for the use in assessment, judgement, and enactment of policies their judgments may be affected. The BECE's authenticity has been questioned by senior high school principals and other concerned educators around the country., claiming that there is no association between high BECE achievement and senior high school success

(Mereku, 2000). Mostly, students pass in the BECE but get to senior high schools and their performance is nothing to write home about, some passed WASSCE but get into the higher level and their performance is abysmal.

The result is mostly withdrawal from the school for poor performance. The misinterpretations, as well as overuses or misuses of test scores, are generally called Test Score Pollution.

Three major sources of test score pollution were identified by Haladyna (1992), namely:

1. Test preparation activities.
2. Situational factors.
3. External factors.

Statement of the Problem

The standardized achievement test score is the currency of education, where it is universally recognized by the public and 'uncritically' accepted by some educators as a valid indicator of educational improvement (Haladyna, 1992). The impact of assessment results on teaching and learning development cannot be overstated (Chapman & Snyder, 2000). The performance of teachers, schools, and education systems is measured using test results. This has therefore become a global measuring instrument especially, in the field of Education.

Etsey (2005) stated that in Ghana, testing has primarily consisted of traditional or formal exams of fundamental abilities and topic knowledge, with the results serving as a gauge to take actions such as transfer of teachers, ranking of schools, promotion of student, graduation and scholarships. This

indicates the fact that the stakes of Standardized examinations in Ghana are indeed high.

As a result, there is a loud cry by test experts regarding dissatisfaction with achievement tests scores, misuse, and overuse of test results, high pressure to produce high test scores, high stakes nature of many test uses, test scores used to determine educational improvements.

In the past decades, experts have signalled that high-stakes testing could lead to unethical forms of test preparation, undesirable teaching practices like cheating in examinations. It could lead to loss of self-esteem, narrowed content teaching from the syllabuses (Anane, 2015).

Some Educators have therefore complained about the recent development. For instance, Anane (2008) explained that, while many people believe that testing gives a variety of educational benefits, most test interpretation and use, as well as efficacy, are questionable especially, the West African Examination Council's tests for both the junior and senior high schools are subjects of increasing debate. Even though it seems not much has been done in the form of research, recent public outcries have raised doubts about whether test score improvements actually indicate improved learning. Second, the Ministry of Education (2002) noted that high-stakes assessments such as the West African Senior Schools Certificate Examination continue to plague the education system since results from these examinations are used to measure Educational quality.

The high stakes nature of standardized examination has resulted in many teachers and testers training and preparing test takers to complete their tests (Chalak & Tavakoli, 2010 as cited in Anane, 2015). WAEC has

consistently reported cases of examination malpractices throughout the country. This resulted in the cancellation of some candidates' results and in some cases withholding of results.

Table 1: *Distribution of Candidates in the two Districts who had at least one Paper Cancelled in BECE*

District/Municipal	Year	Number of candidates
Yendi	2016	31
	2017	27
	2018	19
Saboba	2016	24
	2017	18
	2018	13

Source: GES office, 2020

In other cases, examination centres moved to other places. Yendi centre for private WASSCE was moved to the capital, Tamale after reports on examinations malpractices in 2013. Wulensi Senior High School in the Nanumba south district has suffered similar penalties for reported cases of examination malpractices in 2010.

Research has identified factors that influence test scores. Messick (2004) called these influences "contaminants" but did not specify the sources of them, while Haladyna (2002) specified two major sources of test score pollution:

1. Non-standard conditions under which tests are administered
2. Preparing students to take tests.

According to Haladyna (1992), there are three main sources of test score pollution: Test preparation activities, Situational factors, and External factors.

Previous research on test score pollution as identified put together the opinions of Educators and other shake-holders in Examinations as sources of test score pollution. These perceived sources included test preparation practices, Situational factors (Test Administration) and External factors (Parents and Community). The sources are barely identified in a vacuum without any statistical evidence to indicate that they contaminate test scores. The previous studies on test score pollution failed to demonstrate how these perceived sources of test score pollution polluted students' test scores. Again, previous research on test score pollution also failed to test whether the situation of test score pollution was the same in different geographical areas. This study is therefore set out to produce statistical evidence to either confirm or otherwise that these perceived sources truly contaminate test scores in Ghanaian schools. This study will also test hypotheses to compare the incidence of test score pollution in different Geographical locations (Saboba and Yendi Districts) and Educational settings (Public and Private).

Purpose of the Study

The study's purpose was to look into the perceived sources of test score pollution in Junior High Schools in two selected districts in the Northern Region, Ghana. Precisely, the study sought to:

1. Examine how test preparation practices (preparing students for a test) contribute to test score pollution.

2. Determine how situational factors (test administration or testing conditions) contribute to test score pollution.
3. Determine how external factors (Parents and Community) contribute to test score pollution.
4. Examine the effects of test score pollution on schools in two selected districts.
5. Compare the incidence of Test score pollution in Yendi and Saboba Districts
6. Compare the incidence of Test score pollution in Public and Privates schools in the two Districts

Research Questions

The study was guided by the following research questions:

1. How do test preparation practices (preparing students for the test) contribute to test score pollution?
2. How do situational factors (test administration situation) contribute to test score pollution?
3. How do external factors (Parents and Community) contribute to test score pollution?
4. What are the effects of test score pollution on schools in the two selected Districts?

Research Hypotheses

The research hypotheses below were tested at $p < 0.05$ level of significance.

1. H_o: There is no statistically significant difference in test score pollution between Saboba District and Yendi Municipality.

H_1: There is a statistically significant difference in test score pollution between Saboba District and Yendi Municipality.

2. H_o: There is no statistically significant difference in test score pollution between Public and Private schools

   H_1: There is a statistically significant difference in test score pollution between Public and Private schools.

The basis of the above comparisons was to find out whether the incidences of test score pollution were the same in different geographical areas (Saboba and Yendi) as well as different Educational settings (Public and Private schools).

Significance of the Study

This study is significant for several reasons. Firstly and more importantly, the results will help to clarify how the force to improve test scores results in conditions and practices which pollute or contaminates the interpretations and predictions made using that test scores.

Second, the findings of this research will help the Ghanaian government formulate policies on the use of standardized achievement assessment results and a multilevel approach to assessment based on the integration of quantitative and qualitative assessment techniques to reduce test score pollution.

Thirdly, the results will be useful to teachers and stakeholders concerning the misuse, misinterpretation, and overuse of test results.

Finally, it will also generate knowledge on the negative correlation of raising standardized achievement tests on its outcome, the effects on the validity of outcome and ways of minimizing test score pollution in Junior High Schools in Yendi and Saboba Districts.

Delimitations

This study was confined to only two selected Districts of the Northern Region. The researcher sought responses from trained and untrained teachers in both public and private Junior High Schools in the two selected districts. Again, the study focused on investigating the perceived source of test score pollution, thus, how test preparation, situational factors, and other factors are associated with test score pollution. Information was gathered using a questionnaire. Descriptive statistics (means and standard deviations) and inferential statistics (Independent samples t-test) were considered appropriate for use in data analyzes. The decision criterion was set at 0.05 significant levels.

Limitations

This study was confronted with several limitations. First, the analysis of the study was based on self-reported information and for that matter, the respondents (teachers) could either over-reported or under-reported their perceptions regarding test score pollution in the Districts.

Secondly, the study population was confined to some selected teachers in the Northern Region of Ghana. Therefore, the generalization of the results of the study on other populations (i.e. Districts, Regions, and Municipalities) in the country may be misleading.

Finally, data were collected at the time basic schools in the country were closed down because of the outbreak of the Covid-19 pandemic. It was, therefore difficult to get some teachers to respond to the questionnaire for this study. I ended up using Google forms to get responses from some teachers who were out of the study areas. The use of Google forms resulted in

problems such as cost of the internet, unstable networks, and in some cases lack of network on the part of the researcher and the respondents (especially those who lived in remote areas).

CHAPTER TWO

LITERATURE REVIEW

Introduction

The focus of this chapter is on a review of related literature. The review of literature gave support to the study. For clarity in presentation, literature was reviewed on the following subtopics:

Historical development of test

People who fail to learn from the past are usually compelled to repeat the errors of the past. For these reasons, in reviewing the literature on the state of the art of assessment of human cognitive abilities, it is suitable to look back on some of the forces that have moulded the development of these measures of intellectual ability with the view to getting an understanding of why they have the form and substance they do have. The attempts to measures human cognitive abilities can be traced to a time early in the history of imperial china (DuBois, as cited in Anastasi, 2002; DuBois as cited in Cunningham, 1986; Ebel, as cited in Amedahe, 2006; Flanagan et al., 2007). Flanagan et al, (2007) and Dubois (as cited in Anastasi, 2002& Cunningham,1986) because the Chinese had no hereditary aristocracy, developed a civil service testing programme as far back 220 B.C and this programme lasted for about 4000 years. It was, however, discontinued when Alfred Binet (1905) displayed his scale for measuring intelligence. Researchers have pointed out that the test covered the civil law, military affairs, agriculture, revenue, and geography knowledge of the examinee and that civil servants were tested every three

years for the purposes of initial appointments and continuance in employment (Anastasi, 1982; Cunningham, 1986). The historical development of education measurement dates back to 1840s. For instance, the rules for conductive written exams, the establishment of the examination board, practical mental tests, identification of factors of intelligence, use of objective classroom tests, scholastic aptitude tests, the development of test scoring machine, the taxonomy of educational objectives are some of the reforms up to 1960s (Smith, 2005).

In England, ability testing was adopted during the middle portion of the 19th century (Cunningham, 1986; Flanagan et al, 1997). Cunningham (1986) explained that the Chinese method of selecting government employees was used as a basis for the establishment of the Indian civil service. He concluded that the first British civil commission was set up in 1850 (p.3). Again, according to Yeboah (2017), the primary focus of educational assessment in the first half of the twentieth century was the necessity to create methods of differentiating between persons when distributing access to the scarce and unchanging economic good of education. Besides, European students were examined orally until well after the 12th century, when paper began replacing parchment and papyrus. In the 16th century, the Jesuits began using tests for the evaluation and placement of their students.

In the USA, testing began in the latter part of the 19th century, (Cunningham, 1986; Flanagan et al; 1997) Dubois pointed out that following the successful use in England of the Chinese method of selecting government employees, the method was adopted in the USA. Cunningham also noted that the first civil service was established in 1883. Formal testing in schools (paper

and pencil tests) began with the introduction of paper in the 12th century, (Cunningham, 1986).

According to Cunningham (1986), assessment using written rests was first used by the Jesuits at St Ignatio. He further explained that the development of academic tests was founded in Britain, specifically at the University of London. With its initial charter, testing leading to the award of degrees was accepted as an appropriate basis for evaluation. It is important to note, however, that, before this time, academic testing (oral testing) in school was already in place. Testing was an established auxiliary to the educational process among the ancient Greeks, according to DuBois (quoted in Anastasi, 2002). Tests were used to evaluate both physical and mental abilities. With its interweaving of testing and teaching, the Socratic method of teaching shares a lot with today's programmed learning (Oduro-Okyireh, 2008).

According to Anastasi (2002) and Amedahe (2000), European universities have relied on formal examinations to confer degrees and honours since their founding in the Middle Ages. These examinations, however, were largely oral. Test development like many other aspects within psychology and education is a product of many contributors and disciplines throughout history. Notable among the early thinkers were the following personalities; Charles Darwin (1809-1882), a trained physician and later a clergyman, published the book- The Origin of the species in1859. He was an important factor in the increased acceptance of individual differences (Cunningham, 1986; Flanagan et al; 1997). British mathematician and physicist, Sir Francis Galton (1822-1911), is generally recognized as the founder of formal testing. Golton, most important contributions were his emphasis on individual differences, which is

the cornerstone of the field of psychological measurement, his initial attempts to establish norms and standard scores, and his laying of the foundation for the development of the correlation coefficient. Credit for coining the term- mental test in 1890 however, is given to the American psychologist, James Mckeen Cattel (1860-1944). Galton and Cattel worked together to propel the field of mental testing forward in large and definable units (Anastasi, 1982; Cunningham, 1986; Flanagan et al;1997, as cited in Odokro Okyireh, 2008).

Karl Pearson (1857-1936), a brilliant mathematician and statistician who was a Galton student, deserves special mention. He developed the mathematical foundations of regression (then known as reversion), correlation, and co-variation of observable phenomena in such a way that Gaton could make inferences about unobservable phenomena (Anastasi, 1982; Cunningham, 1986; Flanagan et al, 1997). It's worth mentioning that Galton and Pearson's correlation coefficients are still utilized as the foundation for reliability and validity coefficients in educational and psychological assessment today (Oduro-Okyireh, 2008). A French psychologist named Alfred Binet (1857-1911) created the first intelligence test that measured high-level mental functioning called the Binet-Simon test in 1905 together with Theodore Simon (1872-1961) Louis Terman is credited with the modification of the Binet-Simon test in 1916 and coming out with the Stanford- Binet test which was the first well-standardised and carefully developed intelligence test. With the ongoing development in the field of measurement at the time, the use of the Stanford-Binet test as an individual intelligence test declined after the introduction of the Wechsler tests developed by Wechsler in 1939 (Flanagan et al; 1997).

Exams began in the United Kingdom with the introduction of competitive entry into public service and institutions. Mental testing began as an attempt to more precisely identify students who required special education services, and it was established in the context of "clever plus" grammar school selection (Akplu, 1989). Tyler (2000) also claimed that throughout the early half of the twentieth century, the primary focus of education measurement was the necessity to discover techniques of differentiating between persons when distributing access to the limited and unchanging economic benefit of education. During World War I, Arthur Otis (1886-1964), under the tutorship of Louis Terman in 1917, developed the first group tests of intelligence which were used to screen recruits for intellectual fitness. Arthur Otis is further credited with the design and introduction of multiple-choice and other objective-type test items (Oduro-Okyireh, 2008). It is worth noting that achievement testing in Ghanaian schools today involves the use of multiple-choice and other objective-type tests. Discoveries, innovations, and development continued in the field of educational measurement over the years and by 1945 many of the theories and principles used in educational testing today had been developed (Amedahe, 1989).

Theoretical Review

Classical Test Theory

Classical test theory, according to Hambleton and Jones (1993), is a theory regarding test results that introduces three concepts: test score (also known as the observed score), true score, and an error score. They went on to say that there are many other models of classical test theory, including the "classical test model." This is a straightforward linear model with three

components: observed score (X), true score (T), and error margin (E). According to traditional test theory, a student's Observed score (X) is made up of his or her True score and an Error margin (Hambleton & Jones, 1993). Mathematically the classical test model is represented as X=T+E. This is due to the fact that the true score is not clearly observable and must instead be calculated based on the individual's response to a set of test items. As a result, the problem is unsolvable unless some simplifying assumptions are applied. Etsey, (2005) stated that the error margin here simply refers to the portion of the Observed score that does not represent the behaviour or knowledge possessed by the learner. Unlike the error margin, the true score is the representation or a true picture of the knowledge or mastery of the content or behaviour the learner possesses (Etsey 2005). The above explanation of the true score and the error score clearly shows that the two are uncorrelated. This is one of the major assumptions of this model (Hambleton & Jones, 1993). The assumption clearly spells out that the true score and the error score virtually have no relationship. The assumption of this model is that every examinee has a true score which is obtained if the error margin does not exit (Magno, 2009). Magno's assertion is that, clearly, error margins contaminate the examinee's valid score. Other assumptions of the classical test theory include; 1. The average error score of the examinees is zero and 2. Error scores on the parallel tests are uncorrelated.

The focus of this study in relation to the classical test theory is to examine how errors from the perceived sources, thus, Test preparation practices, Test situation and External factors influence the true score of the examinee. The classical test theory explains further that the higher the error

margin, the more the observed score deviates from the true score and the more the error margin reduces, the more the examinee's observed score moves towards the true score. The true score as the representation or a true picture of the knowledge or mastery of the content or behaviour the learner possesses will inform the right decisions such as placements, promotions, selections, diagnosis among others. It will also inform the best of judgments such as 'excellent', 'very good', 'poor' among others. However, as the true score gets contaminated by the error scores, as a result of the various factors, (Test preparation practices, Situational factors and External factors, decisions or judgments made from these scores could be misleading.

Conceptual Review

Test score pollution

Test score pollution is an increase of test scores which is the result of practices and is usually designed to raise performance rating or assessment scores without upgrading the genuine performance on the attributes being tested and is associated with high stakes testing (Haladyna, 2002). It is a term used to describe test preparation practices that increase or decrease test score pollution test performances without connection to the construct represented by the test. Test score pollution is the condition that affects the validity of uses, interpretation and inferences that are made from test scores. In other words, it is any influence that affects the accuracy of achievement test scores. Messick (2004) called these influences "contaminants" but did not specify the sources of them, while, Haladyna (2002) specified two major sources of test scores pollution:

1. non-standard conditions under which tests are administered and
2. preparing students to take tests.

Sources of test score pollution

Hargett (1992) believed that the most important source of test scores pollution is disproportional importance attached to tests by commentators, editorialists based on a misconception about the role of tests and misinterpretation of results of test scores. He is of the view that the most dangerous pollution is the misinterpretation and over-interpretation of test scores which lead to many of the other sources of contamination. He suggested that standardized tests should be used with care and different methods of testing must be employed in assessments. Haladyna (1992) listed three main sources of test score pollution known as:

1. Test preparation activities
2. Situational factors
3. External factors

Conceptual Framework

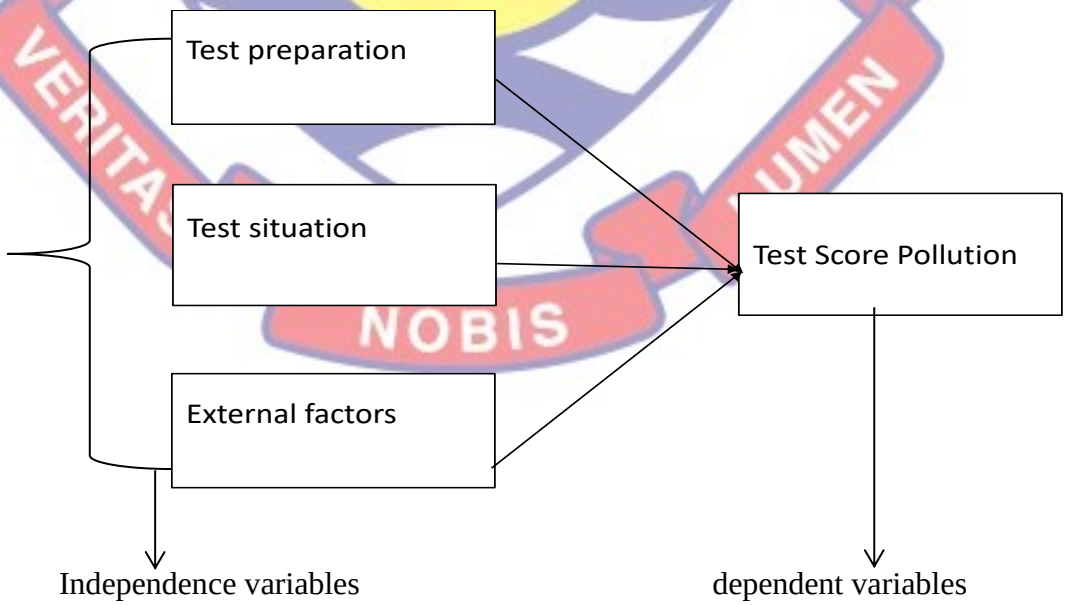

Figure 1: Conceptual Framework
Source: *Researcher's construct 2020*

Test preparation

Test or exam preparation primarily refers to education courses, strategies, and study and exam tips designed to increase students' performance on examination and standardized tests. Smith (2002) and Haladyna (1992) introduced some type of orientation toward preparing students to take high stakes achievement tests and reported some sources of test preparation:

1. Ordinary curriculum with no special preparation
2. Teaching test-taking skills
3. Exhortation (Motivation)
4. Teaching content known to be covered by the test
5. Curriculum matching (in format and content)
6. Stress inoculation
7. Practising test or parallel test items
8. Cheating
9. Test-wiseness training
10. Increasing motivation
11. Changes in the instructional programme
12. Specific inappropriate instruction (scoring high)
13. Excusing low achieving students from taking the test

Among the factors listed above, test-wiseness, presenting items identical, similar parallel to those on the test, and scoring high are the most important type of orientation toward preparing the students to take tests and many researchers have discussed the how of providing test-wiseness instructions to students at all levels (Maylone, 2004; Hoover, 2002; Volante 2006). As a cognitive capacity, test-wiseness can be characterized as a set of

strategies that a test taker can employ to improve his or her result. Dreisbacha and Keogh (2002) investigated the effects of test-taking skills training on the performance of young Spanish-speaking children from low socioeconomic backgrounds on readiness tests. Their findings backed up the theory that test-wiseness has a significant impact on such youngsters and should be taken into account by assessment software.

In their studies, Dolly and Williams (1983) demonstrated that it is possible to teach cognitive strategies which are part of test-wiseness and these strategies maximize test scores. Petty and Harrell (1977) looked into the usefulness of a cognitive approach to anxiety, motivation, and test wiseness, using programmed texts to condition linguistic repertoires related to each of these issues. They looked at the impact of motivation, anxiety, and test-wiseness training on group IQ and test performance, and discovered that there was a substantial effect on scores. They concluded that test-wiseness training can taint and modify the results of the scores. Stricker (2003) looked into test-wiseness on self-reported personality scales, applying accuracy measures in estimating the frequency of endorsement of personality items, estimating their social desirability, and identifying and "keying" items that rated the same factor, as well as indicators of capacity to alter scores on standard personality measures when they were presented with wit. Oakland (2001) attempted to prove that many young children lack the skills required to take standardized examinations and, as a result, do poorly. He identified some prerequisite abilities and designed curricular materials and increased the test-wiseness of them. For six weeks, he worked with the materials twice a week. His experimental group saw a considerable improvement in total score.

In another study, Genshaft and Kirwin (1990) examined different study skills and concluded that effective study skills instructions can help learners to raise test scores. According to Nolen, Haladyna, and Haas (2000), more than 60% of teachers questioned instructed and prepared their pupils for tests. This test preparation was deemed ethical by Haladyna (2002), who asserted that comparing those who teach test-taking skills against those who do not teach test-taking skills results in test score contamination. He dubbed it "exhortation" because he included advice on eating and sleeping before the test, principals' announcements and words of encouragement, and other tactics aimed to "motivate" students to do their best on the "exam."

Practising on actual test items or a similar form, parallel or identical form may also affect the result of the scores obtained through the test. Students may receive answers, hints to answer. These factors may affect the test score and are considered test contaminants. Test preparation activities may be ethical (such as training in test-wiseness skills, checking answer sheets to ensure that each has been properly completed, increasing motivation) or non-ethical (such as scoring high, presenting items similar, identical, or parallel to those on the test), according to Haladyna et al. (1990). However, the most important point is that test preparation activation. Flippo and Caverly (2008) believed that “test’ wiseness is a meaningful but often misunderstood concept of psychological measurement” (p. 205)

Situational Factor

Situational factors refer to all those factors that are specific to the organization and administration of the test and their situations. They are factors affecting test scores, their inferences, interpretations, and validity.

Factors such as test anxiety, motivation, self-esteem, inhibition, stress, fatigue, concentration, attention, interest, setting, policies of the school administration, location, and the examiner effect are among the factors that have an impact on test scores. These factors are specific to the administration of the test and may pollute or contaminate test results or interpretations that are made from test scores. Among these factors, motivation, test anxiety, and stress seem to play more roles. Such factors are referred to as socio-psychological and strategic characteristics of the test takers by scholars in the field (e.g. Gardner & Moorcroft, 2005; Dornyei & Schmidt, 2001 Bachman, Cushing & Purpora 2003). For instance, Bachman et al. (2003) developed some language questionnaires (LLQs). The major goal of constructing such questionnaires was likely to find out the effects of socio-psychological characteristics (regarded as irrelevant factors) on test-takers ' performances on language tests. The motivation, attitude, effort, and anxiety were all measured using the socio-psychological questionnaire battery. Through a validation study of questionnaires to examine personal factors in L2 test performance, Purpora (2004), used structural equation modelling, captured the components of motivation as instrumental, achievement, and anxiety such as class, test, and language anxiety. Purpora (2004) claimed that “these socio-psychological and strategic factors may have a significant impact on test scores” (p. 93) such influences may result in the misinterpretation of information obtained from test scores and lead to test scores pollution.

Furthermore, both teachers and researchers agree that motivation is one of the most important elements influencing the rate and success of second/foreign language learning. Lambert (2003) presented a "social-

psychological model" that highlighted cognitive and affective elements such as language aptitudes and intellect, as well as attitudes and motivation. Gardner and Moorcroft (2005), defined L2 motivation as “the extent to which individual words or strives to learn the language because of a desire to do so and the satisfaction experienced in this activity” (p.10). Motivation is divided into three categories: intensity of motivation, desire to learn the language, and attitude toward the act of learning the language.

Gardner and Moorcroft (2005) looked at how language ability, integrative motivation, and method of presentation affected learning speed. The findings revealed that all three parameters had an impact on learning speed. Subjects with strong language aptitude learnt quicker than those with low language aptitude, those with high integrative motivation learnt significantly quicker than those with low integrative motivation, and learning was faster under visual/written settings than under aural/or alone.

Wu Man-Fat (2004) investigated the link between integrative motivation and L2 achievement among Chinese L2 students in Hong Kong. He claims that a better knowledge of the relationship between interaction motivation and L2 linguistic accomplishment in Hong Kong enables instructors and curriculum planners to enhance teaching practices and policies, resulting in higher test scores. Konig (2006) studied the orientation, motivation, and attitudes displayed by the second foreign language learners in Turkey and found that the informants’ instrumental orientation was very strong. Therefore, they learned a foreign language to get a job, a better position, for graduate studies either at home or abroad, or to realize objectives in life. Their motivation was high and they enjoyed learning the language.

Their attitude was also positive either toward the language or toward the speakers of the foreign language. These studies show that the orientation, motivation, attitude, and aptitude of learners have effects on the scores obtained by learners, on their general achievements, and rates of learning.

As observed in a validation study by Purpora (2004) mentioned above, test anxiety was deemed as an important situational factor and has received a great deal of attention. Test anxiety can affect the results of tests and scores. Test anxiety was mentioned by Haladyna et al (1990), who estimated that over 25% of the school-age population suffers from some form of the disorder. They believed that test anxiety could be treated, but that it could also stem from stressful conditions in the classroom. Test anxiety may increase because of many factors, such as the level of difficulty of the material, teachers' impact of the test results on students' lives and future, the conditions under which tests are administered, parental encouragement, expectations, etc. Some studies showed that test anxiety may be promoted through parental restrictions, blame, overprotection or stress may increase which in turn could affect their results of tests (Mine, 1987; Filson & Brown, 2008; Birenburn & Pinku, 2007; Cassaday & Johnson, 2002).

The time limit may also pollute the results of scores. Timed testing compared to extending time limits or increasing or providing extra time may affect the results obtained through tests (Cheryl & Robin, 2000; Kenworthy, 2006). On the other hand, teachers' attitudes towards teaching, tests, and even students may affect the results of the tests. It can have an impact on students' achievements (Wagaman, 2009). This factor as a situational factional factor may increase or decrease the performance of the students. Standardized tests

may create some kind of discouraging climate for teachers, which in turn affect the profession and their attitude toward the test and consequentially may contaminate the results of the test. Learners' perseverance, in the classroom, may affect their L2 achievement and their test results.

External Factors

External Factors refer to any factor rather than test preparation and situational factors, such as family and home influence. The social, economic, ethical, the cultural status of the families, the level of education of parents, their income, conditions of living, mobility from the learning environment, attitudes of the parents toward education and learning, inadequate parental care regarding learners' health or nutrition are among these factors.

Learners' achievement and exam scores are heavily influenced by their culture. According to Thomas, Hertzig, Dryman, and Fernandez (2001), the ethnic background of students on a test can affect their test scores. The familiarity and attitude of their families or culture towards the target language community may also influence the test results. Social capital, support, money, and opportunities available for them may also pollute test scores' interpretations and uses. Although in reality schools, institutes, universities, and their staff and the personnel are blamed or praised for the success or failure of the students for the test results, in many instances, the external factors may influence the results and inferences of test scores. Therefore, ignoring them as probable sources of test score pollution may affect the accuracy of test interpretations and uses. Haladyana (1992) reported on the role of test score pollution in interpreting test scores.

Empirical Review

Achievement test

Achievement tests, in general, are the universal way of inferring students' abilities, feelings, attitudes, opinions, and achievements of what one has studied within a given period of instruction (Nitko, 2001). This format of assessing students is primarily "paper and pencil''.

Achievement tests are generally classified into two, thus the teacher-made test and standardized test. This teacher-made test is constructed by the classroom teacher while the standardized test is developed by test expects with standard guidelines for administration and scoring. The focus of this study is however on a standardized test, precisely the Basic Education Certificate Examination (BECE).

In Ghana, a standardised test is conducted by the West African Examination Council (WAEC) at the terminal point of the educational system. Examples of the WAEC conducted tests are the Basic Education Certificate Examination (BECE) and the Senior Secondary School Certificate Examination (SSSCE). Linn and Gronlund (1995) outlined the features of standardized achievement tests, which include the following:

1. The test items are of high technical quality. They are developed by educational and test specialists, experimented (pre-tested), and selected on the basis of difficulty, discriminating power, and relationship to a clearly defined and rigid set of specifications.
2. Directions for administering and scoring are clearly stated so that the procedures are standard for different users of the test.

3. Norms based on national samples of students (pilot tested) in the grades where the test is intended for use are provided as aids in interpreting the scores.
4. Equivalent and comparable forms of the tests are provided as well as information concerning the degree to which the tests are comparable.
5. A test manual and other accessory materials are included as a guide for administering and scoring the test, evaluating its technical qualities, and interpreting and using the results.

Educational use of test

The decisions to manage instruction, select from a group, place students in the right places, classify students bade on similar features, give counselling and guild students, and to accredit and certify are all examples of educational uses of tests, according to Amedahe and Asamoah-Gyimah (2003).

Instructional management decisions refer to the classroom decisions that the teacher takes based on the assessment results of students. The classroom teacher constantly needs to diagnose his instruction and remediate the aspects which have been defective (Amedahe & Gyimah, 2003). This is done through feedback from students to the teacher. For diagnosis and remediation, the teacher engages in diagnostic testing to identify students who need remedial. Nitko (1996), pointed out that, diagnosis involves identifying both the appropriate content and the features of the learning activities in which a student should be engaged to attain the learning target.

Another use of tests is the modelling of learning targets. According to Nitko (1996), assessment determines for students what the teacher wants them

to learn (p. 9). He explained further that students can always relate their current performance on the learning targets to the desired performance. The teacher can then teach his students to find out the ways by which their performance is related to the desired performance and how it is deficient. With this, the teacher can focus his teaching on the remediation of the identified deficiencies whiles students are also able to know what is relevant to learn once they can evaluate their performance vis-à-vis the desired learning targets.

Again, tests are used to motivate students, reward those who have prepared well in advance, and provide negative consequences for those who have not prepared well. The frequency of an individual's behaviour is increased by reinforcement. Hence, it is reasonable for one to conclude that tests influence students learning in the sense that the motivation from tests as a result of performing well can activate and direct their learning by sustaining their interest (Cunningham, 1986; Ebel & Gronlund, 1988; Nitko, 1996).

Moreover, tests are used for the assignment of grades to students' performance. The grades or symbols (A, B, C…) that the assessor, in most cases the classroom teacher reports, the formal evaluation or judgments on the quality or worth of his/her students' achievement of the important learning objectives (Amedahe & Gyimah, 2003; Ebel & Frisbie, 1996; Nitko 1996). It is important to note that assessment results constitute the most important part as it is in the Ghanaian educational system that provides the basis for the assignment of grades. Ebel and Frisbie (1991) have warned that to serve effectively the purpose of stimulating, directing, and rewarding students' effort to learn, grades must properly represent the student level of achievement (validity). The highest grades mostly go to those students who have

demonstrated the highest level of achievement with respect to the course objectives.

Selection is one of the uses of the test. For selection decisions, sometimes, an institution decides whether some persons are acceptable for specific programmes while others are not. In selection, some are accepted while others are rejected (Amedahe & Gyimah, 2003; Cronbach, 2008; Nitko, 1996). An educational institution often uses test results to provide part of the information on which selection decisions are based. Typical examples are the selection of candidates for admission into Senior High Schools (SHS) and other higher educational institution in Ghana. This is done by the use of test scores of students at the end of the Junior High Schools and University admissions in Ghana which are based on the test scores of students at the end of the Senior High Schools.

Cunningham (1986) identified the use of the test as a means of grouping children concerning their ability to profit from different types of school instruction and the identification of the intellectually retarded and the gifted. Nitko (1996) has pointed out that sometimes, based on test results, a decision is made that results in a person being assigned to one of several different but unordered categories of programmes. He further explained that these types of decisions are called classification decisions. These decisions result in either assigning students in the same classroom to different groups for effective instruction or assigning students to special education classes. Cunningham (1986) however cautioned test users about the over-reliance on tests in assigning students to special education classes by explaining that

intelligence tests are only one component of the assessment of students referred for possible placement in special classes (p.11).

One of the uses of the test is placement decisions. Cronbach (1975); Kubiszyn and Orich (2006) and Nitko (1996) explained that placement decisions are made after an individual has been accepted into an educational programme. Cronbach, Kubiszyn and Orich, and Nitko continued by noting that placement decisions basically involve using assessment results or test data to determine which programme can best suit an individual. Such decisions are characterized by assigning individuals to different levels of the same general type of instruction or education based on their ability, with no one rejected by the institution (Cronbach &Glaser, cited by Nitko, 1996). Promotion in Ghanaian schools from one class or form to another which in most cases is based on the performance in tests of the previous class is an example of a placement decision.

One cannot ignore guidance and counselling decisions as one of the uses of the test. It involves using assessment results to help students in exploring and choosing careers and in directing them to prepare for the careers they select (Anastasi, 2002; Amedahe & Gyimah, 2003; Kubiszyn & Orich, 2006; Nitko, 1996). Amedahe and Gyimah (2003) have explained that guidance is one of the student's personal services provided in a non-instructional setting to care for the needs of students including educational, emotional, and adjustment needs. Nitko (1996) and Amedahe and Gyimah (2003) have agreed with the fact and argued that due to the complexities involved in guidance and counselling decisions, test data must always be combined with other assessments such as interviews, interest inventories,

various aptitude tests, and personality questionnaire together with additional background information on students and discussed with students in a series of counselling sessions to help students make good decisions.

Finally credentialing and credentialing certification decision is another use of the test, Nitko (1996) and Amedahe and Gyimah (2003) explained that they are concerned with assuring that a student has attained a certain standard of learning and hence deserve a certificate. Credentialing and certification may be regulated by the state as done in the USA. Sometimes it is executed by an external examining body at the state level. In Ghana, certification and credentialing of students at the basic and senior high schools is done by the WAEC. With the introduction of the practice of continuous assessment as a result of the educational reforms in 1987, Ghanaian classroom teachers contribute 30% of the total marks for certification of students at the JSS and SSS levels (Amedahe, 2000; Pecku, 2000).

Testing in Schools

The education purpose of measuring intellectual abilities can be traced back to two lines of historical development, which includes: (a) testing for selection and placement and (b) assessment of educational outcomes.

Individual differences in human intelligence are the focus of selection and placement tests. It expands on Binet's concept of testing as a means of determining a person's capacity or aptitude to benefit from schooling (Glaser & Silver, 1994). Before beginning a course of instruction, selection testing is used to assess human capacities. This information allows people to be properly placed, diagnosed, certified, included, or excluded from a course. Assessments of educational outcomes, on the other hand, are designed to measure the

outcomes of a course of study in terms of intended or unexpected educational consequences. The distinction between these two is not always evident. This is due to the fact that the outcomes of testing and evaluation are frequently used to make decisions concerning the next set of educational experiences (Glaser & Silver, 1994).

Conversely, selection testing is used to predict learning success. In the selection test, predictive validity is crucial. Assessing educational outcomes, on the other hand, tries to describe the type of performance that occurs as a result of learning. This means that content validity and the kind of acquired performance are the most important factors to consider. Selection tests are intended to identify abilities that develop over time as a result of educational and background experiences gained both in and outside of the classroom. The assessment of educational outcomes aims to directly measure school achievement. People can better comprehend their current positive or harmful influences by looking at their previous experience in school settings with selection testing and assessment of educational outcomes (Dietel, Herman & Knuth, 2002).

According to DuBois (2000), proficiency testing to evaluate competence for government employment is said to have started in China around 2200 B.C. The system was steadily perfected over millennia, and despite a focus on literary rather than administrative talents, it became a model for attempts to standardize civil service combat in Europe and the United States during the nineteenth century (McArthur, 1987). In 1905, the Chinese, ironically, abandoned their civil service test system just as it was becoming widely replicated overseas (DuBois 2005).

Selection testing increased in popularity in Europe, just as it did in the United States, in the first part of the twentieth century. The pioneering work of Binet in the establishment of an intelligence measure had a major influence on the creation of selection testing. Binet's significant contribution was the use of tasks that were more similar to those that would be faced in everyday life than the elemental answers, such as discrimination and reaction speed, that Galton, Cattell, and a few other researchers had used to study individual differences in human mental functioning (Carroll, 2008). Binet had been collecting data on individual differences for roughly 15 years by 1905. His papers show that he was unable to come up with a suitable definition of intelligence, but that he remained convinced of the necessity for a tool to assess this attribute (Curtis & Glaser, 1981). His effort in designing an intelligence test paved the way for testing to be used to control the sorting of people for a range of societal reasons. Educational environments provided substantial application contexts for selection testing for much of the twentieth century (Glaser &Silver, 1994).

Binet's groundbreaking work on intelligence evaluation was directly linked to a practical, educational purpose. In the early 1900s, France's minister of public education wanted to ensure that construction resources were not wasted on youngsters who would struggle to learn. Binet and his colleague Simon devised tests (based on the technique used in developing the Binet IQ test) for use in Binet's experimental school in Paris in the early 1900s to identify students unlikely to succeed in regular classrooms and hence in need of special education (Wolf, 2008).

The introduction of compulsory schooling in the United States of America put further strain on the educational system. Selection testing

provides a way to relieve some of the system's stress. By 1926, compulsory education regulations had resulted in a four-fold rise in the percentage of high school-aged pupils who attended school, compared to 1910. (Pintner, 2003). Another significant source of strain on the educational system was the necessity to deal with America's growing population of foreign immigrants (mainly Europeans). The law of obligatory schooling suggested "Americanization" in a country where the population contained a greater proportion of non-native-born individuals. This became one of the most important tasks of education. Southern and eastern Europeans presented fresh challenges to educators and the country as a whole in the first quarter of that century. Concerns were voiced about changes in American culture, and labour organizations became concerned about excess labour as new waves of immigration rose. As a result, tests provided scientific evidence to back up their fears. When compared to their native-born counterparts, immigrants did badly on the test, raising concerns about a decline in national IQ (Pintner, 2003).

Differences in levels of ability and aptitude among kids continued to hunt instructors as the school population grew and diversified as a result of compulsory schooling and immigration in the United States. Learning disabilities of various etiologies were proposed as explanations for the disparities in educational achievement, and school systems began to categorize students based on test scores. Testing to identify children with special needs is increasingly becoming standard practice in education. For those in the late 19th and early 20th centuries who wanted to utilize examinations to construct the one best system of education, testing presented an easy and strong tool of

social control (Tyack, 1976). Even the most progressive educators had to like the idea of individualized instruction to meet the needs of various learners, as Wolf, Bixby, Glenn, and Gardner (2002) pointed out. The ability to respond to variances in students' accomplishments and learning rates aided the widespread acceptance of this type of training. The use of tests for selected purposes was thought to be a key aspect in effective instruction management.

Although educational quality and social justice were frequently cited as reasons for the extensive use of testing to choose and sort kids for educational opportunities, many people afterwards claimed that these practices were based on racial, ethnic, and gender politics (Gould, 1974; Mercer, 2000).

Despite some opposition to the broad use of tests to decide educational possibilities, early in the twentieth century, testing for selection and placement became an institutionalized practice in schools across the United States and other sections of society. Selection testing was employed in primary and secondary schools, as well as in universities for university admissions, government civil service positions, and military career placement (Glaser & silver, 1994).

In a study of 50 years of testing dispute, Cronbach (1975) subsequently reflected on the setting as follows: Wiliam James had advised psychology that writing a guy's biography in advance was not the best way to understand him, but the tests came very near in estimating how much schooling a man could utilize and what vocations thrive in. More seriously, when exams were used to determine who would be placed in the “fast” part of an early grade, exams were used to determine destiny. (Page 11).

Unlike selection testing, which was established based on transparent, if limited, concepts of aptitude and intellect, the philosophy guiding school success evaluation is less apparent. In the past, techniques for assessing success and competence growth tended to rely on psychometric technology, which arose in the domain of selection and aptitude testing. As a result, suitable psychological theories of human competence and performance, which are required for the assessment of achievement, have been lacking in achievement testing (Glaser &silver, 1994).

Significant advancements in assessment have been made in recent decades, both in terms of technology and underlying philosophy. As demands for content validity have grown more insistent (e.g., requests for diagnosis and mastery testing, national assessment and local accountability, and data that describe rather than rate learners' accomplishments and capabilities), technology has improved slightly (Cronbach, 2000). The concept and method of achievement assessment have developed through concepts such as criterion-referenced testing and anchor-point performance reference, and more recently, authentic assessment, portfolio procedures, curriculum-embedded assessment, and analyzing of the subject matter's cognitive process requirements have all been considered.

The underlying psychological theory has progressed from behavioural theories of the mid-twentieth century, which produced behavioural objectives but was unable to accurately explain complex processes of thought, thinking, and problem-solving, to more mental accounts of complicated human performance, laying the foundation for a performance measurement theory and psychometrics (Bennett & ward, 2003; Mislevy, Yamanmoto & Anacker,

1992; Shepard, 1992). In the 1980s, a growing number of long-encouraged (Glaser, 1981) novel techniques and circumstances assessing high levels of competence and reasoning capacities in schoolchildren and adults were launched in the United States. Nonetheless, much of this work is experimental at the moment (the twenty-first century), and the most common procedures in the current national educational system's assessment of accomplishment have altered little in the last 50 years (Glaser &Silver 1994).

In Ghana, the West African Examinations Council (WAEC) was established in 1955 to start a gradual take-over of the running of the overseas school certificate examination, from the Cambridge university local examination syndicate and the take-over became complete in 1960. WAEC since 1960 has been involved among other things in assisting in the selection of students to secondary schools and universities through the BECE and SSSCE respectively. They are also used in the selection of students into higher institutions including colleges of Education and technical universities. Until recently in Ghana, test scores from WAEC examinations had been used for mainly selection purposes, but interest in test scores as a measure of quality and accountability is growing steadily (Tamakloe et al, 2005).

High –Stakes Testing

Stricker, (2003) explained high-stakes test is one that has significant implications for the test taker. Passing the exam has numerous advantages, including admission to university education, a scholarship, and a license to practice a profession. Failure has significant consequences, such as being compelled to retake the test until it is passed, being denied admission to a postsecondary institution, or having difficulties finding work

Moon, Brighton, Jarvis & Hall (2007), underlined the high stakes in terms of the possible consequences of the compulsory test. They went on to say that these examinations were implemented as part of a policy to raise student achievement by holding teachers, schools, and students responsible for fulfilling national standards. According to them, failure to perform at an acceptable level might result in a loss of school accreditation as well as public condemnation. When high-stakes testing is taking place. The findings are utilized to make comparisons between schools and districts. Schools are ranked and branded within districts based on student test pass rates (Fuller & Johnson, 2001; Perreault, 2000), and test results are published in publications and publications, putting pressure on teachers, managers, and school boards to improve test scores (Thomas et'al, 2000). Anane (2015) pointed out that despite these increasing uncertainties about high-stakes tests, the current landscape of education in Ghana prominently features high-stakes testing. He went further to post the following questions, "it is working?" "Does it increase students learning?" "How do teachers respond to high–stakes testing?" He, therefore, concluded that an increase in test scores does not mean familiarity with the test and prepping of students.

The Debate on High-stakes testing

Those in favour of high-stakes testing believe that government-determined and legislated education standards are vital for improving kids' learning across the country (Moon et al, 2007). Supporters of the high-stakes testing claim that standards-based assessments provide a platform of high-quality content and standards on which schools may build a rich, focused curriculum and promote national balance in education by exposing all students

to the same subject and holding them responsible to the same standards. Again, it is thought that when high-stakes assessments are in place, instructors and schools take student accomplishment more carefully since they are held accountable for their kids' test scores (Carnoy & Loeb, 2002; Roderick & Engel, 2001; wright, 2002 as cited in Moon et al, 2007). Proponents for high-stakes testing programs went on to explain the rewards beyond the curriculum implications. Standardized tests are thought to be effective agents of educational reform because they are fairly inexpensive when compared to measures like reducing class size or hiring teacher aides; they can be publicly mandated, making them more efficient than affecting change in individual classrooms; they can be quickly implemented, and they have visible results since they can be easily replicated (Lin, 2002; Smith & Fey, 2000 as cited in Moon et al, 2007). Policymakers and the general public, according to Haladyna, Haas, and Allusion (1998), regard test scores as valid and objective assessments of student achievement. This eliminates grading-related influences such as partisanship and partiality (Airasian, 1993; Hannaway, 2003 as cited in Moon et al, 2007). For example, a poll of over 1,000 parents of school-aged children commissioned by the Association of American Publishers (AAP, 2000, as cited in Moon et al, 2007) found that the majority of parents in the United States support standardized testing. According to the research, over 83 percent of parents believe that standardized tests provide critical information, and over 90 per cent desire access to comparison data about their children and the school they attended (Moon et al, 2007). More unforeseen consequences of high-stakes testing, according to Cizek (2001 as cited in Moon et al, 2007), imply that the focus on high-stakes testing has

produced many beneficial, unanticipated effects in schools, such as enhanced teaching abilities and content-area competence among instructors, increased attention paid to kids with special needs, and enhanced availability of information about student performance. Cizek's assertions seem to be supported by the circumstances in Texas at the time. The adjustment has been dubbed "The Texas Miracle" because it has been connected to such large improvements in the performance of minority pupils and those from low socioeconomic origins (two groups known for their low achievement) (Moon et al, 2007; Klein, Hamiltion, McCaffrey, & Stecher, 2000).

While rising test scores seemed to suggest positive developments in Texas education, sceptics disputed whether the better scores are a result of increased classroom focus on and preparation for state tests, or whether they are a reflection of qualitative improvement among students and institutions (Moon et al, 2007). Increases in test results, according to Anane (2015), are not related to true learning, but rather to students' familiarity with the test and preparation. Students' test results alone, according to statistical research, are not sufficiently accurate and valid indicators of teacher effectiveness to be employed in high-stakes personnel decisions (Economic policy institute, 2010 as cited by Anane, 2015).

Despite the aforementioned considerations, opponents of high-stakes testing continue to doubt the validity of standardized test scores as indicators of school and student progress. They doubt that events such as "The Texas Miracle" can be explained in terms of teaching children how to take an exam. If this is the case, pupils improve their test-taking skills without necessarily improving their learning abilities (Gordon & Reese, 1997; Haney, 2000;

Hoffliman, Assaf, & Paris, 2001; Klein, et al, 2000 as cited in Moon et al, 2007).

An objective examination of the situation in Texas, according to these experts, found that "the magic" could best be defined in terms of creative accounting and instructional adaptability. As a result, test scores in Texas rose, as did the pressure to remove low-achieving students from the testing vehicle. Efforts to raise scores were linked to an increase in suspensions, fails, and drop-out rates (Haney, 2000, as cited in Moon et al, 2007), as well as a small increase in the rate at which poor achievers were unable to take most examinations (McNeil, 2000). In addition to innovative accounting techniques, researchers discovered that pupils in Texas classrooms were spending an increasing amount of time studying for state exams (Goldon & Reese, 1997; Haney, 2000; Assaf, & Paris, 2001; Klein et al, 2000 as cited in Moon et al, 2007). Gordon and Reese (1997) found that teachers in Texas spent a considerable amount of class time preparing kids for the tests rather than the daily lessons that were supposed to connect to the test goals, based on their examination of an open-ended survey completed by 100 public school teachers. Special practice lessons were also included in the curriculum, according to them. Critics stated that scores were unlikely to provide an adequate measure of students' learning, problem-solving capabilities, or critical and higher-order thinking capabilities when education focused entirely on the test. A good assessment should, in theory, drive students and teachers to work hard, resulting in a more enriching learning experience. Working harder, however, does not always lead to higher test scores (Lin, 2002), and high test scores do not always indicate richer learning. According to research, students

tend to do well solely on the test for which they have prepared, with improvements failing to transfer to other kinds of evaluation (Amrein & Berline, 2003; Roderick &,2001 as cited in Moon et al, 2007). Increases in test scores tend to be transitory and artificial, implying that students who are taught to the test do not gain general knowledge about a topic, but rather a knowledge unique to a particular examination (Klein et al, 2000). The validity of state test scores as a measure of student achievement is called into question by findings like these.

Researchers have criticized high-stakes testing for its tendency to reproduce societal inequalities (Diamond & Spillane, 2004 as cited in Moon et al, 2007), its influence on the nature of school curricula, and the resultant impact of test-driven curriculum and instruction on the educational experiences of the nation (Moon, 2001; Moon, Callaghan & Brighton 2002). Critics worry that a curriculum centred on high-stakes testing systems will become restricted, focusing on topics that will likely appear on the test rather than natural inquiry and true learning.

Impact of High-stakes testing on curriculum and Instruction

Opponents of test-based accountability typically believe that it will just influence test preparation activities that focus on increasing test-specific skills at the expense of more general skills, resulting in achievement increases that aren't transferable to other outcome measures (Jacobs, 20015 as cited in Anane, 2015). High-stakes testing is likely to put teachers under a great deal of pressure to raise students' grades, because a failure on the part of students may appear to portray a teacher as having failed. Teachers are likely to view themselves as less effective and probably leave the profession or leave the

school where the pressure is enormous (Harris & Sass, 2011 as cited in Anane, 2015). Anane (2015) pointed out that, despite these increasing uncertainties about high-stakes tests, the current landscape of our education prominently features high-stakes testing. He went further to pose the following questions; "it is working?" "Does it increase students learning?" "How do teachers respond to respond to high stakes testing?" The findings in a study by Moon et al (2007) seem to reveal the answer to the last question post by Anane. They discovered a clear link between teachers' decisions on curriculum and instruction in their classrooms and the perceived pressures associated with high-stakes testing in their study. They went on to say that much of the strain and frustration felt by teachers stems from the perception that they must administer a jam-packed predetermined curriculum that does not allow for flexibility in response to student needs. Teachers frequently express their dissatisfaction with the time constraints imposed by a standardized curriculum, claiming that they are unable to investigate topics in the depth required to enhance student learning or to allow students to follow scholarly interests. Some educators are concerned that the pressure to perform well on a test affects instructional practice (Abrams, Pedulla, & Madaus, 2003, as cited in Anane, 2015) and prevents teachers from caring for students' needs that are unrelated to their test scores.

Moon, et al., (2007) found that many teachers felt that the prescribed curriculum's rigid pacing and sequencing force them to rush over material and proceed on whether or not pupils have attained a deep level of understanding. For children who are predicted to struggle, the time leading up to the state test is marked by an emphasis on drill and preparation in some institutions.

Teachers claim that officials told them to concentrate the weeks leading up to the examination on test preparation, rather than offering new material or engaging students in project work. The findings of this study imply that the strong focus on test preparation abruptly fades when the test is administered (Moon, et al, 2007).

In support of this, Anane (2015) explained that teachers may also change their teaching methods to concentrate on test content and preparation which can lead to test score inflation over time. Even though some test preparation practices are ethical, nevertheless, the increased use of national and international tests as measures for school improvements and accountability for student achievement has placed new demands on teachers to improve students' test scores. These have resulted in teachers changing their instructional methods and using most part their instructional hours to review past questions that come from WAEC's examination (Anane, 2010 as cited in Anane 2015). This situation has been partly captured in the findings in a research conducted by Wright (2002) in the USA: Everything that has to do with testing has been given priority, and nothing else has been given priority. The bottom-line question is, what is going to help them perform better on the test? Also, if it isn't going to help them perform better on the test, we don't have time for it right now (p.10).

The impact of test score pollution on test score validity, interpretation, and application

Test scores are relevant because they are frequently used to back up statements that go beyond (and often considerably beyond) observed results. The argument-based validation technique (Cronbach, 2008; Shepard, 1992)

provides a framework for evaluating assertions based on test results. The main concept is to present the proposed interpretation and usage plainly and in some detail, and then assess their plausibility.

In many circumstances, tests are planned and constructed particularly to support certain decisions concerning test takers (e.g., selection, diagnosis, and placement) by providing information about test-taker characteristics that are important to the decision. In some cases, tests are designed to evaluate characteristics that are relevant to a variety of judgments in a variety of situations. A test of communicative competence in a language, for example, could be interpreted in terms of the test taker's ability to use the language effectively in a variety of situations, and the test could then be utilized by various institutions to make various choices (Chatterji, 2003).

The fuel of measurement-driven training is a high-stakes outcome grafted to test performance. While the instructional engine is powered by the high stakes associated with test performance, the equity and justice of an individual's or institution's ward or sanction is totally dependent on the degree to which the inference, decision, or description is drawn from test performance is correct. (Madaus, p. 34, 1990).

Tests, whether they are high-stakes or not, are instruments that use a sample of questions or tasks from a topic domain to represent a larger total. The degree to which a student's performance on the sample supports an inference about whether the student knows or has mastered the larger subject is what makes a test worthwhile. The single most crucial idea in testing, preserving the measurement's reliability, is the correctness of this inference's validity. If an unethical practice artificially inflates test scores, it is no longer

possible to conclude that a high score indicates mastery of the larger content domain sampled by the test. As a result, the test's validity, or whether the test results can be understood as intended, is important (Mehrens, 2004).

Test validity refers to how relevant and meaningful an inference and any subsequent judgements or characterizations of individual students, teachers, or institutions based on test results can be (Madaus, 1990). When analyzing test validity, however, we must evaluate whether test results are useful for specific populations and use

When a testing program results in significant decisions or consequences, whether real or perceived, a process might start that corrodes the test's ability to represent the relevant domain. Any inferences drawn from a test that no longer represents the relevant area, as well as any decisions based on those inferences, are invalid. The more the influence of a testing program on such judgments, the more it distorts what it is supposed to measure. This is a perception effect that occurs when students, professors, or administrators believe that the test results will have significant implications for them as individuals or for their institution, regardless of whether or not their perception is correct (Madaus, 1990). There is a lot of pressure to "teach to the test." Although this pressure does not warrant a clearly unethical reaction, educators can choose to resist it; yet, it may impact their decisions in planning and conducting the evaluation. The influence may be mild at first, but as it expands, it can have negative consequences. Subjects and intellectual activities not measured by the evaluations may receive less attention. Higher-order thinking skills may be overshadowed by rote memory skills. Because of the assessment's format, an instruction may be forced to focus on that format,

measuring just those skills that will assist students to obtain "correct" answers within that context. Finally, and perhaps most importantly, all engaged administrators, teachers, parents, and students may feel that improving test scores is the fundamental purpose of education, rather than merely a valuable indicator of student learning.

The idea is that when the student's knowledge domain approaches that of the test's sample, focusing on the exam may taint the validity of any conclusions drawn from it about the larger domain. Finally, restricting the scope of what is taught may leave children more prepared for higher test scores, but not necessarily for the issues, they will confront later in life.

The effects of test score pollution on schools

The impact of test score contamination in the two districts chosen is concerning. Many schools, particularly Junior High Schools in the Districts, are affected by the consequences of test score contamination. Because test results are increasingly being considered as the primary indicator of how well an educational institution or program is performing, the number of passes a Junior High School receives in the BECE reflects the quality of its teachers and instruction. The lower the pass rate, the weaker the school's instruction and the more slacker the teachers are believed to be. Many parents who hold this viewpoint follow the crowd and relocate their children from school to school in search of the best school for them. Some schools risk closure due to their failure to achieve the requirements for good or high-performing schools. Some schools are also overcrowded due to the fact that they are the best or perform well. Teachers and headmasters are under pressure to raise test scores

and keep schools open by doing anything they can, including teaching to the test.

Reducing Test Score Pollution

Inappropriate or unethical test preparation or delivery, inappropriate use of the exams, and variables beyond the control of schools and their personnel compromise the validity of testing systems. "Test score pollution" has been coined to describe these impacts (Haladyna, 2002). Indeed, even good testing techniques can "pollute" the validity of testing programs, because the individuals and institutions that the tests evaluate and compare do not necessarily adhere to the same standards (Haladyna, 2002).

To preserve the integrity of the assessment process and the validity of the decisions made as a result of that evaluation, test pollution must be kept to a minimum. This document outlines various approaches to reduce test score pollution for individuals involved in the formulation and implementation of assessment policy.

Security of a Test

In order to undercut efforts to boost test scores through incorrect preparation activities, it is vital to ensure the security of the assessments being reviewed while choosing an assessment (Madaus, 1990). Tests should be maintained secure during the development phase for the same reason. Signed agreements by those with access to the tests to protect the security, limited access to the location where the test is being developed and to the test development materials themselves, collecting and destroying notes and drafts, ensuring that no copies are lost or stolen during the printing process, and

accounting for all test development materials before they are distributed are just a few of the steps that should be taken.

Preparing Students for Testing

The test administrator could also be in charge of preparing pupils for the exam. In order for students to give their best effort on an educational assessment, those in charge of administering it must provide them with basic information such as when the assessment will be given, the content and abilities that will be assessed, what the assessment will emphasize, the standard or level of performance expected, how the assessment will be scored, and how the assessment results will be used (Mehrens & Lehmann, 2002; Nitko, 2001).

The best preparation for a standardized educational achievement exam that teachers can give their pupils is to focus lessons on the local curriculum rather than the format and content of a specific test. Schools can also help students prepare for standardized tests by teaching general test-taking skills like paying attention to oral and written instructions, writing responses or marking answers neatly, and making efficient use of assessment time to complete all needed activities (Nitko, 2001).

Administering the Test

Everyone involved in a testing program wants it to be carried out with caution. Those who have an interest in the outcome must be confident in the accuracy of the data provided by the evaluation. To that aim, every effort should be made to ensure that an assessment's reliability and validity are not jeopardized during its administration. Furthermore, the need for test security

must be reinforced on a regular basis. Breach of security or willful attempts to tamper with test findings is a major issue that must be addressed.

The administration of a testing program must be consistent and secure in order for the assessment to be reliable and meaningful. Many doors may be opened to people who, in reaction to pressure, cheat or otherwise inappropriately raise their students' test scores if security is weak. Because the inferences and uses for which test developers have validated their assessments are often inherently linked to the way the test is administered, if a test is not administered uniformly, the inferences and uses for which test developers have validated their assessments may become meaningless.

Varying Assessment Procedures

According to Nitko (2001), a multi-level assessment approach in the classroom based on the integration of quantitative and qualitative evaluation techniques appears to be critical. He believes that language testers should widen their perspectives from testing to assessment, moving away from psychometric to statistical approaches and from tests to various forms of assessment procedures. Self-assessment, observation, portfolio, peer evaluation, numerous sources of information both qualitatively and quantitatively, conference and discussion based on students' performance are some of the activities he offered for implementing the multi-level approach of assessment in the classroom. It appears that taking a multi-level approach to second language assessment can mitigate some of the impacts of test score pollution and aid the process of interpreting or inferring from test scores, resulting in more equitable results.

Impact of Literature Review on this Study

Much research on the prevalence of test score pollution has been conducted globally, according to the literature review. Test score pollution is now recognized as a concern in Ghana, Africa, and the rest of the world. The researcher was able to determine which areas of test score pollution had been looked at and where to go from there thanks to the literature review. As a result, the literature review led and directed the research scope.

CHAPTER THREE

RESEARCH METHODS

Introduction

The methodology used to conduct the research is described in this chapter. The research design, study area, population, sample and sampling technique, research instruments, data gathering technique, and data analytics are all examples of these.

Research Design

The main rationale for the study was to examine the perceived source of test score pollution in two selected districts in the Northern Region of Ghana. It again sought to find out how test preparation practices, situational factors, and external factors (Parents and Community) cause test score pollution as well as the effects of test score pollution on schools in the two selected districts. A quantitative approach using a descriptive survey was adopted for this study. The quantitative approach is deemed appropriate for this study because, according to Cohen, Manion, and Morrison (2000), the quantitative approach is concerned with attempts to quantify social phenomena, collect and analyze numerical data. The focus of this study was to provide statistical evidence in an attempt to quantify the perceived sources of test score pollution in the two selected districts.

Descriptive research focuses on describing the characteristics of a phenomenon (Amedahe, 2004). According to Gay (1992), descriptive research entails gathering data to test hypotheses or answer research questions on the

current status of the study's subjects. In line with Gay's assertion, this study quantified the perceived sources of test score pollution by answering research questions and testing hypotheses. The justification of The descriptive research design is suited for the study because, according to Best and Khan (2007), descriptive research is concerned with the nature of existing situations or interactions, such as determining the nature of prevalent conditions, practices and attitudes, held opinions, ongoing processes, or developing trends. This is the main purpose of this study, thus, to collect data to answer research questions concerning the current status of test score pollution in the two selected Districts of the Northern Region. Another reason for the adoption of the descriptive research design is that it is suitable for either quantitative or qualitative research where there is the formulation of hypotheses or research questions to be tested or answered to describe situations (Amedahe, 2004).

Again, the descriptive survey allows the researcher the opportunity to select a sample from the population being studied (Ary, Jacobs & Razavieh, 1990; Gay, 1992). The descriptive study design has the advantage of frequently employing the process of randomization, which allows for the estimation of error when population characteristics are inferred from sample observation (Amedahe, 2004). Finally, in the social sciences, where huge populations are studied, the descriptive design is highly appreciated, and it is often employed in educational research since data collected through a descriptive survey represents field settings (Osuala, 2001).

With all the strengths of the descriptive survey identified, it will be misleading to say that the descriptive survey is without challenges. Fraenkel and Wallen (2000) identified some weaknesses of the descriptive survey as:

1. the challenge of ensuring that the questions to be answered are straightforward and not deceptive
2. getting people to critically and genuinely answer questions is a challenge.
3. Obtaining a sufficient quantity of completed and returned questionnaires to allow for critical interpretation is also a hurdle. In support of Fraenkel and Wallen, Osuola (2001) points out that creating a quality inquiry necessitates special attention to two essential factors: suitable sampling processes and accuracy in defining words in eliciting information. He went on to say that while descriptive research is necessary for discovering answers to issues, it isn't adequate in and of itself to give answers, and it can't offer cause-and-effect links.

   Despite the drawbacks of descriptive research noted previously, it is nonetheless regarded appropriate for this study because of the hopes it provides for reaching the study's major goal.

Study Areas

Saboba District is one of the sixteen districts that make up Ghana's Northern Region. It was once a part of the larger Saboba Chereponi District, which was formed in 1988 from the then-East Dagomba District Council, until the northern part of the district was split off, resulting in the formation of Chereponi District on February 29, 2008 (which is now part of the North East Region; the remaining section has been renamed Saboba District). The Saboba District Assembly is in the northeastern region of the Northern Region, with Saboba as its capital. The District is bordered on the north by Chereponi District, on the south by Tatale Sanguli District, on the west by Yendi

Municipal and Gushegu Municipal, and the east by the River Oti, which serves as the international border between Ghana and Togo.

The District's population is 65,706 with 32,320 males and 33,386 females (GSS, 2010). The people are predominantly Konkombas with few minority tribe groups including Ewes, Mossi, Dagombas and others. The major occupation among the Konkombas is farming and hunting which is believed to have been passed down to them by their forefathers. The traditional dance of the Konkombas is 'Kinachung' which is mostly performed by the youth during festivals, funerals and other traditional occasions. With almost 68.0 per cent of the population proclaiming Christianity as their faith, Christianity is the district's majority religion. Islam is the second most popular religious belief, with 16.0 per cent, followed by traditional religion with 9.0 per cent. The majority of the people do not belong to any organized religion (GSS, 2010).

In the area of Education, Saboba District has a total population of 216 basic schools. Out of this number, 27 are private schools while the rest are public schools. There are two secondary schools and one technical school.

Yendi is another municipal assembly among the sixteen districts that make up Ghana's Northern Region. It was a part of the former East Dagomba District Council, which was later renamed Yendi District and became an ordinary district assembly in 1988. On February 29, 2008, it was given the status of a municipality, becoming Yendi Municipal District; later, in June 2012, the western section of the district was broken off to form Mion District. It has a landmass of 1,446.3 square kilometres and is positioned in the heart of the Northern Region's eastern corridor (GSS, 2010)

The Greenwich Meridian runs through a number of communities in the municipality, including Bago, Laatam, Lumpua, Gbetobu, Gbungbaliga, and Nakpachei. The Municipality is bordered on the east by Saboba District, on the south by Zabzugu Districts, on the north by Nanumba North District, and on the west by Gushegu and Mion Districts.

Yendi Municipality has a population of 117,780 people, accounting for 4.8 per cent of the Northern region's total population of 2,479,461. Males and females share a population proportion of 50.0% each, with rural areas accounting for more than half of the population (56.1%) in the Municipality (GSS, 2010).

Ya-Na is Dagbon's traditional overlord, with Yendi serving as the traditional capital. The skin of Ya-Na has two entrances. The Municipality celebrates two festivals: the ‘Bugum' (fire) and the Damba festivals. The ‘Bugum'festival is a yearly event that residents of the municipality participate in. It falls in the Dagbani lunar month of ‘Bugum' and the Islamic month of Muharram. The first month of the Dagomba lunar calendar is known as ‘Bugum.'

Islam is the most widely practised religion in the Yendi Municipality, with more than two-thirds of the people adhering to it. Traditionalists (13.2%) are in second, followed by Catholics (13.2%). (7.2 per cent).

In the area of Education, Yendi Municipality has a total population of 296 basic schools. Out of this number, 58 are private schools while the rest are public schools. There are three secondary schools and one technical school. Yendi municipality also has two professional tertiary schools, one as a College of Education and the other as a College of Nursing.

Target Population

The term "study population" refers to the total number of instances that match a set of criteria (Polit & Hungler 1996). They went on to say that the population is the group of people on whom the researcher wants to base his conclusions after the study.

The study's target demographic included both trained and untrained teachers from both public and private Junior High Schools in the two Northern Region districts. This population was targeted because there was the incidence of test score pollution leading to the cancelling of some papers among some students in BECE in the two Districts in the years 2016, 2017 and 2018. The total population of the study was 869 teachers in the two districts. Out of this number, 339 were teachers in Private schools while 530 were teachers in Public schools representing 39% and 61% respectively. The data in Table 2 presents the distribution of the teacher population in the two selected districts.

Table 2: *Distribution of Teachers in the two Districts*

Districts	Private school Teachers	Public school Teachers	Total Number of Teachers	Percentage (%)
Yendi	205	290	495	56.96
Saboba	134	240	374	43.04
Total	339	530	869	100.0

Source: GES Office 2020.

Sample Size and Sampling Procedure

Quota-proportionate and convenient sampling techniques were used to obtain the sample. This research took place in the midst of Covid 19 and data was collected at the time basic schools were closed down. Most teachers were

not at post and were away from the study centres. A probability sampling technique will have been difficult to use in this situation. Again Quota and convenient samplings even though a non- probability sampling technique, was considered for this study because of the possibility of refusals and or difficulty in contacting respondents selected through any probability sampling techniques. Speaking in favour of quota sampling, Moser (1952) stated that sampling errors are minimal in comparison to the large and intractable non-sampling errors that occur during data collecting. On the basis of sampling error, it is a minor concern.

A quota of 151 teachers was assigned to Yendi Municipal while a quota of 114 was assigned to Saboba District. The quota was determined by the percentage of the teacher population in each district. A quota of 104 teachers was assigned to Private schools whiles a quota of 161 was assigned to Public schools. The quota was determined by the percentage of the teacher population in both Private and Public schools in the two districts. A quota of 62 and 89 teachers was assigned to private and public teachers respectively in Yendi and finally, a quota of 41 and 73 teachers was assigned to private and public teachers respectively in the Saboba district. According to Amedahe (2002) quota sampling involves sampling to find the people, but you must make sure you have the right number of people for each quota. After the quota was assigned to the districts, a convenient sampling technique was used in the selection of teachers from each of the districts. Convenient sampling involves choosing the nearest or readily available individuals to serve as respondents for a study (Amedahe, 2002). Convenient sampling was considered appropriate because at the time of data collection schools were closed down as

a result of the Covid-19 pandemic, therefore most teachers were out of the study area. Teachers who were readily available were contacted and administered the questionnaire. For those who were out of the study area, the researcher made efforts to reach out to those he could contact through Google forms.

The study's sample size was 265 teachers from the two districts chosen. The sample size was calculated using Krejcie and Morgan's sample size determination table. For a population of 850 people, a sample size of 265 is appropriate (Krejcie & Morgan, 1970).

Table 3: *Distribution of Teachers for the Sample in the Selected Districts*

Districts	Number of selected teachers	Percentage (%)
Yendi	151	56.98
Saboba	114	43.02
Total sample size	265	100.0

Source: Researcher's construct 2020

Research Instrument

A 50 item questionnaire developed by Agbenyo (2017) with a reliability coefficient of .765 was reduced to a 33 item questionnaire for this study. An adapted questionnaire was therefore used to collect data for this study. (see appendix A). The items that were removed from the 50 item questionnaire were seen not to be relevant to this study. The questionnaire that was used to collect data for this study was divided into five sections. That is sections A, B, C, D, and E. Section A consisted of bio data on respondents and Section B addressed how test preparation practices contribute to test score pollution. Section C also addressed how situational factors (Test

Administration Situation) pollute test scores in the two selected districts. Section D addressed how external factors contribute to test score pollution and finally, Section E addressed the effects of test scores on schools. Close-ended questions were considered more on the questionnaire.

On a four-point Likert-type scale, the items on the questionnaire were multiple-scored. A handful of the objects were scored dichotomously. For positive statements, the Likert scale items were scored on a scale of four (4) for Strongly Agree to one (1) for Strongly Disagree. Negative statements, on the other hand, were scaled in the opposite direction. According to Asamoah-Gyimah (2002), the Likert type scale is the most straightforward, but equally effective, way for gauging the perspectives and perceptions of teachers on an ongoing practice when compared to social-distance scales, Thurstone scales, and scalogram analyses. It is adopted to ensure effective analysis of the data even though it restricts the free expression and perception of respondents in a study.

Adapted Instrument Validation

First and foremost the questionnaire was validated by my colleagues in the area of measurement and evaluation where they helped to establish the content-related evidence validity of the questionnaire. After my colleagues had gone through the instrument, I submitted the questionnaire to lecturers of the Department of Education and Psychology who are in the area of educational measurement and evaluation and research methods for expert knowledge in the validation of research instruments. Then I finally summited it to my supervisor to go through. These helped to establish the content-related evidence validity of the questionnaire.

Again, the instrument was pre-tested in the Saboba North circuit, one of the education circuits that have not been selected for the study. This was basically to check the validity and reliability of the instrument. Respondents of the pre-test were tasked to complete the questionnaire and to provide comments or suggestions for revising any ambiguous items. They were also told to discuss openly with me any ambiguity, incoherence, or incomprehension that they experience in any aspect of the draft questionnaire. The respondents were JHS teachers of all the schools in that particular circuit. 80% of teachers in the circuit were involved in the pre-testing exercise. Schools in the selected circuit have similar socio-cultural characteristics to that of the selected schools for the study. Feedback from the pre-testing helped to revise items that were either ambiguous or appear not to serve their purpose (to measure what it was intended for). The reliability (internal consistency) of the questionnaire was determined using Cronbach's coefficient alpha. The Cronbach's alpha determined whether the instrument was suitable for use or not. The internal consistency, using (Cronbach's alpha) of each section of the questionnaire was estimated. The Cronbach's alpha obtained for section B (Test preparation practices) of the instrument was .765. For section C (Situational factors), a Cronbach coefficient alpha of .711 was obtained. For section D (External factors), alpha was .786 whiles section E (Effects of test score pollution) produced an alpha of .765 and finally, the overall Cronbach's alpha obtained was .871 (see appendix B)

Ethical Consideration

Researchers must be aware of ethical considerations, particularly in social research, which deals with personal data. In social research, moral

considerations and respect for participants are critical (Punch, 2009). Several ethical problems were taken into account in this study. Informed consent, secrecy, and anonymity were all part of the deal. One of the major topics addressed in this study was informed consent. Informed consent permits potential participants to choose whether or not to participate in the study (Seidman, 2006). It explains the requirement for participants to comprehend the goals, objectives, and potential consequences of their participation. It emphasizes that respondents have the right to withdraw consent even after it has been granted; this is consistent with Mertens (2010), who argued that informed consent originates from the participant's right to freedom. All participants in this study were given a thorough explanation of the study's purpose before they were asked to participate.

With respect to confidentiality, measures were taken to ensure that the participants' responses remained private. Participants were promised that their responses would be kept private, that no one they knew would have access to them, and that no one's identity would be revealed in the study.

The study also took into account the anonymity of the responders. Anonymity, according to Oliver (2010), is a critical issue in research ethics since it permits participants' identities to be hidden. To preserve the privacy of the data in this investigation, codes were utilized. Prior visits to selected schools were performed before data collecting began to avoid invading participants' privacy needlessly.

Data Collection Procedure

The Department of Education and Psychology and the Ethical Review Board at the University of Cape Coast provided a letter of introduction (see

appendix C) and an ethical clearance letter (see appendix D), respectively. The study's aim, the need for individual involvement, as well as the anonymity and confidentiality of respondents' responses to the questions, were all stated. The selected schools' principals were contacted and received the introductory letter. Permission was obtained for the instrument's administration. For data collection, four research assistants were chosen and trained. The research assistants were educated on how to communicate with respondents, how to clarify difficult questions to respondents, and other equally critical details that allowed the researcher to collect uniform data. The data was gathered when the covid-19 outbreak forced the closure of basic schools. As a result, respondents who were out of the study areas were given the questionnaire through Google forms. Respondents were informed about the study's goal and the process for filling out the questionnaire. I personally administered the questionnaire to the respondents who were present in the research locations, along with the assistants, to guarantee clarity on how the questionnaire should be completed. A maximum of four weeks was used for the data collection in the two selected districts.

Data Analysis

The questionnaire responses were edited and scored. The editing technique primarily focused on ensuring that respondents followed instructions accurately and that all questions were answered.

Section A (respondent bio data) was analyzed with frequency and percentage tables as well as graphical representation like pie and bar charts. For positive assertions, sections B, C, D, and E received scores of 4, 3, 2, and 1 for

Strongly Agree, Agree, Disagree, and Strongly Disagree, respectively. Negative statements, on the other hand, were scaled in the other direction.

Table 4: *Summary of how Data was Analyzed*

Research Questions/Hypothesis	Statistical Tool Used
1. How does test preparation practices (teachers) associated with test scores pollution?	Means and standard deviations
2. How do situational factors (test administration situation) associated with test score pollution?	Means and standard deviations
3. How do external factors (parents and the community) contribute to test score pollution	Means and standard deviations
4. What are the effects of test score pollution on schools in the two selected districts?	Means and standard deviations
5. There is no statistically significant difference in test score pollution between Saboba district and Yendi municipal	Independent samples t-test
6. There is no statistically significant difference in test score pollution between Public and Private schools	Independent samples t-test

Source: Researcher's construct 2020

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

The focus of this chapter is on analysis, presentation, as well as interpretation of the findings resulting from this study. The rationale for this study was to explore the perceived source of test score pollution in two selected districts in the Northern Region, thus, how test preparation practices, situational factors, and external factors cause test score pollution and the effects of test score pollution on schools in the two selected districts. The results of the research questions and hypotheses set for the study were used to conduct data analysis and interpretation. The analysis was conducted on all data collected from 265 teachers in two selected Districts in the Northern Region.

For all research questions, descriptive statistics (means and standard deviations) were used, and for the research hypotheses, inferential statistics (Independent samples t-test) were used. The demographic features of respondents are described in the first section of this chapter (teachers). The research findings are reported in the final section, which is based on the study's research questions and hypothesis.

Demographic Features Respondents

Section A of the research instrument was created to elicit the teachers' biographical information. These data include the teacher's gender, professional status, academic qualification, place of work and school status.

Pie charts, bar graphs and frequencies, and percentages were used to present the analysis.

Data on the gender of the respondents is presented in Figure 2.

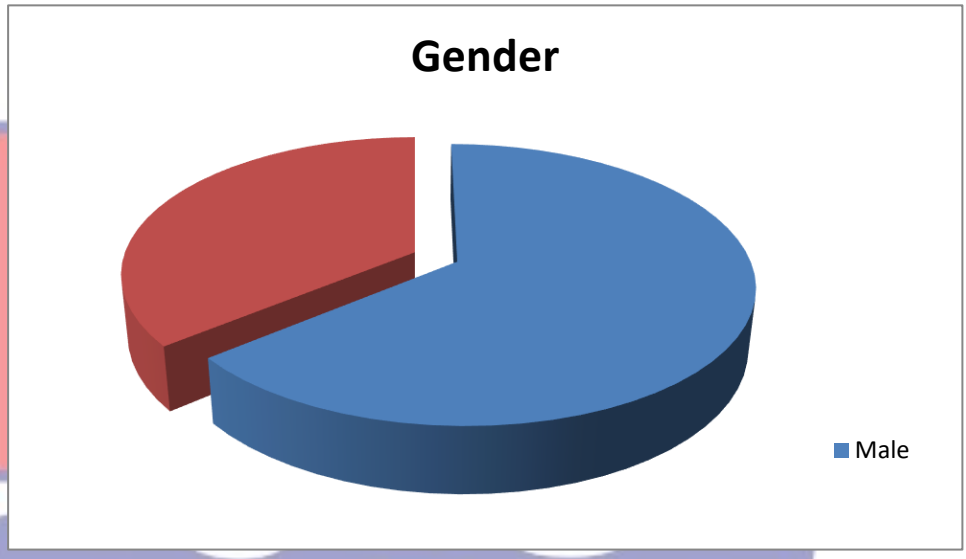

Figure 2: A pie chart showing the gender of the respondents

Source: Field survey, Abukari (2021)

From Figure 2, 170 of the respondents were male teachers while 95 of them were female teachers. This implies that there were more males teachers than their females' counterparts.

Data on the school status is presented in Figure 3

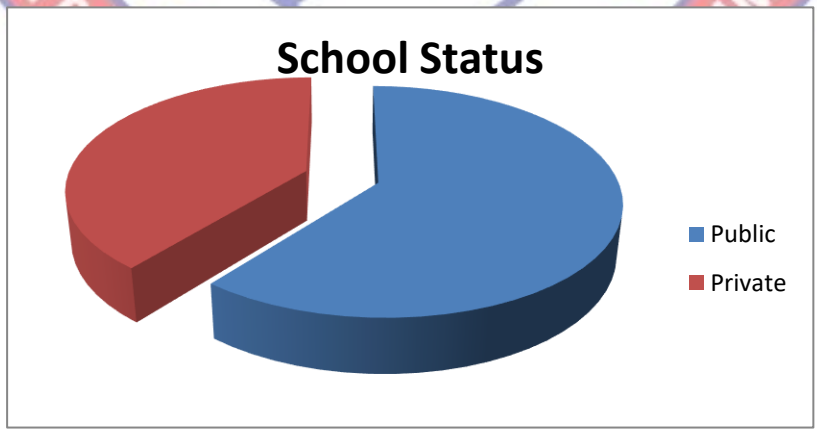

Figure 3: A pie chart showing the school status of the respondents

Source: Field survey, Abukari (2021)

From Figure 3, the Majority of the respondents were teachers in public schools while a few of them were teachers in private schools.

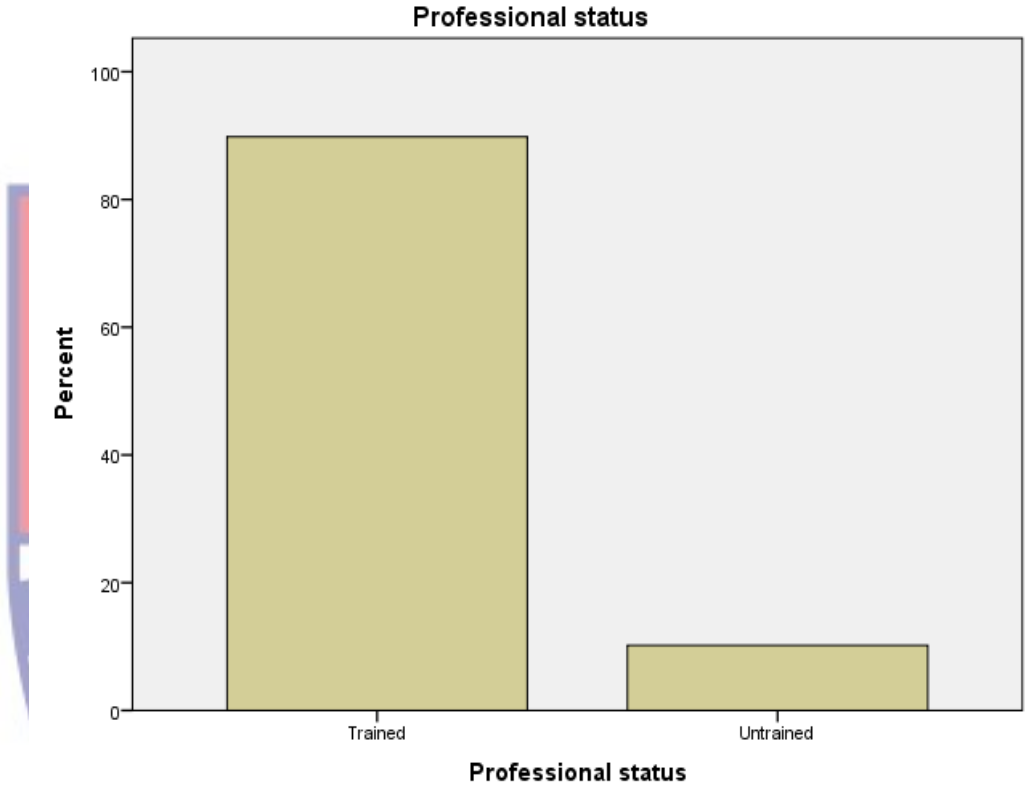

Figure 4: presents a bar graph showing the professional status of the *teachers.*

Figure 4 indicates a majority of the teachers that took part in the study were trained while untrained teachers were the minority. This indicated that most of the teachers held professional certificates for teaching and very few of them were untrained.

Table 5: *Distribution of the Teachers by their Academic Qualification (n=265)*

Academic Qualification	Frequency (Number.)	Percentage (%)
Teachers'Cert A	3	1.2
Diploma/HND	101	38.1
First degree	149	56.2
Master's	12	4.5
Total	265	100

Source: Field survey, Abukari (2021)

From table 5, the majority of the teachers, 149 were first degree holders followed by teachers with either diploma and/or HND, 101 teachers. Few teachers held a master's degree, 12, (M.Ed, M.A, MPHIL or M.Sc). Teachers' Certificate 'A' in education recorded the least number of teachers (3 teachers). This, therefore, implies that the majority of the teachers in the two selected districts are first degree holders as well as a diploma in education holders.

Analysis of the Research Question and Hypotheses

To answer the research questions and test the hypothesis, descriptive statistics (means and standard deviations) and inferential statistics (Independent samples t-test) were considered appropriate for the analyses. As a result of a four Likert scale used and the scores allocated to responses (4 for strongly agree, 3 for agree, 2 for disagree, 1 for strongly disagree, and the reverse for negative items), a mean of 2.50 and above was considered positive responses of teachers while a mean of 2.49 and below was considered negative responses of teachers.

Research Question 1: How do test preparation practices (preparing students for the test) contribute to test score pollution?

The researcher sought to find out how test preparation practices (preparing students for the test) contribute to test score pollution in the two selected Districts in the Northern Region. To achieve this, means and standard deviations were computed from the teachers' responses. On a four-point Likert scale, the teachers were asked to indicate their levels of agreement or disagreement with statements concerning how test preparation practices contributed to test score pollution in the districts. For the fact the four Likert

scales were scored 4 for strongly agree, 3 for agree, 2 for disagree and 1 for strongly disagree, the average for the four scores thus (4+3+2+1) divided by 4 will be 2.50. Therefore a mean of 2.50 and above indicated that respondents are practising the item (positive responses) while a mean of 2.49 and below indicated that respondents are not practising the item (negative responses)

The results on how test preparation practices contributed to test score pollution in the two districts are presented in Table 6.

Table 6: *Descriptive Analysis (Means and Standard Deviation) of how Test Preparation Practices Contribute to Test Score Pollution (n=265)*

Test preparation practices	Mean	Std. D
I spend more time preparing students for WAEC examinations	3.45	.595
I prepare my scheme of work based on what is included in the BECE	3.45	.589
I teach students on test-taking strategies tests	3.28	.624
I don't recommend or register academically weak (low-achievers) for WAEC examinations	2.76	.892
I give past examination questions from WAEC to my students to practice	3.34	.526
Most teachers modify the order of the curriculum based on what is mostly included in the BECE	3.20	.627
I encourage students to perform in the examination	3.58	.539
I encourage Parents and my colleague teachers to help motivate students to perform during the examination.	3.47	.544
Parents and the community edge and help teachers to increase the performance of students in BECE by any means possible	3.07	.682
Means of means/SD	3.28	.624

Source: Field survey, Abukari (2021)

From the data in Table 6, the means and standard of (M=3.45, SD= .595) showed that teachers spend more time preparing students for WAEC examinations. Again, the mean and standard of (M=3.28, SD= .624) show that most teachers teach students test-taking strategies. Also, the data in Table 6 revealed that the majority of the teachers M=3.20, SD=.627 agreed to the fact that they modify the order of the curriculum based on what is mostly included in the BECE. The mean and standard, M=3.07, SD=.682 gives statistical evidence that Parents and the community encourage teachers to help increase the performance of students in the BECE by any means possible. This was identified as one of the variables under test preparation practices that teachers engage in which lead to test scores pollution in the districts.

Research Question 2: How do situational factors (test administration situation) contribute to test score pollution?

I again sought to find out how situational factors (test administration situation) contribute to test score pollution. To achieve this, descriptive statistics (Means and standard deviation) were deemed appropriate for the analysis. Means and standard deviations were used to determine how situational factors (test administration situation) contribute to test score pollution. On a four-point Likert scale, the respondents were asked to indicate their levels of agreement or disagreement with statements concerning how situational factors (test administration situation) contribute to test score pollution in the districts. The four Likert scales used in this were scored 4 for strongly agree, 3 for agree, 2 for disagree and 1 for strongly disagree, the average for the four scores thus (4+3+2+1) divided by 4 will be 2.50. Therefore a mean of 2.50 and above indicated that respondents are practising

the item (positive responses) while a mean of 2.49 and below indicated that respondents are not practising the item (negative responses). Therefore A mean of 2.50 and above indicates a teacher's positive responses while a mean below 2.50 indicates the teacher's negative responses. The result on situation factors that contribute to test score pollution is presented in Table 7.

Table 7: *Descriptive Statistics (Means and Standard Deviation) of how Situational Factors (test administration situation) Contribute to Test Score Pollution (n=265)*

Situational factors (test administration situation)	Mean	Std. D
I observe anxiety (fear or panic) among my students during examination	3.21	.658
I mostly motivate my students to do their best during examinations	3.49	.551
I mostly try to help my students build their self-esteem in other to perform well in examinations	3.40	.588
I encourage students not to stress themselves up during examinations	3.21	.614
I often see students with foreign materials during examinations	2.85	.801
My students sometimes report other students copying from friends in the examination hall to me	2.82	.738
Means of means/SD	3.16	.658

Source: Field survey, Abukari (2021)

On the individual variables, most teachers M=3.21, SD = .658 confirmed that students are mostly anxious during standardised Examinations.

Again, a good number of teachers M=3.49, SD= .551 accepted that they mostly motivate their students to do their best during examinations. Also, the majority of teachers, M=3.40, SD=.588, confirmed that they help their students build their self-esteem in other to perform well in examinations. Again the majority of teachers, M=3.21, SD = .614, attested to the fact that they encourage students not to stress themselves during an examination. Furthermore, the majority of the teachers, M=2.85, SD=.801 confirmed that they often see students with foreign materials during examinations. On the issue of copying, the data further revealed that most students copy from friends during examinations, M=2.82, SD=.738, the overall mean and standard deviation.

Research Question 3: How do external factors (parents and community) contribute to test score pollution?

I again sought to find out how external factors (parents and community) contribute to test score pollution. To achieve this, descriptive statistics (Means and standard deviation) were considered appropriate for the analysis. Means and standard deviations were used to determine how external factors (parents and community) contribute to test score pollution. On a four-point Likert scale, the respondents were asked to indicate their levels of agreement or disagreement with statements concerning how external factors (parents and community) contribute to test score pollution in the districts. A mean of 2.50 and above indicates a teacher's positive responses while a mean below 2.50 indicates the teacher's negative responses. The result is presented in Table 8.

Table 8-*Descriptive Analysis (Means and Standard Deviation) how of External Factors (parents and community) Contribute to Test Score Pollution (n=265)*

External factors (Parents and community)	Mean	Std. D
Parents discuss BECE results at PTA meetings	3.24	.669
Parents persuade teachers to organize extra classes for their wards when the time for BECE is not getting closer	2.86	.788
Parents persuade teachers to organize extra classes for their wards when the time for BECE is drawing near	3.10	.711
Some parents encourage teachers to solve past questions from WAEC with their wards	3.03	.665
Parents mostly blame teachers for the failure of their wards	3.31	.658
Parents' attitudes mostly put pressure on teachers to try any means possible for their students to perform well during BECE	3.15	.705
Some parents reward teachers when their children get higher scores in BECE	2.87	.802
Parents sometimes want to know if there is leakage of the examination questions	2.68	.784
Parents sometimes want to know if there is the possibility of buying examination questions for their wards	2.57	.837
Some parents buy examination questions for their wards	2.73	.845
The community is always interested in the performance of students in BECE	3.32	.608
Means of means / SD	2.98	.733

Source: Field survey, Abukari (2021)

From the data in Table 8, some teachers M=2.86, SD= .788 confirmed that Parents persuade teachers to organize extra classes for their wards when the time for BECE is not drawing near. The table also attests to the fact that some parents buy examination questions for their wards. The M=2.73, SD=.845 provides statistical evidence to back the claim. Again most teachers

M=3.31, SD = .658 revealed that parents mostly blame teachers for the failure of their wards. Finally, the table also reveals M=3.15, SD=.705 that the attitude of parents mostly put pressure on teachers to try any means possible for their students to perform well during BECE.

Research Question 4: What are the effects of test score pollution on schools in the two selected districts?

As a way of achieving the purpose of the study, I further investigated the effects of test score pollution on schools in the two selected districts. Means and standard deviations were used to determine the effects. On a four-point Likert scale, the respondents were asked to indicate their levels of agreement or disagreement with statements concerning the effects of test score pollution on schools in the selected districts. A mean of 2.50 and above indicates a teacher's positive responses while a mean below 2.50 indicates the teacher's negative responses. The result is presented in Table 9

Table 9: *Descriptive Analysis (Means and Standard Deviation) Effects of Test Score Pollution on Schools in the Two Selected Districts (n=265)*

Effects of Test Score Pollution	Mean	Std. D
Poor performance of students in the BECE brings a bad image for the school	3.49	.530
Teachers in poor-performing schools are seen as not teaching effectively	3.37	.634
Schools with good performance in BECE are mostly overpopulated	3.08	.836
Schools that perform poorly in the BECE are mostly underpopulated	2.90	.856
The administrators (headmasters) of schools with poor performance in BECE are seen as not effective	3.21	.604
Students leave schools that perform poorly to schools that perform better	3.29	.508
Poor performing schools in BECE do not receive much attention from stakeholders and the community at large	3.14	.679
Means of means / SD	3.21	.663

Source: Field survey, Abukari (2021)

Table 9 presents the effects of test score pollution on schools in the two selected districts. On individual variables, it was revealed that the inability of a school to produce higher grades in BECE creates a bad image for the school M=3.49, SD=.530 confirmed that fact. The result also reveals that teachers in poor-performing schools are seen as not teaching effectively. This was evident after the recorded M=3.37, SD=.634. Furthermore, many students, at the point of completing school, leave poor-performing schools to schools that score higher grades in BECE. This is shown in Table 9, thus, M=3.29, SD=.508.

Research Hypothesis 1

H_0: There is no statistically significant difference in test score pollution between Saboba District and Yendi Municipality.

H_1: There is a statistically significant difference in test score pollution between Saboba District and Yendi Municipality.

The researcher further tested the hypothesis to find out whether differences exist in test score pollution among the two selected districts (Saboba and Yendi). To achieve this, the Independent samples t-test was considered appropriate because two independent groups (Saboba and Yendi) were involved and a dependent variable, that is scores on test score pollution measured on an interval scale of measurement. The dependent variable, thus scores on test score pollution was obtained by an accumulation of the scores of the responses by each respondent on sections, B, C and D of the instrument.

However, to perform independent samples t-test, assumptions such as normality of data and homogeneity of variances were checked. Table 10 present the normality of the data set.

Table 10 *Normality of Data (Saboba and Yendi)*

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Place of work		Statistic	Df	Sig.	Statistic	Df	Sig.
Test score pollution	Saboba	.073	114	.196	.991	114	.661
	Yendi	.071	151	.063	.987	151	.194

Source: Field survey, Abukari (2021)

For the test of normality of the data in the two districts, Shapiro-Wilk Test results as shown in Table 10 indicates that Saboba scores on test score pollution are normally distributed since the sig. value of .661 is greater than p. value of 0.05 *(p>0.05)*. Again the Shapiro-Wilk Test results as shown in Table 10 indicates that Yendi scores on test score pollution were also normally distributed since the sig. value of .194 is greater than p. value of 0.05 *(p>0.05)*.

Table 11: *Means Plot (Saboba and Yendi)*

	Place of work	N	Mean	Std. Deviation	Std. Error Mean
Total scores on test score pollution	Saboba	114	80.35	7.666	.718
	Yendi	151	82.20	6.486	.528

Source: Field survey, Abukari (2021)

The output in table 11 is a comparison of the means scores for the districts. It is shown in output that Yendi recorded the highest mean (M=82.20, SD=6.486) whiles Saboba recorded the least mean (M=80.35, SD=7.666). This implies that Yendi had a higher level of test score pollution as compared to Saboba. However, further, checks were done using Independent samples t-test to determine whether these differences in means score were statistically significant. The results are presented in Table 12.

Table 12: *Independent Samples t-test*

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	T	Df	Sig. (2-tailed)
Test score pollution	Equal variances assumed	5.725	.017	-2.122	263	.035
	Equal variances not assumed			-2.074	219.791	.039

Source: Field survey, Abukari (2021) ** Significant at p=0.05 (2-tailed)

From Table 12, the Significance value (Sig) for Levene's test is .017 which is less than the alpha or critical value of p= 0.05. This shows that the assumption of homogeneity has been violated for this sample. Hence, the sig value (p=.039) for "Equal variances not assumed" was appropriate for interpretation. The sig value (p=.039) signifies that. This suggests that the Null hypothesis thus, "There is no statistically significant difference in test score pollution between Yendi and Saboba districts" is rejected in favour of the alternate hypothesis thus "There is a statistically significant difference in test score pollution between Yendi and Saboba districts." The analysis gives reasons to conclude that Yendi (M=82.20, SD=6.486) had the highest level of test score pollution than Saboba (M=80.35, SD=7.666); t(219.791)= -2.074, p=.039

Research Hypothesis 2

H_0: There is no statistically significant difference in test score pollution between Public and Private schools.

H_1: There is a statistically significant difference in test score pollution between Public and Private schools.

The researcher again tested the hypothesis to find out the differences between test score pollution among Public and Private schools. To achieve this, the Independent samples t-test was considered appropriate because two independent groups (Public and Private) schools were involved and a dependent variable, that is scores on test score pollution measured on an interval scale of measurement. The dependent variable, thus scores on test score pollution was obtained by an accumulation of the responses submitted by each respondent on sections, B, C and D of the instrument.

Normality of data and homogeneity of variance assumptions were first checked. The results of the normality assumption are presented in Table 13

Table 13: *Normality of Data (Public and Private)*

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
School Status		Statistic	Df	Sig.	Statistic	Df	Sig.
Test score pollution	Public	.088	161	.004	.976	161	.006
	Private	.096	104	.020	.973	104	.032

Source: Field survey, Abukari (2021)

For the test of normality of data in both Public and Private schools, the sig values on Shapiro-Wilk Test as shown in Table 13 are both significant indicating that both Public and Private schools' scores on test pollution violated normality since the sig. value of .006 and .032 respectively are each less than p-value of 0.05 *($p>0.05$)*.

Table 14: *Means Plot (Public and Private schools)*

	School Status	N	Mean	Std. Deviation	Std. Error Mean
Test score pollution	Public	161	79.43	7.549	.595
	Private	104	84.45	4.891	.480

Source: Field survey, Abukari (2021)

The output is a comparison of the scores for public and private schools. It is shown in the output that private schools recorded the highest mean (M=84.45, SD=4.891) whiles public schools recorded the least mean (M=79.43, SD=7.549). This indicates that private schools had the highest level of test score pollution as compared to public schools. However, further, checks were done using the independent samples test to determine whether these differences in means score were statistically significant. The results are presented in Table 15.

Table 15: *Independent Samples Test*

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	T	Df	Sig. (2-tailed)
Test score pollution	Equal variances assumed	18.583	.000	-6.010	263	.000
	Equal variances not assumed			-6.566	262.994	.000

Source: Field survey, Abukari (2021)

From Table 15, the sig value for Levene's test is significant at .000 which is less than the alpha or critical value of p > 0.05. This shows that the

assumption of homogeneity has been violated for this sample. As a result, the sig value for "Equal variances not assumed" was appropriate for interpretation (p=.000). The sig value was statistically significant since p=.000. This indicates a significant difference in the mean scores of test score pollution between Public and Private schools. The null hypothesis - "There is no statistically significant difference in test score pollution in Public and Private schools' is rejected in favour of the alternate hypothesis -"There is a statistically significant difference in test score pollution in Public and Private schools'. In conclusion, Private schools (M=84.45, SD=4.891) had the highest level of test score pollution than Public schools (M=79.43, SD=7.549); t (262.994) = -6.566, p=.000 (two-tailed).

Discussion of Findings

This study aimed at examining the perceived sources of test score pollution in the two selected districts. The three main factors were all revealed as the sources of test score pollution in the two districts. The findings are in line with the assertion of Hargett (1992) who believed that the most important source of test score pollution is attributed to teachers, test preparation activities, situational factors, and external factors. He claimed that the most serious pollution is the misinterpretation and over-interpretation of test scores which lead to many of the other sources of contamination. Similarly, this finding goes a long way to confirm Haladyna's (1992) idea that the three main sources of test score pollution include, test preparation activities, situational factors, and external factors. As a result the conceptual framework truly reflects the findings of this study as shown below.

The final model

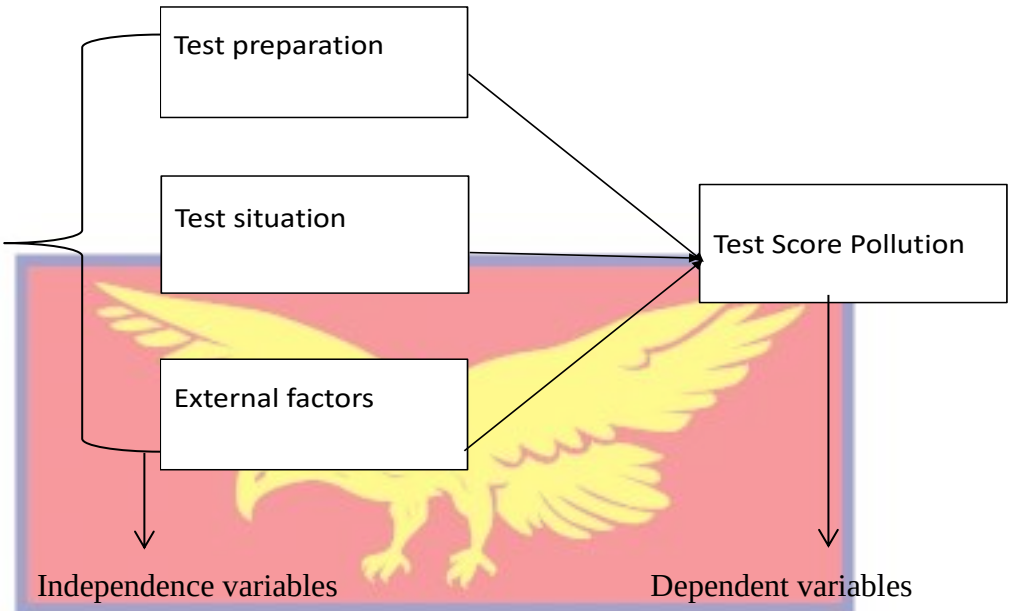

Figure 5: Observed Framework

Precisely, this study examined how test preparation practices, situational factors and external factors contributed to test score pollution. The effects of test score pollution on schools were examined as well as. The comparison of the incidence of test score pollution in both Yendi and Saboba districts as well as Public and Private were examined. The revelations of all the above are discussed below.

1. In this study, one of the objectives was to find out how test preparation practices (preparing students for the test) contribute to test score pollution in the two selected Districts in the Northern Region. The revelations supported notable research findings. For instance, this study revealed that teachers teach students test-taking strategies. Haladyna (1992) referred to these strategies as 'test wiseness'. Most writers agreed that Test wiseness is one of the serious factors that contaminates test scores. For example, Dreisbacha and Keogh (2002)

revealed that test-wiseness is an important influence for school children and should be considered in assessment programmers. They seem to endorse this practice because they consider it ethical but this still contaminates test scores. In support of this, Haladyna (2002) considered this test preparation as ethical and claimed that comparisons between those teaching test-taking skills and those not teaching test-taking skills introduce test scores pollution. This finding implies that the majority of teachers focus more on teaching students Examination writing skills. This means that pupils are taught basically how to write and pass their Examination. These learners are supposed to learn the values of life that will prepare them for the challenges in future unfortunately this is substituted for skills in passing Examinations. Again, the study revealed that the majority of the teachers agreed to the fact that teachers modify the order of the curriculum based on what is mostly included in the BECE. This finding supports Anane's (2008) study that teachers reported that testing and test results, to some extent influence their instructional planning. These, give reasons to believe that the practices of teachers in preparing students for standardized examinations contribute to test score pollution. It is a common practice that teachers basically teach topics that mostly appear in BECE to the neglect of topics that are equally important in the growth and development of the learners. The reason is just for the students to pass their examinations. Again the majority of the teachers agreed to the fact that students normally send pieces of papers and other materials that are not to be taken to the

examination hall to the exam hall. This finding agrees with Haladyna (1992) who indicated some types of orientation toward preparing students to take high stakes achievement tests and reported some sources of test scores pollution which include cheating on the part of the students. In support of the above findings, Thomas, Hertzig, Dryman, and Fernandez (2001) explained that it is a reality in schools, institutes, universities, and their staff and personnel are blamed or praised for the success or failure of the students for test results but in many instances, the external factors may influence the results and inferences of test scores. This influences the preparation given to students before a standardized examination. Similarly, Haladyana (1992) revealed that school officials and parents try to blame the teacher if the class does poorly on a standardized test without considering the multiple questions such as did the content of the test match the breadth and depth of the content taught in the classroom?

2. Secondly, this study was to find out how situational factors (test conditions and administration) contribute to test score pollution in the two selected Districts in the Northern Region. The revelations supported notable research findings. For instance, this study revealed that teachers often see students with foreign materials during examinations. This has a great influence on the performance of students in the examination. This finding agrees with Haladyna (1992) who indicated some sources of test scores pollution to include cheating on the part of the students. This is one of the commonest examination malpractices in our schools today. Students are usually seen with

foreign materials in examination halls some of them these materials include; mobile phones, programmable calculators, written notes on papers, portions of textbooks among others. The use of these materials brings about to students scoring marks that they do not deserve. This increases the error margins of their test scores leading to contamination of the results. Therefore judgements made or decisions taken with these contaminated scores could be misleading since it does not represent the true performance of the students. In relation to the finding above, the study further revealed that most students copy from friends during examinations. The finding lends support to the previously reviewed literature. Other studies reported that when teacher-specific and/or classroom specific accountability existed, teachers reported instances of cheating giving direct hints to students or changing student responses on tests and feelings of anxiety and low self-esteem and sometimes use the tests as a filter for making instructional decisions (Dorr-Bremme, Burry, Catterall, Cabello & Daniels, 1983 as cited in Anane, 2008).

3. The researcher again sought to find out how external factors (Parents and Community) contribute to test score pollution in the two selected Districts in the Northern Region. The revelations supported notable research findings. For instance, this study revealed that teachers Parents, and the community encouraging teachers to help increase the performance of students in the BECE by any means possible was also identified as one of the factors that contribute to test scores pollution in the districts. In the work of Thomas, Hertzig, Dryman, and Fernandez

(2001) they postulated that culture plays an important role in the success of the learners and their test scores. According to their studies, the ethnic background of the tester may have an impact on his/her test results. The familiarity and attitude of their families or culture toward the target language community may also have an influence on test results. Social capital, support, money, and opportunities available for them may also pollute test score interpretations and uses. The above arguments simply point to the reality. Schools institutes, universities, and their staff and personnel are blamed or praised for the success or failure of the students for test results but in many instances, the external factors may influence the results and inferences of test scores. Similarly, Haladyana (1992) revealed that the Community and Parents put blame on the teacher if the students do poorly on a standardized test without considering the multiple questions such as did the content of the test match the breadth and depth of the content taught in the classroom? All the above are consistent with the current findings.

4. Furthermore, the researcher sought to find out the effects of test score pollution on schools in the two selected Districts in the Northern Region. The revelations supported notable research findings. For instance, this study revealed that the inability of a school to produce higher grades in BECE creates a bad image for the school. The findings give evidence to the study of Madaus (1990) who revealed that the effect of bad test scores brings about a bad image to the students, school, and teachers. He explained that the perception of

students, teachers, or administrators is that the results of the test will have a relevant impact on them as individuals or on their institution. Revelations from the study indicated that many students leave schools with lower grades to schools with higher grades in BECE when they are about to complete school. The result also shows that headmasters of low-performing schools are seen as not effective.

5. The study again sort to compare the incidence of test score pollution in the two selected districts. The incidence of test score pollution was found to be higher in Yendi than in Saboba. This finding indicates that the test score pollution may vary from one geographical location to the other. Even with the two districts having similar characteristics, there was still a significant difference in test score pollution. This clearly points to the fact that the incidence of test score pollution may even differ among schools in the same geographical location.
6. Finally, the study sorts to compare the incidence of test score pollution in two Educational settings, thus, Public and Private schools. To the expectation of many Educators, the incidence of test score pollution was found to be higher in Private schools than in the Public schools. This finding indicates that the test score pollution may vary from one Educational setting to the other. With the hidden motives of most private schools to make money, one would definitely expert that they raise test scores (both ethical and unethical in either to satisfy parents. In most of these private schools parents can easily have their way out in contributing their quota to pollute test scores since they are always willing to pay money in support of teachers in the private schools to

raise test scores of students. Currently in Ghanaian basic schools were teachers are not allowed to take any form of money from parents, the situation is not the same in the private schools. These and many other reasons could have contributed to the significant difference in test score pollution in Private schools than and the Public schools.

CHAPTER FIVE

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Overview

This chapter summarizes the study's findings, as well as the study's conclusions, recommendations, and research ideas. As a result, the focus of the chapter is on the policy implications of the study's findings and future research. The study's key findings and major conclusions form the basis for the recommendations.

Summary of the Study

The study assessed the sources of test score pollutions in Junior High Schools in the two selected districts in the Northern Region, Ghana. Specifically, the study examined the perceived factors that contribute to test score pollution thus, how test preparation practices, situational factors, and external factors contribute to test score pollution and examine the effects of test score pollution on schools. The quantitative approach using the descriptive survey was adopted for the study. This implies that the study employed quantitative approaches through the use of questionnaires to elicit responses from a sample of teachers. In all, there were 265 teachers selected from the two districts (Yendi and Saboba). The data were analyzed using the Statistical Package for Social Sciences (SPSS version 22) to generate descriptive statistics (means and standard deviations) and inferential statistics (Independent sample t-test).

Summary of Key Findings

The following is a summary of the study's findings:

1. The research discovered that the incidence of test scores pollution can take place at the time of preparing students for standardized examinations. From the data in Table 6, the overall mean and standard deviation (M=3.28, SD=.624) indicates strongly that test preparation practices contribute to test score pollution in the two districts. Some of these significant factors include the fact that teachers spend more time preparing students for WAEC examinations, most teachers teach students on test-taking strategies, teachers prepare their scheme of work based on what is included in the BECE. The study again revealed that teachers modify the order of the curriculum based on what is mostly included in the BECE, teachers encouraging and motivating students to perform in an examination, students cheating in the form of copying from friends during the examinations, and finally, teachers encouraging parents and their colleague teachers to help motivate students to perform during the examination as well as Parents and community edging and encouraging teachers to increase the performance of students by any means possible. These factors were among other factors identified as key preparation practices that contribute to test scores pollution.
2. The data in Table 7 shows an overall mean and standard deviation M=3.16, SD=.658. This indicates strongly that situational factors (test administration situation) contribute to test score pollution in the two districts. It was found that anxiety among students during standardised

Examinations lead students to send foreign materials into examination halls, Copying from friends during examinations are the key situational factors (test administration) that contribute to test score pollution in the two districts.

3. The study also revealed that external factors (parents and community) contribute to test score pollution in the two districts. The data in Table 8, indicates an overall mean and standard deviation (M=2.98, SD=.733) reports strongly that external factors (parents and community) contribute to test score pollution in the two districts. Some of these identified factors include Parents persuading teachers to organize extra classes for their wards when the time for BECE is drawing near, the attitude of parents mostly put pressure on teachers to try any means possible for their students to perform well during BECE, parents and the community mostly blaming teachers for the failure of their wards and some parents buying examination questions for their wards.
4. The overall mean and standard deviation M=3.21, SD=.663 as shown in Table 9 indicates that the majority of the teachers affirmed the fact that test score pollution affects the schools in many ways. The study revealed through its findings that the effect of test score pollution includes, the creation of a bad image for the school as a result of its inability to produce higher grades in BECE, teachers in poor-performing schools are seen as not teaching effectively, many students leaving poor-performing schools to schools that score higher grades in

BECE and the administrators (headmasters) of schools with poor performance in BECE are seen as not effective in their schools.

5. There was a statistically significant difference in test scores pollution between Yendi and Saboba Districts. The results indicated that test score pollution was higher in Yendi municipality than in Saboba district.
6. There was a statistically significant difference between test scores pollution between Public and Private schools. The results show that test score pollution was higher in Private schools than public schools.

Conclusions

Based on the research findings, the researcher came out with the following conclusions;

First of all, teachers are so much involved in some practices at the time they are preparing students for a standardized test. These practices pollute test scores. This confirms survey results that consistently indicated that teachers' practice of testing in the schools is not at its optimal best. However, in as much as teachers are faulted for poor practices that pollute test scores, it is a result of the pressure of accountability where test scores are used as a measure of quality and improvement on Education. It appears from the study that the influence of testing affects meaningful teaching and learning within schools. The reason for the test score pollutions in the two districts could be as a result of the perception of teachers that the WAEC examination is a tool of accountability of the quality of teaching in their school.

Again this study revealed a number of test administration factors that influence test scores. It can therefore be concluded that authorities

(examination bodies) seem to be poor in the implementation of examination rules. This is because some Examination malpractices such as the sending of foreign materials into examination halls by students, students copying from each other, parents buying examination questions for their wards were revealed in this study as contributors to test score pollution.

Finally, the pressure from parents and the community on schools to increase test scores is another revelation of this study. One will not be wrong to conclude that Parents just want to see their wards score high grades at all costs, either ethical or unethical. The result of this attitude of parents is the contamination of test scores.

Recommendations

The following recommendations are made based on the findings of the investigation and the conclusions reached.

In the first place, the government should consider reviewing the Educational system that measures the quality of educational improvement with the use of test scores since this increases the stakes of standardized Examinations and as a result mount unbearable pressure on teachers to increase test scores to the neglect of meaningful learning.

The Ghana Education Service (GES) should collaborate with institutions like the University of Cape Coast (Department of Education and Psychology) to provide appropriate in-service training or programmes on testing practices (courses) to teachers since the initial training of most of these teachers lacked vital assessment practices. This will help to prevent the situation of teachers spending much time in class in preparing students for external tests to the neglect of meaningful learning.

Again, the government should charge WAEC to strengthen the implementation of examination standards in the country. WAEC should be charged to produce valid and reliable test results by minimising malpractices among students and schools during examinations.

Finally, parents should be educated on the reasons for assessment and also on the effective ways of monitoring and discussing their wards' performance with teachers. Parents should be given education on the need for their wards to be given meaningful learning but not higher grades.

Suggestion for Further Research

Apart from the general recommendations which have been elaborated in this research report, it is suggested that other academic studies be conducted around this present topic and in different locations. The following are some suggested areas that can be considered for further studies.

1. A similar study in other districts in the Northern region and other parts of the country is recommended. This will give more generalisations of findings.
2. Again, I suggest a replication of this study at higher levels like the Senior High schools. This would be necessary to identify the effect of test scores pollutions on teachers and students at the senior high levels and beyond.

REFERENCES

Adeyemi, T. O. (2010). *Examination malpractices among secondary school students in Ondo State, Nigeria: Perceived causes and possible solutions:* Nigeria. *Journal of Education Administration and Policy Studies, 2(3), 48-55.*

Agbenyo, S. (2017). Determinants of test score pollution in Junior High Schools in three Districts in the Volta Region of Ghana. *Unpublished master's thesis*, University of Cape Coast, Ghana.

Akaranga, S. I., & Ongong, J. J. (2013). *The phenomenon of Examination Malpractice: An Example of Nairobi and Kenyatta Universities,* Kenya *Journal of Education and Practice*, *4* (18), 87 – 96.

Akpalu, P. (1986). Predictable losers in testing schemes. *School Administrator, 57*(11), 6-9.

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

Amedahe, F. K. (2000). *Continuous assessment.* Unpublished paper, University of Cape Coast, Ghana.

Amedahe, F. K. (2004). *Notes on educational research.* Unpublished document, University of Cape Coast, Ghana.

Anane, E. (2008). *High stakes testing, curriculum implementation and instruction.* University of Cape Coast. Ghana.

Anane, E. (2015). Influence of accountability pressure on science, mathematics and English language teachers' practices in senior high schools in Ghana. *Ghana Journal of Education, volume 1.* University of Cape Coast. Ghana.

Anastasi, A. (2002). *Psychological testing*. New York: Macmillan Publishing Company.

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

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

Azizeh, C., & Mansoor, T. (2010). *Sources of test score pollution: State of the art*. Published master's thesis. Islamic Azad University, Khorasgan Branch (Isfahan).

Bennett, R. E., & Ward, W. C. (2003). *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment.* Hillsdale, NJ: Erlbaum.

Best, J. W., & Kahn, J.V. (2007), *Research in education*. New Delhi, Hall of India Private.

Birenbaum, M., & Pinku, P. (2007). Effects of test anxiety, information organization, and testing situation on performance on two test formats**.** *Contemporary Educational Psychology*. 22/1, 23-38.

Carroll, J., B. (2008). *Human cognitive abilities: A survey of factor-analytic studies. Cambridge,* England: Cambridge University Press. Available at http://www.questia.com/PM.qst?

Cassady, J. & Johnson, R. (2002). Cognitive test anxiety and academic performance. *Contemporary Education Psychology*, 27 (2). 270.

Chapman, D. W., & Snyder, C. W. (2000). Can high stakes national testing improve instruction: Re-examining conventional wisdom. *International Journal of Educational Development, 20*(5), 457-474.

Chatterji, M. (2003). *Educational assessment*. Boston: Allyn and Bacon.

Cheryl, W., & Robin, D. (2000). *Effect of increased test-taking time on test scores by ethnic groups, age, and sex.* Princton, NJ: Educational Testing Service Publications Order Services.

Cohen, L., Manion, L., & Morrison L. (2000). *Research methods in education* (5th ed.). London Routledge Falmer.

Cronbach, L. J. (1975). Five decades of public controversy over mental testing. *American Psychologist, 30*, 1-14.

Cronbach, L. J. (2008). *Essentials of psychological testing* (2nd ed.). New York: Harpers and Brothers Publishers.

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

Curtis, E. M., & Glaser, R. (1981). Changing Conceptions of Intelligence *Review of Research in Education, 9*,111-148.

Dietel, R. J., Knuth, R. A. & Herman, J.L. (2002).*What does the research say about assessment?* NCREL, Oak Brook. Available online: www.ncrel.org/sdrs/areas/stw_esys/4assess.htm.

Dolly, J., & Williams, K. (1983). *Teaching test wiseness*. Paper presented at the Annual Meeting of the Northern Rocky Mountain Educational Research Association, Jackson Hole, WY.

Dorr-Bremme, D., Burry, J., Catterall, J., Cabello, B., & Daniels, L. (1983). The costs of testing in American public schools (CSE Tech. Rep. No. 198). Los Angeles, CA: University of California, Center for the Study of Evaluation.

Dreisbacha, M., & Keogh, B. (2002). Testwiseness as a factor in readiness test performance of young Mexican-American children. *Journal of Educational Psychology, 74* (2). 224-229.

DuBois, P. H. (2000). A test-dominated society: China 115 B. C. – 1905 A. D. Proceedings of the 1964 invitational conference on testing problems. In Anastasi (Ed.). *Testing problems in perspective*. (pp. 51-63). Washington D. C. American Council on Testing.

DuBois, P. H. (2005). *A history of psychological testing.* Boston: Allyn and Bacon Inc.

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

Etsey, Y. K. (2005) Assessing Performance in Schools: Issues and Practice. *Ife Psychologia 13*(1): 123-135.

Filson, Y., & Brown, M. (2008). Effects of test anxiety and testing conditions on intelligence scores of elementary school boys and girls, *Psychological Reports, 32* (2), 643-49

Flippo, B., & Caverly, D. C. (2008). (Eds.), *Handbook of college reading and study strategy research* (pp. 121-144). New York: Rouledge.

Flanagan, D, Genshaft, J. L. & Harrison, P. L. (2007). *Intellectual assessment, tests, and issues.* New York: The Guilford Press.

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

Gardner, R., & Moorcroft, R. (2005). The role of attitudes and motivation in second language learning: Correlational and experimental considerations. *Language Learning, 35(* 207 – 227).

Gay, R. L. (1992). *Educational research: Competencies for analysis and application.* (4th ed.) New York: Macmillan Publishing Company.

Genshaft, J., & Kirwin, P. (1990). Improving the study skills of at-risk students. *Special Services in Schools*, 5 (3), 109-130.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36*, 923-936.

Glaser, R., & Silver, E. (1994). Assessment, testing, and instruction: Retrospect and prospect. In L. Darling-Hammond, (Ed.), *Review of Research in Education,* (pp. 393-419): Washington, D.C.: American Educational Research Association.

Gordon, S. P., & Reese, M. (2007). High-stakes testing: Worth the price? *Journal of School Leadership,* 7, 345-368.

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

Hambleton, R. K., & Jones, R. W. (1993).Comparison of Classical Test Theory and Item Response Theory and their Applications to Test Development. Educational Measurement: Issues and Practice, 12(3), 38-47.

Haladyna, T. (1992, August). Test score pollution: Implications for limited English proficient students. In *Proceedings of the Second National*

Research Symposium on limited English proficient student issues: Focus on evaluation and measurement (Vol. 2, pp. 135-163).

Haladyna, T. (2002). Generic questioning strategies for linking teaching to testing. Educational technology: *Research and Development*, *39* (l), 73-82.

Haladyna, T., Haas, N., & Nolen, S. (1990). *Test score pollution*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Haladyna, T., Nolen, S., & Haas, N. (2000). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher, 20* (5), 2-7.

Hargett, G. (1992). *Test score pollution: Implications for limited English proficient students.* In *Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues: Focus on Evaluation and Measurement*. Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Language Affairs.

Hoover, J. P. (2002). A dozen ways to raise students' test performance. *Principal, 81* (*3*), 17-18.

Kamin, L. J. (1974). *The science and politics of IQ*. Hillsdale, NJ: Erlbaum.

Kenworthy, R. (2006). Timed versus at-home assessment tests: Does time affect the quality of second language learners' written compositions? *TESOL,10*(1).

König, G. (2006). Orientation, motivation and attitudes of Turkish University students learning a second foreign language. *Zeitschrift für*

Interkulturellen Fremdsprachenunterricht, *11* (1), 10. Retrieved October from http://www.ualberta.ca/~german/ejournal/Koenig6.htm.

Krejcie, R. V., & Morgan, D. W. (1970). *Determining sample size for research activities. Educational and Psychological Measurement, 30,* 607-610.

Kubiszyn, T., & Orich, G. (2006). *Educational testing and measurement: Classroom application and practice.* New Jersey: Scott, Foresman and Company.

Lambert, W. (2003), Psychological approaches to the study of language Part II: On the second language learning and bilingualism. *Modern Language Journal 14*, 114-121.

Linn, R. L., Baker E. L., & Dunbar. S. B. (2002). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20 (8)*, 15-21.

Linn, R. L., & Gronlund, S. J., (1995). *The mismeasure of man*. New York: W. W. Norton.

Madaus, G. (1990). Testing and the curriculum. In L.N. Tanner (Ed.), (pp. 83-121). *Cultural issues in curriculum. 87th Yearbook, Part I, National Society for the Study of Education*. Chicago: University of Chicago Press.

Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory using Derived Test Data. The International Journal of Educational and Psychological Assessment, Vol.1, Issue 1. Pp. 1-11

Maylone, N. (2004). Do tests show more than "Test Think"? *The Education Digest*, *69* (8), 16-20.

McArthur, D. L. (1987). *Educational assessment: A brief history*. In D. L. McArthur (Ed.), (15-26). Alternative approaches to the assessment of achievement. Boston: Kluwer.

Mehrens, W. A. (2004). National tests and local curriculum: Match or mismatch. *Educational Measurement: Issues and Practices, 3* (3), 9-15.

Mehrens, W. A. & Lehmann, I. J. (2002). *Measurement and evaluation in education and psychology.* (4th ed.) New York: Holt, Rinehart and Winston Inc.

Mercer, J. R. (2000). Alternative paradigms for assessment in a pluralistic society.

Mereku, K. (2000). *The BECE Grading System Committee Report: Implications for Minimum Educational Qualifications for Basic Education Certificate*. Thousands Oaks, CA: Sage.

Mertens, D. M. (2010). *Research and evaluation in education and psychology, integrating diversity with quantitative, qualitative, and mixed methods* (3rd ed.). Thousands Oaks, CA: Sage.

Messick, S. (2004). The psychology of educational measurement. *Journal of Educational Measurement, 21,* 215-237.

Mine, H. (1987). *A study of the effects of child-rearing patterns on test anxiety in late adolescents*. Paper presented at the biennial meeting of the International Society for the Study of Behavioral Development. Tokyo, Japan.

Ministry of Education (2002). *Meeting the challenges of Education in the twenty-first century*. Report of the President's Committee on review of education reforms in Ghana.

Mislevy, R. J., Yamamoto, K., & Anacker, S. (1992). Toward a test theory for assessing student understanding. In R. A. Lesh & S. Lamon (Eds.), (pp. 293- 318). *Assessments of authentic performance in school mathematics.* Washington, DC: American Association for the Advancement of Science.

Moser, C. A. (1952). Quota sampling. Journal of the Royal Statistical Society. Series A (General), 115(3), 411-423.

Moon, T. R., Brighton, C. M., Jarvis, J. M., & Hall, C. J. (2007). *State Standardized Testing Programs: Their Effects on Teachers and Students* (RM07228). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.

Nitko, A. J. (1996). Designing tests that are integrated with instruction. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed), (pp. 447-471). New York: American Council on Education.

Nitko, A.J. (2001). *Educational assessment of students* (3rd ed.) New Jersey: Prentice-Hall.

Oakland, T. (2001). The effects of test-wiseness materials on standardized test performance of preschool disadvantaged children. *Journal of School Psychology, 10* (4), 355-360.

Oduro-Okyireh, K., (2008). *50 years of educational progress and challenges in Ghana.* Paper presented at the Educational Expansion and Access in

Ghana: A review of fifty years of challenge and progress, Parliament offices, Westminster, London.

Oliver, P. (2010). *The student's guide to research ethics*. Berkshire: Open University Press.6, Retrieved, from http://smj.sma.org.sg/5103/5103.

Onyibe, C. O., Uma, U. U., & Ibina, E. (2015). Examination malpractice in Nigeria: Causes and effects on national development: Nigeria. *Journal of Education and Practice*, 6(26), ISSN 2222-1735 (Paper), ISSN 2222-288X (Online)

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

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

Petty, N., & Harrell, E. (1977). Effect of programmed instruction related to motivation, anxiety, and test wiseness on Group IQ test performance. *Journal of Educational Psychology*, *69* (5), 630-635.

Pintner, R. (2003). Supplementary Guide for the Revised Stanford-Binet Scale *Applied Psychology Monographs, 3*. Stanford, CA: Published for the American Association for Applied Psychology by Stanford University Press.

Polit, D. F., & Hungler, B. P. (1999). *Nursing research: Principles and methods* (6th ed.). Philadelphia, Lippincott.

Punch, K. F. (2009). *Introduction to research methods in education*. Thousand Oaks, CA Sage Publications Ltd.

Purpura, J. E. (2004). *Assessing Grammar*. Cambridge: Cambridge University Press.

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Columbia: Teachers College Press.

Shepard, L.A. (1992). *Will national tests improve student learning?* (CSE Technical Report 342). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Smith, M. L. (2002). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20* (5), 8-11.

Stricker, L. J. (2003). Test-wiseness on personality scales. *Journal of Applied Psychology, 53* (3), 1-17.

Tamakloe, E. K., Atta, E. T., & Amedahe, F. K. (2005). *Principles and methods of teaching*. (2nd ed.). Accra: Black Mask Ltd. (p. 181).

Thomas, A., Hertzig, M., Dryman, I., & Fernandez, P. (2001). Examiner effect in IQ testing of Puerto Rican working-class children. *American Journal of Orthopsychiatry, 41* (5), 809-821.

Tyack, D. (1974). *The one best system: A history of American urban education.* Cambridge, MA: Harvard University Press.

Tyler, R. W., (1986). *Testing, teaching, and learning*. Report of a conference on research on testing. Washington, DC: U.S. Department of Health, Education, and Welfare.

Ushie, O. S., & Ishanga, A. R. (2016). Examination malpractice: Causes, effects and possible ways of curbing the menace. A case study of Cross River University of Technology: Nigeria. *International Journal of Managerial Studies and Research (IJMSR), 4*(1) 59-65.

Volante, L. (2006). Toward appropriation for standardized achievement testing. *Journal of Educational Thought, 40* (2). 129-144.

Wagaman, J. (2009). R*aising high school reading test scores: factors that affect reading scores*. Retrieved October 21, 2017, from http://www.suite101.

Wiyalat, N., (2009). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28* (4), 499-509.

Wolf, D., Bixby, J., Glenn, J. G., & Gardner, H. (2008). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), (pp. 11- 16) *Review of research in education*. Washington, D.C.: American Educational Research Association.

Worthen, B. R., & Spandel, V. (1991). *Putting the standardized test debate in perspective. Association For Supervision And Curriculum Development, Educational Leadership* (Feb. 1991 ed).

Wu Man-Fat, M. (2004). An exploration of the role of integrative motivation in the achievement of English language learners in Hong Kong. *Karen's Linguistics Issues*. Retrieved from http://www3.telus.net/b linguisticsissues/motivationhk

APPENDICES

Appendix A

RESEARCH INSTRUMENT

UNIVERSITY OF CAPE COAST

COLLEGE OF EDUCATION STUDIES

DEPARTMENT OF EDUCATION AND PSYCHOLOGY

Dear respondent, the purpose of this questionnaire is to seek information on "the perceived sources of test score pollution in Junior High Schools in two selected Districts". The information is needed purely for academic purposes. **Any information provided will be treated with the utmost confidentiality.** Your full participation will help make informed decisions about sources of test score pollution. You are kindly requested to be candid with your responses. Thank you for accepting to be part of this study.

Introductions:

The questionnaire is in five sections, section A consists of personal data, section B is on how test preparation practices influence test score, section C is on Situational Factors (Test Administration Situation), section D focuses on Other Factors (Parents and the Community) and section E deals with the Effects of Test Score Pollution on Schools. **Please tick (√) as appropriate using the following:**

SA- Strongly Agree, A- Agree, D- Disagree, and SD- Strongly Disagree

SECTION A: Demographic data of respondents

1. Sex: Male [ ] Female [ ]
2. Professional Status: Trained [ ] Untrained [ ]

3. Academic qualification: Certificate [ ] Diploma [ ] First Degree [ ] Higher [ ]
4. School Status: Public [ ] Private [ ]

SECTION B: How do test preparation practices influence to test score?

	SSA	AA	DD	SSD
I spend more of class time preparing students for WAEC examinations				
I prepare my scheme of work based on what is included in the BECE				
I teach students test-taking strategies ('test wiseness')				
I don't register students who are academically weak (low achievers) for the WAEC examination				
I give past questions from WAEC to my students to practice				
Most teachers modify the order of the curriculum based on what is mostly included in the BECE				
I encourage students to perform in the examination				
I encourage Parents and my colleague teachers to help motivate students to perform during the examination.				
Parents and the community edge and help teachers to increase the performance of students in BECE by any means possible				

SECTION C: How do Situational Factors (Test Administration Situation) contribute to test scores pollution?

	SSA	AA	DD	SSD
I observe anxiety (fear or panic) among Students during an examination				
I mostly motivate my students to do their best during examinations				
I mostly try to help my students build their self-esteem in other to perform well in examinations				
I encourage students not to stress themselves up during examinations				
I often see students with foreign materials during examinations				
My students sometimes report other students copying from friends in the examination hall to me				

SECTION D: How do external Factors (Parents and the Community) contribute to test score pollution?

	SSA	AA	DD	SSD
Parents discuss BECE results at PTA meetings				
Parents persuade teachers to organize extra classes for their wards when the time for BECE is not drawing near				
Parents persuade teachers to organize extra classes for their wards when the time for BECE is drawing near				

Some parents encourage teachers to solve past questions from WAEC with their wards				
Parents mostly blame teachers for the failure of their wards				
Parents' attitudes mostly put pressure on teachers to try any means possible for their students to perform well during BECE				
Some parents reward teachers when their children get higher scores in BECE				
Parents sometimes want to know if there is leakage of the examination questions				
Parents sometimes want to know if there is the possibility of getting exams questions to buy				
Some parents buy examination questions for their wards				
The community is always interested in the performance of students in BECE				

SECTION E: What are the Effects of Test Score Pollution on Schools?

	SSA	**AA**	**DD**	**SSD**
Poor performance of students in the BECE brings a bad image for the school				
Teachers of schools with poor performance in the BECE are seen as not teaching effectively				
Schools with good performance in BECE are mostly overpopulated				
Schools that perform poorly in the BECE are mostly underpopulated				
The administrators (headmasters) of schools with poor performance in BECE are seen as not effective				

Students leave schools that perform poorly to schools that perform better				
Poor performing schools in BECE do not receive much attention from stakeholders and the community at large				

Appendix B

RELIABILITY TEST RESULTS

Case Processing Summary

		N	%
Cases	Valid	85	100.0
	Excluded[a]	0	.0
	Total	85	100.0

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.871	.896	33

APPENDIX C

LETTER OF INTRODUCTION

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

Telephone: 233-3321-32440/4 & 32489/3
Direct: 033 20 91697
Fax: 03321-30184
Telex: 2552, UCC, GH.
Telegram & Cables: University, Cape Coast
Email: edufound@ucc.edu.gh

UNIVERSITY POST OFFICE
CAPE COAST, GHANA

Our Ref:

Your Ref:

18th June, 2020

TO WHOM IT MAY CONCERN

Dear Sir/Madam,

THESIS WORK
LETTER OF INTRODUCTION: MR. UMAR SAFIANU ABUKARI

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

Mr. Abukari is researching on the topic: **"PERCEIVED SOURCES OF TEST SCORE POLLUTION IN TWO SELECTED DISTRICTS IN THE NORTHERN REGION."**

He has opted to collect or gather data at your institution/establishment for his Thesis work. We would be most grateful if you could provide him 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 (Mr.)
Principal Administrative Assistant
For: **Head**

APPENDIX D

ETHICAL CLEARANCE

UNIVERSITY OF CAPE COAST
COLLEGE OF EDUCATION STUDIES
ETHICAL REVIEW BOARD

UNIVERSITY POST OFFICE
CAPE COAST, GHANA

Our Ref: CES-ERB/ucc-edu/v4/20-75
Your Ref: .....
Date: 1st October, 2020

Dear Sir/Madam,

ETHICAL REQUIREMENTS CLEARANCE FOR RESEARCH STUDY

The bearer, Umar Safianu Abukari, Reg. No. EF/MPHIL/19/0001 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:

Perceived sources of test score pollution in Junior High Schools in two selected districts in the Northern Region, Ghana.

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)

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

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

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